

Facial expression recognition (FER) survey: a vision, architectural elements, and future directions

Sana Ullah, Jie Ou, Yuanlun Xie and Wenhong Tian

School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China

ABSTRACT

With the cutting-edge advancements in computer vision, facial expression recognition (FER) is an active research area due to its broad practical applications. It has been utilized in various fields, including education, advertising and marketing, entertainment and gaming, health, and transportation. The facial expression recognition-based systems are rapidly evolving due to new challenges, and significant research studies have been conducted on both basic and compound facial expressions of emotions; however, measuring emotions is challenging. Fueled by the recent advancements and challenges to the FER systems, in this article, we have discussed the basics of FER and architectural elements, FER applications and use-cases, FER-based global leading companies, interconnection between FER, Internet of Things (IoT) and Cloud computing, summarize open challenges in-depth to FER technologies, and future directions through utilizing Preferred Reporting Items for Systematic reviews and Meta Analyses Method (PRISMA). In the end, the conclusion and future thoughts are discussed. By overcoming the identified challenges and future directions in this research study, researchers will revolutionize the discipline of facial expression recognition in the future.

Submitted 31 January 2024

Accepted 8 April 2024

Published 3 June 2024

Corresponding authors

Sana Ullah, sanauom1@gmail.com

Wenhong Tian,

Tian_wenhong@uestc.edu.cn

Academic editor

Ivan Miguel Pires

Additional Information and
Declarations can be found on
page 25

DOI 10.7717/peerj-cs.2024

© Copyright
2024 Ullah et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Human-Computer Interaction, Algorithms and Analysis of Algorithms, Computer Vision, Neural Networks

Keywords Computer vision, Emotion recognition technology, Basic & compound emotions, Cloud computing, Internet of Things (IoT)

INTRODUCTION

Emotions are important components of our lives, based on which we make numerous life decisions. Facial expression recognition (FER) is critical for computer games, recommendation systems, human-computer interaction, stress control, mental health monitoring, autism monitoring, and many more. Facial expression recognition is crucial for understanding human intentions and emotions, facilitating effective human-machine interaction and communication. Communication is a critical tool in promoting societies to permit humans to adapt, progress, and survive. It is the most powerful signal through which humans exchange intentions and emotional states, and 55% of the communication is performed through facial expressions (*Saganowski, 2022; Haryanto & Agastya, 2022*). Commonly, emotions and intentions are conveyed by humans either through involuntary language, facial expression, and gestures. Based on this, various FER systems have

been designed to address real-world challenges such as driver safety, human-machine interaction, medical assessment, and lie detection ([Appasaheb Borgalli & Surve, 2023](#)). Facial expression recognition technology has applications in different fields, such as access control, law enforcement, entertainment & gaming, transportation, security & surveillance, advertisement & marketing, health & safety, education, finance & banking, and many more. Emotions are divided into two categories, namely basic and compound emotions. The seven basic emotions include disgust, anger, neutrality, sadness, fear, surprise, and happiness. Compound emotions, such as 'happily surprised,' are formed by combining basic emotions, representing the simultaneous appearance of two emotions on a human face. Researchers have worked on basic emotions, but recognizing compound emotions is challenging due to their complex features. Image recognition technology utilizes computers to understand, analyze, and process diverse patterns of objects and targets ([Huo, Yu & Liu, 2023](#)). The evolution of convolutional neural networks (CNNs) revolutionized computer vision. Diverse CNN models have been applied to computer vision tasks like segmentation, object detection, and image recognition. CNNs are popular because they can automatically learn features from an image. Different filters are used in the convolutional layers, and techniques such as pooling are employed to reduce dimensions and increase robustness against variations. With the rapid evolution of technology, different methods have been used at different times to recognize emotions. Traditional feature extraction-based methods, including LBP, HOG, NMF, K-NN, Random Forest, and SVM, were utilized in the early stages. These methods were effective for selected tasks, and manual feature engineering was time-consuming and inefficient for complex image classification ([Mohammadi & Kalhor, 2021](#); [Dargan et al., 2020](#); [Jiao & Zhao, 2019](#); [Mahesh, 2020](#)).

The integration of deep learning methods like CNNs, recurrent neural networks (RNNs), transfer learning, Siamese networks, and generative adversarial networks (GANs) addressed several limitations of traditional machine learning methods. Furthermore, different attention mechanisms were introduced further to improve the performance of deep learning methods, and deep learning methods outperformed the traditional machine learning methods in image recognition and classification. Despite their benefits, attention mechanisms integrated into deep learning methods for improving performance across diverse tasks still struggle with capturing contextual information in data and managing long-range dependencies ([Chen et al., 2021](#); [Mathew, Amudha & Sivakumari, 2021](#); [Kunapuli, 2023](#); [Diniz et al., 2021](#)).

A new type of neural network, the transformer, was introduced, which utilized a self-attention mechanism to extract intrinsic features and showed potential performance in many applications of AI ([Mazzia et al., 2022](#); [Han et al., 2022](#)). When applied for the first time, the transformer achieved significant performance in many tasks related to language processing (NLP). With the strong representation capabilities and achieving significant breakthroughs by the transformer-based models without any fine-tuning in the NLP, the researchers are now inspired and applied transformer-based models to computer vision tasks. CNNs are counted as a fundamental component, but currently, transformers are utilized as the best hybrid component with CNNs for better performance. Digital images are complex and contain high-level information like patterns, scenes, and objects.

The information in digital images can be extracted and analyzed through vision-based algorithms and models; the important insights about the image include extracting features, tracking movements, and recognizing objects. Computer vision is a critical area of research, as its applications extend across nearly all fields. However, extracting high-level meaningful information from an image of compound facial expressions of emotion is challenging for researchers due to background clutter, pose, and brightness variations (*Xie, Tian & Ma, 2020; Tunstall, Werra & Wolf, 2022; Chitty-Venkata et al., 2022*).

This article presents the current trends in facial expressions of emotion technologies research, propelled by applications and the need for convergence in several interdisciplinary technologies. The main contributions and novelties of this article are as follows:

(1) We provide a comprehensive overview of the future vision of facial expression recognition technology and discuss the cutting-edge technologies that will enable its realization. This forward-looking perspective sets the stage for the advancements presented in the article (section ‘Research methodology’).

(2) We offer clear and concise definitions of related terminologies, elements, and trends in face recognition technologies. Furthermore, we have discussed various models and algorithms utilized in FER. By clarifying these key concepts, we establish a solid foundation for understanding the state-of-the-art in this field (section ‘Facial expression recognition technology in the next decade’).

(3) We present a wide range of applications for facial expression recognition technologies and highlight the top companies around the globe that are leading the way in this domain. This industry-focused analysis demonstrates the practical relevance and potential impact of our research (section ‘Facial expression recognition models/frameworks’).

(4) We delve into the interconnection among cloud computing technology, Internet of Things (IoT), and facial expression recognition technologies, providing a detailed discussion on how these technologies converge to enable powerful new applications. This interdisciplinary perspective is a unique contribution of our article (section ‘Facial expression recognition, Internet of Things (IoT), and cloud computing’).

(5) We identify and analyze the open challenges faced by facial expression recognition technologies and propose future research directions to address these challenges. By highlighting these research gaps and opportunities, we set the agenda for advancing the field (section ‘Research methodologies issues’).

(6) We conclude the article with a summary of our findings and offer insightful thoughts on the future of facial expression recognition technologies. This synthesis and forward-looking perspective distinguish our work from existing literature in the field (section ‘Conclusions’).

By presenting these unique contributions and novelties, our article stands out in the crowded field of facial expression recognition research. The comprehensive overview, interdisciplinary analysis, and future-oriented perspective make this work a valuable resource for researchers and practitioners alike.

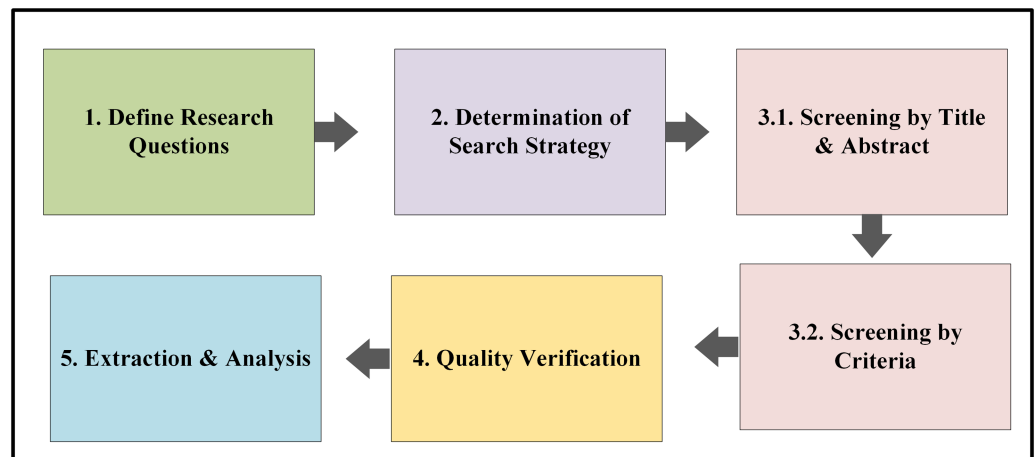


Figure 1 PRISMA review protocol.

Full-size DOI: [10.7717/peerjcs.2024/fig-1](https://doi.org/10.7717/peerjcs.2024/fig-1)

RESEARCH METHODOLOGY

We have utilized the Preferred Reporting Items for Systematic reviews and Meta-Analyses Method (PRISMA); review protocols are developed based on PRISMA guidelines, shown in Fig. 1. Based on PRISMA guidelines, five steps have been followed: (1) defining research questions (RQ), (2) conducting a comprehensive literature search using predefined keywords and databases, (3) screening and selecting articles based on predefined inclusion and exclusion criteria, (4) assessing the quality of selected articles, and (5) extracting and analyzing data from the included studies. This systematic approach ensures the transparency, reproducibility, and credibility of the review process (Page et al., 2021; Lu et al., 2024).

Research questions

Considering the research objectives, the following key research questions are formulated:

RQ 1: What will facial expression recognition technology be like in the next decade, and what about the recent trends in face recognition technology?

RQ 2: What are the standard architectural elements of the facial expression recognition system?

RQ 3: What are the different models/algorithms/architectures used in facial expression recognition?

RQ 4: What are the broad applications of face recognition technology in various fields, and what about the world's top FER technologies companies?

RQ 5: What is the interconnection between FER, the IoT, and cloud computing?

RQ 6: What are the open challenges to FER technology in integrating into various real-world fields?

RQ 7: What are the limitations and challenges concerning FER technology and the future direction for new research in the field of computer vision?

Table 1 shows the search strings with mapping of PICO elements for each request question.

Strategy for searching articles

Searching for relevant research articles from online libraries is an important task. We designed search strings according to SLR guidelines (Kitchenham et al., 2009) and selected six libraries for article searching: IEEE Xplore, ACM Digital Library, Scopus, Web of Sciences, ScienceDirect, Springer, and Google Scholar. The keywords and their synonyms were derived from our research questions (Section ‘Strategy for searching articles’) and included terms such as “facial expression recognition”, “emotion recognition”, “deep learning”, “computer vision”, and “IoT”. We conducted both forward and backward snowballing to identify additional relevant articles.

We screened around 3,245 articles from the target databases according to the guidelines of PRISMA. The details of the number of articles screened from different libraries through the listed keywords used for searching in each library are shown in Table 2.

Criteria for article selection

In the article selection criteria, we followed two phases: primary selection and secondary selection. The primary selection was based on title, keywords, and abstract screening. In the secondary selection phase, we applied the following inclusion and exclusion criteria:

Inclusion criteria:

1. Peer-reviewed conference papers and journal articles
2. English-language articles
3. Studies focusing on facial expression recognition, emotion recognition, deep learning, and computer vision

Exclusion criteria:

1. Non-peer-reviewed articles, such as preprints, theses, and dissertations
2. Articles written in languages other than English
3. Studies focusing on synthetic image generation techniques and other unrelated topics
4. Duplicate articles.

Quality assessments of articles

Apart from the previously followed inclusion and exclusion criteria, we have analyzed each article based on its quality, content, and publication venue.

Data extraction and synthesis

We extracted different information from the selected articles based on our research questions. By following the the guidelines of PRISMA for article selection criteria and articles quality assessment, we have extracted results based on 141 articles.

FACIAL EXPRESSION RECOGNITION TECHNOLOGY IN THE NEXT DECADE

Facial expression recognition technology has a rich history, which evolved over decades. From the 1960s till the present, it has significantly evolved. In the early 1960s, the researchers

Table 1 Search strings with mapping of PICO elements.					
Research question	Population	Interest	Comparison	Outcome	Search string
What will facial expression recognition (FER) technology be like in the next decade, and what about the recent trends in face recognition technology?	Humans or machines	FER in future, trends of FER	No specific comparison group required	Future of FER, recent trends of FER	("facial expression recognition" OR "emotion recognition from faces") AND ("FER trends")
What are the standard architectural elements of the facial expression recognition system?	*	FER architectural elements	*	FER architectural elements	("Standard FER architecture" OR "FER architecture elements")
What are the different models/algorithms/architectures used in facial expression recognition?	*	FER models and architectures	Comparison among FER architectures/Models	FER Models/architectures	("FER architecture" OR "FER models" "FER algorithm")
What are the broad applications of FER technology in various fields, and what about the world's top FER technologies companies?	Humans or machines or company	FER applications, FER tech companies	No specific comparison group required	FER applications, FER tech companies	("FER application" OR "FER technology companies")
What is the interconnection between FER, the internet of things (IoT), and cloud computing (CC)?	Humans or machines	FER, IoT, CC	*	Interconnection among FER, IoT & CC	("FER" AND "IoT" AND "CC")
What are the open challenges to FER technology in integrating into various real-world fields?	*	FER challenges in diverse domains	*	Extract FER challenges in diverse domains	("FER challenges OR "FER limitations")
What are the limitations and challenges concerning FER technology and the future direction for new research in the field of computer vision?	*	FER technology challenges, FER future directions	*	Extract FER challenges & future directions	("FER challenges OR "FER future directions")

Table 2 Summary of search strategy and results.

Database	Keywords	Number of papers
ACM Digital Library	Facial expression recognition, emotion recognition, deep learning, computer vision, IoT, compound emotions,	708
IEEE Xplore	*	230
ScienceDirect,	*	821
Springer	*	146
Google Scholar	*	720
Scopus	*	357
Web of Science	*	263

were more focused on developing algorithms for basic features like mouth, nose, and eye for identification and analysis from photographs. [Andrejevic & Selwyn \(2022\)](#) developed a system for extracting certain photo features.

In the 1980s and 1990s, the Eigenfaces techniques were developed to recognize facial expressions of emotion. It was working based on the principal component analysis (PCA) approach. Other methods, such as sparse representation, manifold learning, and linear subspace, were also developed for facial recognition. In this period, most methods were features-based, like the Elastic Bunch Graph Matching (EBGM) algorithm and the Active Appearance Model (AAM) ([Turk & Pentland, 1991a](#); [Turk & Pentland, 1991b](#)).

In the years following 2,000, local feature-based methods like local binary patterns (LBP) and Gabor features started gaining popularity in computer vision for face recognition. In this era, machine learning and statistical methods such as hidden Markov models (HMM) and support vector machines (SVM) were introduced for robust and accurate identification. In the 2010s and onwards, the deep learning-based models era started. The convolutional neural networks (CNNs) models were introduced, and later, in the 2012s, the deep learning-based models gained more popularity due to the success of AlexNet in the ImageNet competition ([Xue-fang & Tao, 2013](#); [Krizhevsky, Sutskever & Hinton, 2012](#); [Taskiran, Kahraman & Erdem, 2020](#)).

There is more research on basic emotions but less on compound emotion recognition, and researchers have been working more recently on compound emotion recognition. Most of the recently introduced compound emotions combine two basic emotions. However, compound emotion can be the combination of more than two basic emotions, so future research on compound emotion will be considered and observed. The compound emotions introduced a few, and new compound emotions will be introduced in future research. The recognition is based on visual modality, and it will be better to introduce more modalities of detailed and complementary feature extraction ([Li & Li, 2020](#)).

The perception and production of compound emotion categories open a new level of complexity in computer vision for human–computer interfaces, designing computer vision, social communication, and human cognition for researchers. The important areas of researchers’ interest in compound emotion recognition are cognitive impairments and psychiatric disorders. It is also the point of investigation for researchers in future research

on which the researchers have focused on whether the cognitive processes and cognitive representation are the same for compound emotions and basic emotions or these are different for basic and compound emotions (Du, Tao & Martinez, 2014; Martinez & Du, 2012).

The latest research in computer vision shows that emotional features and emotions change from culture to culture. Computer scientists have focused on developing deep learning models to overcome cultural issues in emotional features due to culture. The cultural evolutionary research on emotions is in the early stage, and researchers are developing strategies and methods for basic and compound emotion recognition through applying evolutionary hypotheses in emotions study (Jack et al., 2012; Lindquist et al., 2022; Brooks et al., 2023).

Novel deep learning models will be developed while considering intra-class variations, including noise, age variations, illumination variations, pose, and occlusion. The deep learning models will be evaluated on the datasets with diverse images. New deep-learning approaches will be developed for real-time emotion recognition, which will detect emotions for untrained classes in the future (Kuruvayil & Palaniswamy, 2022).

Numerous research works have been performed on facial expression recognition and body pose, and some methods have been performed efficiently in specific settings. However, their performance could be better in unconstrained and natural real-life environments. In the future, the researchers will be focusing on developing new models while considering the combined information of the person (body bounding box) and scene context information (whole image) for better performance in natural and unconstrained environments (Kosti et al., 2019).

The facial expressions of emotion changed with changes in gender, culture, and age. An ideal perfect dataset contains images to facilitate cross-cultural, cross-gender, and cross-age research through deep learning methods. The research on pose variant issues and occlusion in emotion recognition received less attention from researchers due to the unavailability of a large dataset having head-pose and occlusion annotations. The researchers are trying to create a dataset that fulfills the above mentioned issues with images of complex scenarios and natural environments instead of controlled ones. Some emotions are created through the combination of basic emotions with complex features. To properly recognize compound emotions, there is a need for a novel deep learning model to extract compound emotion features appropriately (Benitez-Quiroz, Srinivasan & Martinez, 2016; Barsoum et al., 2016; Li & Deng, 2020).

Diverse facial expression recognition algorithms have been developed for practical and theoretical applications. Recently, the researchers' focus shifted from a controlled environment to challenging wild environmental conditions. There are two main challenges in the wild environmental conditions, including pose variation and occlusion. Some of the algorithms have been developed for the wild environment. However, those algorithms are built for specific defined conditions, adapt to limited variations, and can lead to low performance under some challenging wild environment conditions. In the future, researchers are working to develop new models to overcome the challenges of pose

variation and occlusion and apply them practically in a complex real-world environment ([Huang et al., 2019](#)).

Recently, in computer vision, researchers focused on sex differences in facial expressions of emotion. Little research has been done on it, but it illustrates that females outperformed males in accuracy over males. In the coming research, researchers are focusing on intensity variations and sex differences in emotion recognition and sex differences will be considered while developing or proposing new models ([Wingenbach, Ashwin & Brosnan, 2018](#); [Cherbonnier & Michinov, 2021](#); [Helgeson, 2020](#)).

The security of life, assets, and data is the utmost concern. Face recognition biometrics and IoT technology-based security authentication have gained prominence. A face recognition-based door lock system has been developed. In the future, it can be extended to large buildings and organizations where more people are living. Furthermore, the Payroll system can also be integrated into the system for daily wage workers based on face recognition ID. For effectiveness and efficiency, the iris system can be integrated into the current system ([Bagchi et al., 2022](#)).

The smart city makes the traditional services for the inhabitants sustainable, efficient, and flexible through smart technology, smart healthcare, smart energy, smart transportation, and smart infrastructure. To make cities more responsive and efficient in the future, three emerging technologies, such as face recognition, the IoT, and cloud computing, can be appropriately integrated ([Ibrahim Khalil & Kamel, 2024](#)).

Intelligent face recognition devices are utilized in autonomous vehicles. A Face Recognition and Emotion Detection based on IoT (FRED-IoT) is proposed for tracking the driver's face recognition and emotional state in autonomous vehicles. It can be more efficient if we collectively utilize visual and sensory knowledge for identification and managing multimodal production of knowledge in the future digital era ([Chen, Feng & Zhang, 2022](#)).

Recently, deep convolutional neural networks have achieved an accuracy of around 99% on the large dataset collected in the wild. Several studies demonstrated that the performance of deep learning models is recorded low in adverse situations, such as images with noise, blur, or illumination variations. Researchers have focused on developing deep learning models to be robust in adverse conditions ([Wang & Deng, 2019](#); [Grm et al., 2018](#); [Pala & Erdem, 2019](#); [Mohapatra et al., 2019](#)).

The IoT technology greatly contributes to our daily lives in almost all sectors. Different IoT devices, such as cameras, are utilized in hospitals for surveillance scenarios. For the hospital's security, face recognition is an important element to be used in patient traffic analysis, patient fraud detection, hospital security, emotions detection, and patient sentiment analysis. The utilized automatic intelligent face recognition systems work in a real environment, but the accuracy still needs to be improved in a controlled environment. Scientists are working on an intelligent face recognition system to enhance accuracy in real environments like smart healthcare ([Masud et al., 2020](#)).

Diagnosis of autism disease at the initial stages is very challenging as it is an abnormality in the brain. It can be diagnosed only through facial expression recognition because the facial patterns of autistic children are different from those of neurotypical children.

Different face recognition-based systems have been developed based on six basic emotions. In the future, researchers will work to modify and enhance the existing systems for basic emotions and compound emotions to make them more effective and advanced ([Talaat, 2023](#)).

Integrating edge computing deep learning-based face recognition technologies can enhance data processing, instant decisions, and real-time processing with time delay. In facial expression recognition, edge computing technology is a “missing concept.” The researchers are focusing on integrating edge computing in complex emotion recognition systems. All devices are interconnected for information sharing, and real-time instant decisions are made on the edge with limited process power ([Yang et al., 2021](#)).

The 3D geometric details of the human face helped in 3D face recognition. It utilized a 3D sensor for the collection of data. The 3D facial model has improved facial expression recognition in the last ten years. Computer vision researchers have focused on resolving the issue of face recognition like makeup or beautification, pose, and sensitivity to light conditions ([Adjabi et al., 2020](#)).

In autonomous driving, the face recognition technology plays a vital role. In the future, the researchers have focused on making the system more intelligent by integrating an efficient recognition system. However, there are still numerous challenges at the infrastructure, architectural, algorithm, and system levels, which the researchers have focused on fixing ([Gill et al., 2022](#)).

Tremendous research works have been conducted on facial expressions of emotion to improve through the utilization of deep learning models. In the future, scientists are trying to develop an emotionally oriented deep-learning model by additionally integrating Internet-of-Things sensors to improve the recognition accuracy even for micro-expressions or compound emotions ([Ko, 2018](#)).

Huge research has been conducted on emotion recognition, but no standard data annotation rules exist. The commonly utilized annotation techniques are continuous and discrete, and there is no compatibility among datasets due to uniform standards for data annotation. For the description of different emotions, there is a need for different intensities. The currently available datasets are annotated with eight, six, or four emotions. The datasets are incompatible, and recognition models do not perform equally on these datasets. The researchers are trying to overcome the issues above by introducing a hybrid data annotation technique in which individual emotion is annotated according to its energy. While developing a hybrid data annotation technique, the researchers should care about selecting new data annotation standards because it may lead to imbalanced data. The biggest challenge in designing a model is generalizing it for new datasets or unseen data. It happened mostly due to no standard data-splitting techniques and limited sample data. Subject-independent and subject-dependent data splitting techniques have been used; the research shows that subject-independent techniques show better generalization. However, the subject-independent models are not more intelligent and cannot perform better in real-time and realistic applications. The researchers are trying to design subject-independent models that perform better in realistic, challenging environments ([Ahmad & Khan, 2022](#)).

The available existing datasets have limited samples or need to be more balanced. The deep learning models are data-hungry models that need huge datasets for better performance. Different augmentation techniques have been used to overcome overfitting issues. The datasets are required to capture the expressions from all possible angles for better realistic recognition, and the models should be efficient in real environments ([Sajjad et al., 2023](#)).

The education sector continuously evolves for an online connected environment by integrating the latest technologies. For future education models, diverse advanced technologies such as IoT, digital simulation tools, virtual reality, and artificial intelligence (AI) must collaborate. The researchers are working to make virtual or online learning management systems efficient by integrating the latest facial emotion recognition technology ([Ahaidous, Tabaa & Hachimi, 2023](#)).

Definitions, trends and elements

Definitions

Emotion and basic emotion: Emotion can be defined as the transient agitation produced due to joy, surprise, fear, and others ([Suri & Gross, 2022](#)). It can also be defined as an emotion, a complex physiological and psychological state produced in response to a stimulus, experience, or specific event ([Cabanac, 2002](#)). Descartes said that emotion is an important element used for communication and interaction among people and is a mediator between stimulus and response ([Singh, 2012](#)). The six basic emotion categories, such as surprise, sadness, happiness, fear, disgust, and anger, were first introduced by [Ekman, Sorenson & Friesen \(1969\)](#) and [Ekman \(1992\)](#).

Compound emotion: The compound emotion is the combination of two basic emotions simultaneously. For example, if a person is simultaneously happy and surprised, happy and surprised emotion features will appear on his face. There are 12 compound facial expressions of emotions, including disgustedly surprised, angrily disgusted, angrily surprised, fearfully disgusted, fearfully surprised, fearfully angry, sadly disgusted, sadly surprised, sadly angry, sadly fearful, happily disgusted, and happily surprised ([Du, Tao & Martinez, 2014](#); [Pendhari et al., 2022](#)).

Architectural elements of facial expression recognition system

Recognition systems, also known as pattern recognition systems, are designed to classify and identify features or patterns in data. These systems have been widely used in various fields, such as handwriting recognition, face recognition, speech recognition, and computer vision. The basic architectural elements of a face recognition system include face detection, preprocessing, feature extraction, and face recognition ([Koutroumbas & Theodoridis, 2008](#); [Rao & Reddy, 2011](#); [Kasar, Bhattacharyya & Kim, 2016](#)).

The face recognition process begins with the face detection step, where the input image is analyzed to determine the presence of face images. If face images are verified, the system proceeds to the preprocessing step, which involves the application of various techniques to remove unwanted effects such as shadowing, varying lighting, blur, and noise. Once the image has been cleaned and optimized, it moves on to the feature extraction phase.

During feature extraction, different algorithms are employed to extract important features from the image. This process involves noise cleaning, salience extraction, dimension reduction, and information packing. The facial patch is then transformed into a set of fiducial points along with their corresponding locations. These extracted features serve as the basis for the final step: face recognition.

In the face recognition step, the system performs automatic identification and verification based on the extracted features. Face recognition methods can be further classified into global feature-based and local feature-based approaches. Global feature-based methods consider the entire face as a single entity and extract features from the whole face, while local feature-based methods focus on specific facial regions and extract features from these localized areas.

The architectural elements of a face recognition system work together to ensure accurate and efficient identification and verification of individuals. By following this structured approach, face recognition systems can be applied in a wide range of scenarios, such as security, surveillance, and access control. As research in this field continues to advance, we can expect to see further improvements in the accuracy, robustness, and versatility of face recognition systems ([Rao & Reddy, 2011](#); [Kasar, Bhattacharyya & Kim, 2016](#)). [Figure 2](#) shows the architectural elements of FER system.

Trends

FER is a hot and trending topic in computer vision due to its tremendous applications in different fields. The applications of FER are not only limited to human–computer interaction (HCI) and computer vision, but it has diverse applications in different fields such as robotics and games, security, medicine, marketing, education, and many more. Intensity estimation and emotion recognition are the recent focus of the research ([Ekundayo & Viriri, 2021](#)). The popularity of different paradigms changes with time. We have searched Google search trends from 2004 till date for the Internet of Things, deep learning, facial expression, facial recognition systems, and machine learning from Google Trends. Recently, facial expression and facial recognition systems are gaining popularity ([Google Trends, 2024](#)).

Applications

Due to the rapid innovations and integration of facial expression recognition technologies, companies are rapidly growing worldwide. These companies have healthy competition due to the strong commercial potential of the applications of facial expression recognition technologies ([Ali et al., 2021](#); [Karnati et al., 2023](#)). [Figure 3](#) shows the different applications of facial expression recognition.

Education

With the COVID-19 pandemic, traditional education has abruptly shifted towards online learning systems, presenting significant challenges. In face-to-face interactions, teachers can easily determine whether students are learning properly. The emerging facial expression recognition systems and diverse IoT devices, such as vision sensors and cameras, create an efficient student-teacher online interaction. Online learning is improved by integrating

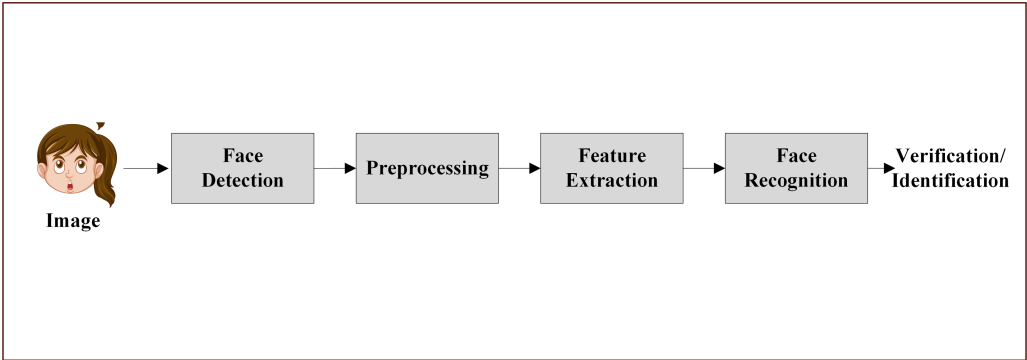


Figure 2 Architecture elements of FER system. Image source credit: Girl, Image by brgfx on Freepik (<https://www.freepik.com/>).

Full-size DOI: 10.7717/peerjcs.2024/fig-2

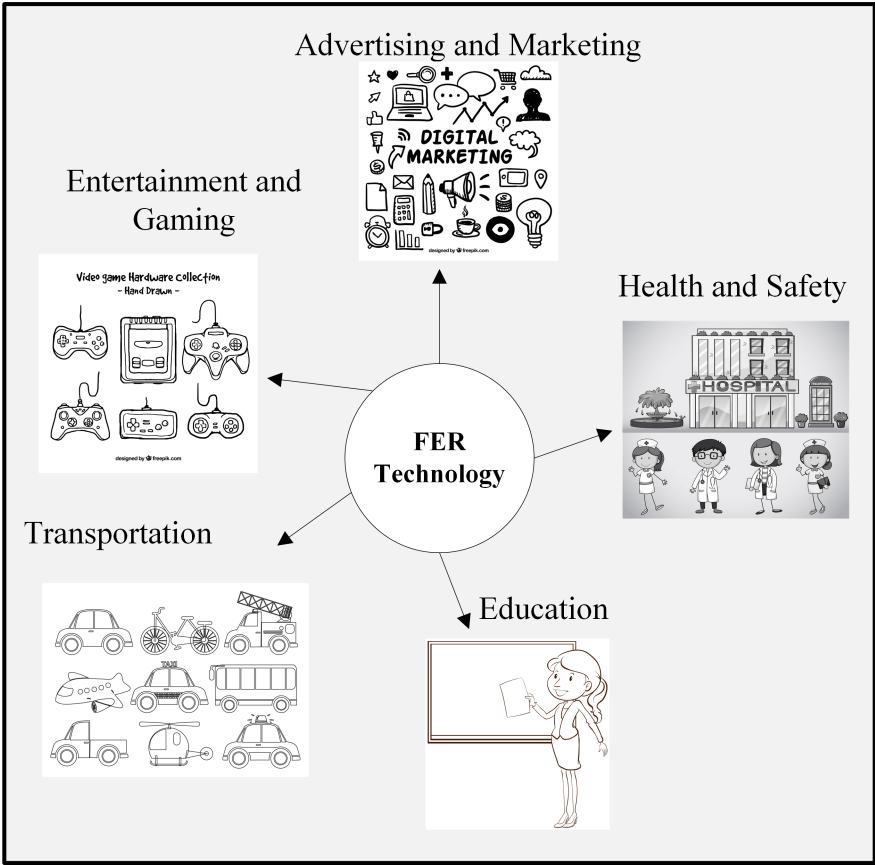


Figure 3 Applications of facial expression recognition. Image source credits: Doctors and nurses working at the hospital, Image by brgfx on Freepik; Digital marketing doodles, Image by Freepik; Collection of hand-drawn video game, Image by Freepik; Different kinds of transportations, Image by brgfx on Freepik; A plain sketch of a teacher, Image by brgfx on Freepik (<https://www.freepik.com/>).

Full-size DOI: 10.7717/peerjcs.2024/fig-3

FER Enabled Education Aids IoT System, which can detect students' emotional states and engagement levels, enabling personalized learning experiences ([Geng, Meng & Dou, 2022](#)). For efficient learning in distance learning and smart classrooms, a deep learning model is utilized based on facial expression recognition to detect the facial expressions of the students in real-time for student concentration analysis. This FER-based framework enabled teachers to create an efficient online learning environment ([Hou et al., 2022](#); [Wang et al., 2020](#)).

Advertising and marketing

Facial expression recognition systems have been used in marketing and customer experience. The FER system observes customers' reactions to services, products, and advertisements, which enables the companies to build the best marketing strategies based on customer feedback ([Van Thanh, 2023](#)). Facial expression recognition architectures are used in diverse emotions recognition which the market researchers used for customer feedback analysis and sentiment analysis for planning future marketing and business strategies of companies ([Srivastava & Bag, 2023](#); [Van Thanh, 2023](#)).

Entertainment and gaming

Facial expression recognition technology can also be utilized in games and the entertainment sector. It is embedded in virtual reality and augmented reality applications-based games for an extensive and realistic experience. It can be utilized in social media platforms for relevant content recommendations based on the collected information ([Ansari et al., 2022](#)). FER in virtual reality (VR) revolutionized the entertainment sector. In VR games, a dynamic world is created to respond according to the player's movements and facial features. The FER revolutionized the entertainment industry by delivering engaging and unique experiences. FER technology is also used in getting audience feedback from facial expressions ([Zhang & Fort, 2023](#)).

Health

Facial expression recognition algorithms can identify a wide range of emotions, such as surprise, anger, sadness, and happiness. Mental health professionals can utilize these algorithms to monitor patients' emotional well-being. Experts in mental health monitoring, customer feedback analysis, and market research can utilize it ([Srivastava & Bag, 2023](#)). The FER system is employed in patient health monitoring. It detects the patient's pain or abnormal condition based on facial features and informs professionals about on-time treatment and health monitoring ([Van Thanh, 2023](#)).

Transportation

FER is one of the most powerful signals of human beings for conveying emotional state. For real-time monitoring of the driver, FER is integrated for detecting drowsiness, fatigue, distraction, or other unfavourable emotional state which affect driving performance. It significantly decreased accidents as it timely alerted drivers ([Xiao et al., 2022](#)). A facial expression recognition-based system is developed called Vibraimage. It detects micro-facial expressions of a passenger in an airport station to declare a passenger suspect or non-suspect.

If a person due to some accidental facial expressions is declared as suspected by the system then he will be passed from screening. It significantly improved transportation security ([Wright, 2023](#)).

FACIAL EXPRESSION RECOGNITION MODELS/ FRAMEWORKS

For the last three decades, facial expression recognition has been one of the prominent research areas in image processing, artificial intelligence, and machine learning communities. Diverse recognition frameworks and models are proposed for the recognition and classification of facial expression recognition. Classical approaches were the initial models introduced around the 1950s. The 2D features were extracted from photographs, and These methods worked on geometrical relationships among facial points. The methods faced many challenges but they provide a basis for advanced face recognition technologies research. Another face recognition system was developed by [Kanade \(1973\)](#). It utilized traditional image processing procedures to extract a vector of 16 facial parameters ([Kanade, 1973](#); [Roberts, 1963](#); [Zhao et al., 2003](#)).

The holistic methods are also called global features-based methods. It performs recognition based on global characteristics. It focuses on the complete face for recognition and not on individual components of the face, such as the nose, eyes, and mouth. The global features extraction was first utilized by [Sirovich & Kirby \(1987\)](#) in principal component analysis (PCA). It was used in independent component analysis (ICA), subspace linear discriminant analysis (SLDA), and linear discriminant analysis (LDA). Later on, another method called support vector machine (SVM) was introduced to recognize emotions ([Bartlett, Movellan & Sejnowski, 2002](#); [Turk & Pentland, 1991a](#); [Turk & Pentland, 1991b](#); [Etemad & Chellappa, 1997](#); [Zhao & Chellappa, 1999](#); [Guo, Li & Chan, 2000](#)).

The hybrid methods are the concatenation of holistic and local methods simultaneously. It is a common trend in research work and utilizes the strengths of the combined methods for improved performance. [Tolba \(2000\)](#) proposed a method that was the hybrid of learning vector quantization (LVQ) and radial basis function (RBF). [Lu, Wang & Jain \(2003\)](#) extended it by the combination of LVQ, RBF, and local binary patterns (LBP). [Lavanya & Hannah Inbarani \(2018\)](#) proposed a method combining tolerance rough similarity (TRS) and PCA. TRS is utilized for extracting similarity index, and PCA for extracting feature matrix ([Gutta & Wechsler, 1997](#); [Lu, Wang & Jain, 2003](#); [Tolba, 2000](#); [Hashemi & Gharahbagh, 2015](#); [Hu et al., 2017](#)).

With the advances in computer vision, different new models have been developed in the modern world of emotion recognition. The modern approaches are categorized into fuzzy logic-based methods, dictionary learning-based methods, and deep learning-based methods. Machine learning methods such as k-nearest neighbors (kNN), artificial neural networks (ANNs) including multi-layer feedforward networks, logistic regression, decision trees, and SVMs are used for classification after the feature extraction phase. These methods have achieved moderate success in their use. Many studies have utilized this classification

pipeline of hand-crafted features and popular classifiers to perform classification (*Rincy & Gupta, 2020; Barata et al., 2014; Blum et al., 2004; Rastgoo et al., 2015*).

Transfer learning (TL) is regarded as a machine learning technique where information obtained from one task is applied to a specific task to enhance learning. TL performs efficiently in distinct feature areas or information patterns when huge data samples are available for training. Inferential TL, unsupervised TL, and supervised TL are three major categories for TL with different subsections. TL techniques using a pre-trained model and convolutional network are the most common strategies in object detection and classification. A pre-trained model is developed using huge benchmark datasets such as ImageNet, comprising enhanced feature attributes at all levels of abstraction. Simply by incorporating new convolutional or dense layers in a pre-trained model as a feature extractor and then fine-tuning the new model *via* continuous backpropagation, these feature representations may be partially or wholly reused in subsequent tasks. Numerous applications have demonstrated that this method may be used to create cutting-edge object detectors (*Yu, Xiu & Li, 2022; Ribani & Marengoni, 2019; Ebrahim, Al-Ayyoub & Alsmirat, 2019*).

Machine learning is a field that is raised out of artificial intelligence (AI). It involves using statistical models to automatically learn patterns from data without defining the patterns explicitly. Machine learning is defined as “the automated detection of meaningful patterns in data”, meaning programs learn to perform simple and complex tasks after gaining more experience from past information. There are a lot of practical applications of machine learning ranging from computer vision applications (such as face detection, image retrieval, *etc.*), speech processing (such as speech synthesis, language modeling, *etc.*), natural language processing (NLP) (such as context-free parsing, named-entity recognition, *etc.*) and other problems including fraud detection, *etc* (*Shalev-Shwartz & Ben-David, 2014; Murdoch et al., 2019*).

Another renowned neural network is the RNN, which uses sequential or time-series data and feeds the results of the prior stage as input to the current stage. Recurrent networks, such as feedforward and CNN, learn from training input, but they stand out due to their “memory”, which enables them to influence current input and output by drawing on data from earlier inputs. The output of an RNN depends on previous components in the sequence, in contrast to a typical DNN, which presumes that inputs and outputs are independent of one another. This feature is essential to a variety of applications since the embedded structure in the data sequence provides valuable information (*Dupond, 2019*).

Deep learning is another field of machine learning in artificial intelligence concerned with algorithms capable of learning from data. The term deep learning simply means training neural networks or deep neural networks to learn a general representation of data. Furthermore, deep learning aims to estimate a function, f^* . For example, all a deep learning classifier $y = f(x)$ does is to map an input x to its corresponding category y . Given a deep learning network, there is a defined mapping $y = f(x; \theta)$ and appropriate value of parameters that needs to be learned in order to get the best estimation function. In, their deep learning model outperforms human performance, which shows the promising future of deep learning algorithms for computer vision tasks. Deep learning usually consists of

several hidden layers, while shallow neural networks are made up of fewer hidden layers. There have been many variant forms of deep learning architectures used to solve a lot of diverse problems in fields such as computer vision, natural language processing and speech recognition. Some of these architectures are deep CNN, RNN and deep belief networks. Recently, with the recent availability of large amount of data and computational power, deep learning gained more attention ([He et al., 2016](#); [Szegedy et al., 2015](#)).

FACIAL EXPRESSION RECOGNITION, INTERNET OF THINGS, AND CLOUD COMPUTING

In recent years, the IoT technology has been integrated into various application domains, including transportation, video surveillance, healthcare, education, and many more, to enable object/device connectivity and data sharing. A smart city comprises different components such as cloud servers, wireless networks, smart healthcare, smart shopping, smart traffic systems, and smart homes. Smart devices and IoTs capture different signals, including facial expressions, and send them to the cloud server for processing and decision-making. The decisions are then forwarded to the stakeholders for necessary actions ([Masud et al., 2020](#); [Yassine et al., 2019](#)).

Cloud computing (CC) is an emerging technology with ubiquitous characteristics such as seamless accessibility, high scalability, and online storage, ultimately reducing workforce and capital costs. It has attracted organizations to run their financial activities and businesses over the cloud. However, despite its strengths, cloud computing has some drawbacks in terms of security concerns ([Tahirkheli et al., 2021](#)). Facial expression recognition technology, when combined with the IoT and cloud computing, has been integrated into autonomous driving systems for monitoring drivers' emotional states. Intelligent and automated facial expression recognition devices are more efficient in detecting the emotional states of the driver, enabling real-time safety measures and personalized driving experiences ([Chen et al., 2021](#)).

Integrating edge computing into IoT devices can make facial expression recognition algorithms more convenient and faster, improving sustainability and decreasing costs. The development of action unit (AU) based expression recognition algorithms for edge devices has enhanced processing time by eliminating the intermediate transfer process ([Yang et al., 2021](#)).

Integrating physical equipment into facial expression recognition systems can improve efficiency, productivity, and services. For example, numerous IoT-enabled vehicle recognition systems have been developed in which UHF-RFID windshield tags are integrated for efficient recognition ([Mufti et al., 2023](#)). These systems can be extended to incorporate facial expression recognition, enabling emotion-aware intelligent transportation systems. Wildlife conflict (HWC) is one of the critical global conflicts. A deep convolutional neural network (DCNN) based recognition system has been developed, which captures images of animals, saves them in .jpg format, and sends the images to the base station *via* a radio frequency (RF) network. It classifies and extracts features from the images, and whenever an animal is identified, it sends a message alert. In such systems,

IoT play a significant role and can be implemented by forest officials for the safety of people and wildlife ([Surya & Selvaperumal, 2022](#)). Similar systems can be adapted for facial expression recognition to detect and respond to human emotions in wildlife conservation efforts.

With the COVID-19 pandemic, traditional education has abruptly shifted towards online learning systems, presenting significant challenges. In face-to-face interactions, teachers can easily determine whether students are learning properly. The emerging facial expression recognition systems and diverse IoT devices, such as vision sensors and cameras, create an efficient student-teacher online interaction. Online learning can be improved by applying FER Enabled Education Aids IoT System, which can detect students' emotional states and engagement levels, enabling personalized learning experiences ([Geng, Meng & Dou, 2022](#)).

The advent of facial expression recognition technology and the IoT technology is making human life smarter and more efficient. For example, the door of an apartment can recognize the owner's facial expressions and open automatically through smartphone activation, enhancing security and convenience. Additionally, this technology can help in identifying lost people and criminals by analyzing their facial expressions ([Medapati, Murthy & Sridhar, 2019](#)). The smart city is illustrated in [Fig. 4](#).

Open challenges

With the tremendous contributions of face recognition, there are still challenges faced by real-world face recognition-based systems that need to be addressed. Most challenges are faced in an uncontrolled environment, where subjects are dynamic, and variations are difficult to catch in machine perspectives due to certain causes. Different challenges to face recognition have been discussed below in detail.

Lab-made compound facial expressions

In some situations, Ekman's basic emotions cannot represent human emotions. For example, at a surprise birthday party, the individual simultaneously feels a mixture of happy and surprised emotions. [Du & Martinez \(2015\)](#) proposes new compound emotions classes, combining basic emotions such as angrily and happily surprised. [Du, Tao & Martinez \(2014\)](#) images of different compound emotions were collected by recruiting diverse volunteers, and the images were taken by training the volunteers about the specific scenario in a controlled manner to act accordingly for each compound emotion. This approach is called 'lab-made facial expressions.' The images collected are captured based on the individual capability and simulated according to the specific direction. It is a challenge as such expressions are rare in real physical environments. The current compound facial expression image datasets may not express internal human emotion, so the compound emotions images can be collected from a real environment and should represent an individual's real compound emotions and feelings ([Du & Martinez, 2015](#); [Du, Tao & Martinez, 2014](#); [Win, Siritanawan & Kotani, 2023](#); [Parent, 2005](#)).

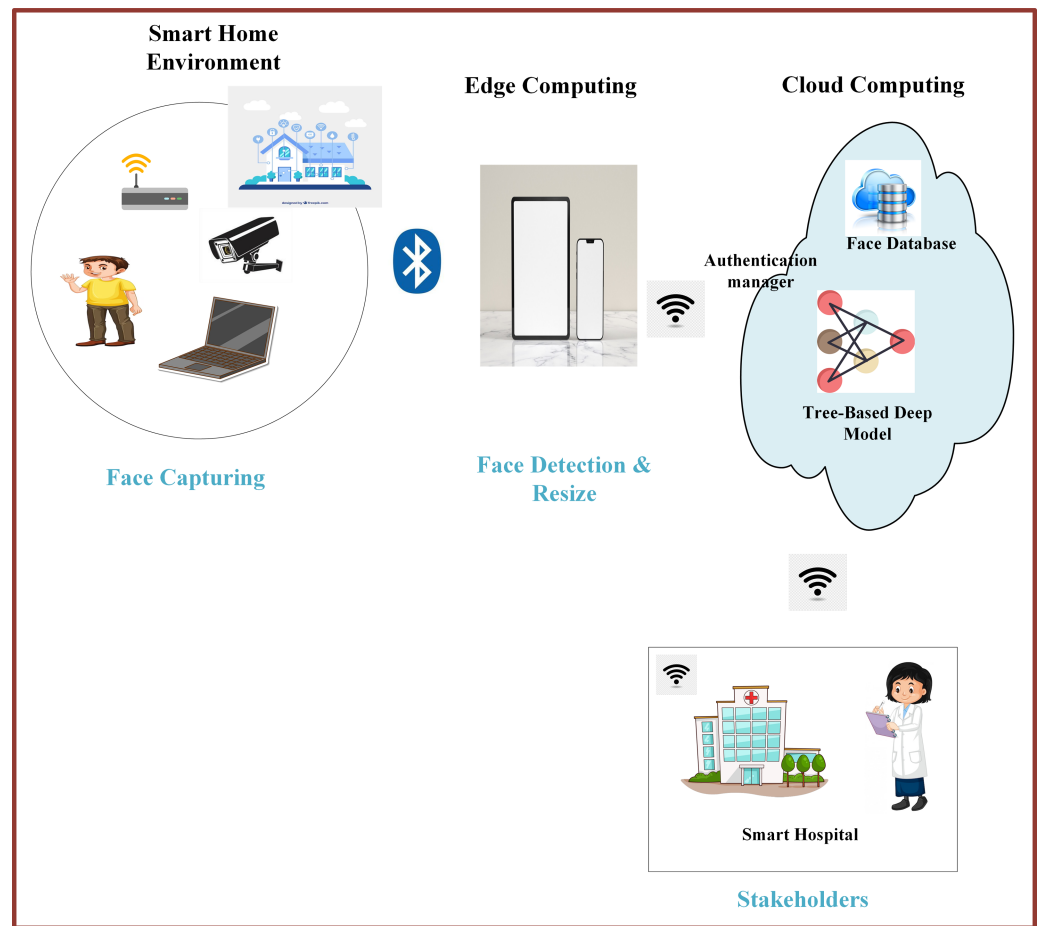


Figure 4 Smart city framework. Image source credits: Smart home, Image by Freepik; Boy in yellow shirt, Image by brgfx on Freepik (<https://www.freepik.com/>). CCTV, <https://www.rawpixel.com/image/6755502/png-sticker-public-domain>, CC0; Bluetooth, Hospital, <https://clipart-library.com/clipart/36700.htm>, Non commercial use only; Devices on marble table, Image by Freepik; A sticker template with a laptop isolated, Image by brgfx on Freepik; Wifi Icon Vector, Image by iyikon on vecteezy.com; Wifi icon, Image by seetwo on vecteezy.com; Cloud computing concept, Image by macrovector on Freepik; Learning Deep Algorithm Data, Image by flatart on vecteezy.com; Girl in science gown writing note on white, Image by brgfx on Freepik.

Full-size DOI: 10.7717/peerjcs.2024/fig-4

Pose variation

The pose variation means the rotation of the subject image in diverse 3D or 2D perspectives. It is challenging when a suspect, terrorist, or thief is in the airport or public crowd. Identifying suspects in public crowds with different poses is difficult, especially for law enforcement agencies. In a stored passport database, a single photo cannot correctly identify a suspect in a crowd with different poses. So, pose variation issues need to be addressed by researchers in the future (*Zhang, Zhang & Huang, 2013*).

Illumination variation

Illumination variation is another serious challenge for face recognition systems in real-world face recognition-based systems. The illumination and environmental conditions

are not under the control of machine intelligence. Tradition machines and deep learning models cannot address the illumination variation (*Zhao & Chellappa, 1999*).

Occlusion

Occlusion is a crucial challenge for computer vision scientists in real-world face recognition systems. In occlusion, the important features of the face are hidden due to the presence of other objects, such as a mask, scarf, hand, sunglasses, helmet, or hat. Tremendous research has been done on it, but it can still be enhanced (*Gautam & Mukhopadhyay, 2019; Guo et al., 2017; Shahid, 2023*).

Aging

Aging is the most challenging for face recognition systems as, with age, the features of the face are changing continuously, and it is difficult for intelligent machines to recognize faces with varying aging effects. It is a harder challenge than other face recognition challenges, and few efforts are available in the literature to address the aging issue of face recognition systems (*Fu, Guo & Huang, 2010*).

Lack of compound facial expression publicly available datasets

There are different challenges and limitations concerning recognizing compound expressions of emotion. The available databases for recognizing compound expressions of emotion could be more extensive. Furthermore, the publicly available datasets contain a very small number of categories, 23 (EmotionNet) and 22 (CFEE), which cover a small portion of conceivable compound emotions. Furthermore, it requires a large amount of training data, especially in quality and quantity training deep neural networks. Currently, the datasets utilized are of a limited number of categories and unbalanced data distributions (*Benitez-Quiroz, Srinivasan & Martinez, 2016; Benitez-Quiroz et al., 2017; Li & Li, 2020*).

RESEARCH METHODOLOGIES ISSUES

The running techniques concerning complex emotion profiling have limitations concerning detection and need a large dataset for the emotion profiling model. Other findings show that compound emotions have yet to be analyzed deeply. Applying deep learning methods requires a large and balanced training dataset for optimal performance. For the recognition of compound expressions of emotion, diverse traditional machine learning approaches are utilized, like SVM and kNN, others, but the utilized traditional methods have different limitations; for the manual tuning of parameters, there is need of more labor work. Another area for improvement is the incapability of detecting images, which is performed spontaneously and uncontrolled. It is argued that most of the recognition of compound expressions of emotion has been done through databases and is based on algorithms and unbalanced classifications, which ultimately lead to a lack of precision. Compound emotion labeling is performed automatically through machine learning-based algorithms, which leads to inaccuracies. The main cons of using the kNN algorithm is the utilization of whole features equally in the field of computing for similarities, which ultimately leads to classification errors (*Jarraya, Masmoudi & Hammami, 2020; Thuseethan, Rajasegarar & Yearwood, 2020; Guo et al., 2018*).

Recognition of limited compound emotions

The compound expressions of emotions are not explored due to their complexities. Compound emotions are mainly created by combining basic emotions. In some methods and databases, it can only recognize a limited number of compound expressions of emotion. Currently, the datasets utilized are of a limited number of categories and unbalanced data distributions ([Liliana, Basaruddin & Widyanto, 2017](#); [Liliana, Basaruddin & Oriza, 2018](#)).

Computational time issue

The recognition of compound facial expressions of emotion is a critical job in large databases. For the recognition of specific image in large databases, diverse algorithms have been implemented. In addition, it takes more computational time and space for dealing with large databases ([Taskiran, Kahraman & Erdem, 2020](#)).

FER in video surveillance

FER-based recognition systems are becoming common and have varied applications, especially video surveillance. In this setting, the performance of FER systems is mostly reliant on image acquisition conditions, mainly when the posture changes, and because the acquisition techniques may include artifacts. We mainly discuss camera focus problems that can lead to image blurring, low-resolution, or compression-related errors and block effects. In this case, the challenge of FER systems is to distinguish individuals from photographs captured employing video surveillance cameras, presenting blurred, low-resolution, block artifacts or faces with variable poses. This challenge remains an unsolved problem and requires further research ([Zhang, An & Zhang, 2018](#)).

Facial expression recognition and smartphones

Adopting FER on mobile devices offers many advantages. In addition to the employment facility, the users do not have to remember the PIN or password; it can be conveniently implemented on tablets and smartphones because only the frontal camera is required. FER systems have been used in recent years to secure devices and control access to many different services through smartphones, such as purchases or online payments on the store. While adopting FER systems on smartphones provides many advantages, many challenges must be addressed. The user's facial image should be captured in a comfortable or constrained environment. Many factors, such as pose and ambient lighting due to various ways of interacting with mobile technology and imaging distance, can restrict facial image quality ([Adjabi et al., 2020](#)).

Cross-cultural variability

Emotional expressions vary considerably from culture to culture, region, and language. It is very challenging for researchers to develop robust models to overcome the cross-cultural variability in expressions of emotion. New robust models need to be developed to integrate and consider cross-cultural variability. Furthermore, it is more challenging in the case of compound facial expressions of emotions ([Vashishth et al., 2023](#)).

Multimodal challenges

Multimodal machine learning-based models are also utilized for emotion recognition, integrating diverse modalities such as facial expressions, speech, text, and others. Due to its complexity due to different modalities, It is challenging for the model to extract relevant important features from each modality appropriately ([Liang & Morency, 2023](#)).

Low resolution

Research shows that facial recognition models have achieved an accuracy of 99.63% on high-resolution quality images, but low-resolution image datasets for automatic facial expression recognition are a challenge. Despite remarkable research on developing diverse models for high-resolution images, little attention is given to low-resolution images. The low-resolution issue mostly occurred to models in real-life environments, as the image captured in a real environment will have the issue of low resolution, illumination, and pose variation ([Lo et al., 2023](#); [Cheng, Zhu & Gong, 2019](#); [Massoli, Amato & Falchi, 2020](#)).

Ethnic differences

In the 21st century, numerous technologies have been utilized to deal with emotion recognition with key breakthroughs. The facial expressions of emotion recognition systems face tremendous issues, but ethnic differences are one of the critical hurdles. Charles Darwin also mentioned that emotional expressions are invariant and can be changed from one ethnic group to another for the same facial expression of emotion. It is a challenge for diverse experimental, methodological, and theoretical grounds ([Ahmad et al., 2022](#); [Costantini et al., 2022](#)).

Future directions

With the innovations in computer vision, numerous advanced deep learning models have been developed to recognize facial expressions of emotions. However, researchers face diverse challenges concerning emotion recognition. While computer vision scientists have overcome various challenges, new challenges continuously arise with new environmental situations and the installation of recognition systems in complex environments. In this section, we discuss different challenges that need to be addressed in future research and explore their implications for the field.

New models for compound emotions recognition

Compound emotions, also called mixed emotions, blend emotions with complex features compared to basic emotions. Accurately recognizing each class of compound emotions is challenging for existing deep learning models due to the complexity in feature extraction and the difficulty in differentiating between complementary and dominating emotions. Future research should focus on developing novel dedicated models specifically for compound facial expressions of emotions. Collaboration among computer scientists, psychologists, neuroscientists, linguists, and social scientists can lead to more comprehensive and accurate emotion models. Addressing this challenge will significantly advance the field of emotion recognition, enabling more nuanced and context-aware applications in various domains, such as healthcare, education, and human–computer interaction ([Wagner, 2019](#); [Zhang et al., 2020](#)).

New cross cultural adoptive models

Facial expressions vary across genders, cultures, and ages. An ideal database and cross-culture adaptive model is necessary to account for these differences. Future research should focus on developing a new cross-culture-adaptive model that efficiently recognizes basic and compound emotions while considering cross-cultural differences. This will require the creation of datasets that include all basic and compound facial expressions of emotions with appropriate attribute labels, ethnicity, gender, and age. Addressing this challenge will enable the development of more inclusive and equitable emotion recognition systems, with applications in fields such as marketing, advertising, and social robotics ([Jack et al., 2012](#)).

Multi-pose facial recognition model

Pose variation is a significant challenge for face recognition systems, particularly in crowded environments where individuals continuously change poses. Future research should focus on developing novel models that can effectively handle pose variations. Potential solutions include the use of 3D face modeling, pose-invariant feature extraction, and multi-view face recognition techniques. Overcoming this challenge will significantly improve the performance and reliability of face recognition systems in real-world scenarios, such as surveillance, security, and law enforcement ([Ding & Tao, 2016](#)).

Enhancing dataset for compound emotion recognition

To advance research on cross-cultural, cross-gender, and cross-age range facial expression recognition, there is a need for enhanced datasets that include high-resolution images, more compound emotion categories, and multi-pose and occlusion annotations. Creating such datasets through collaboration among multi-field experts, including computer scientists, psychologists, neuroscientists, linguists, and social scientists, can lead to more comprehensive and accurate datasets that are universally acceptable. Addressing this challenge will provide a solid foundation for the development of more robust and generalizable emotion recognition models ([Karnati et al., 2023](#)).

Efficient novel models for low and high resolution images

Most deep learning models perform efficiently on high-resolution images but struggle with low-resolution images, which are common in real-time recognition systems. Future research should focus on developing new models that can efficiently handle both low and high-resolution images, making them more adaptable and efficient for real-time recognition-based systems. Potential solutions include the use of super-resolution techniques, multi-scale feature extraction, and attention mechanisms. Overcoming this challenge will enable the deployment of emotion recognition systems in a wider range of applications and environments ([Lo et al., 2023](#)).

Novel models based on merging human brain emotion perception

Integrating advanced imaging techniques, such as functional Magnetic Resonance Imaging (fMRI), into emotion recognition models can provide insights into real-time brain changes during emotion perception. By linking brain neuronal activity to emotion recognition, future models can achieve better performance and incorporate additional

contextual behavioral cues. This interdisciplinary approach, combining neuroscience and computer vision, has the potential to revolutionize emotion recognition and deepen our understanding of human emotions. Addressing this challenge will require close collaboration between researchers in both fields and the development of novel algorithms that can effectively integrate fMRI data into emotion recognition models ([Tipirneni & Leal, 2022](#)).

Advance context-based models

While significant research has been conducted on facial expression recognition, current models often struggle to accurately trace the real feelings of users in context. Future research should focus on developing emotion recognition models that can recognize an individual's real feelings in relation to the context. This will require the development of advanced artificial intelligence-based systems that can integrate multiple modalities and optimize architectures for real-time systems. Addressing this challenge will enable the creation of more empathetic and context-aware emotion recognition systems, with applications in fields such as mental health, customer service, and social robotics ([Cîrneanu, Popescu & Iordache, 2023](#)).

Gait based intelligent classification model

Gait-based emotion recognition, which interprets and analyzes emotions based on walking patterns, is an emerging area of research. Deep learning-based models have yet to be extensively utilized in this domain, presenting an opportunity for future research. The fusion of different deep neural networks for gait-based emotion recognition could lead to significant breakthroughs. Addressing this challenge will expand the scope of emotion recognition beyond facial expressions, enabling the development of more comprehensive and multi-modal emotion recognition systems ([Xu et al., 2022](#)).

Suppression of noise labels

Databases of wild conditions often contain noisy and subjectively labeled data due to occlusion and illumination issues, leading to lower performance of facial expression recognition models. Future research should focus on developing techniques for noisy label suppression to overcome this challenge ([Shao & Luo, 2021](#)). Potential solutions include the use of unsupervised or semi-supervised learning, data augmentation, and robust loss functions. Addressing this challenge will improve the reliability and generalizability of emotion recognition models in real-world scenarios.

CONCLUSIONS

Significant research has been conducted on FER based systems. It is utilized in diverse fields, and computer vision scientists constantly encounter new advancements and challenges. In this article, we have discussed the basics of FER-based systems, architectural elements, FER applications, FER-based global leading companies, Interconnection between FER, IoT, and cloud computing, Open challenges to facial expression recognition technologies, and future directions. Overcoming the identified challenges and future directions in this

research through researchers in the future will bring evolution in the field of computer vision.

ACKNOWLEDGEMENTS

We express our gratitude to team members for their discussions and extend appreciation to anonymous reviewers for their valuable comments.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research is supported by the National Key Research and Development Program of China with ID 2018AAA0103203 and the Chengdu Technology Project with ID 2022-YF05-02014-SN, and Science and Technology Plan Project of Luzhou City, Sichuan Province (No. 2022-XDY-192). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

National Key Research and Development Program of China: 2018AAA0103203.

Chengdu Technology Project: 2022-YF05-02014-SN.

Science and Technology Plan Project of Luzhou City, Sichuan Province: 2022-XDY-192.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Sana Ullah conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Jie Ou performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Yuanlun Xie performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Wenhong Tian performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

This is a literature review.

REFERENCES

Adjabi I, Ouahabi A, Benzaoui A, Taleb-Ahmed A. 2020. Past, present, and future of face recognition: a review. *Electronics* 9(8):1188 DOI [10.3390/electronics9081188](https://doi.org/10.3390/electronics9081188).

- Ahaidous K, Tabaa M, Hachimi H. 2023.** Towards IoT-big data architecture for future education. *Procedia Computer Science* **220**:348–355 DOI [10.1016/j.procs.2023.03.045](https://doi.org/10.1016/j.procs.2023.03.045).
- Ahmad K, Wang S, Vogel C, Jain P, O'Neill O, Sufi BH. 2022.** Comparing the performance of facial emotion recognition systems on real-life videos: gender, ethnicity and age. In: *Proceedings of the future technologies conference (FTC) 2021, vol. 1*. Cham: Springer International Publishing, 193–210.
- Ahmad Z, Khan N. 2022.** A survey on physiological signal-based emotion recognition. *Bioengineering* **9**(11):688 DOI [10.3390/bioengineering9110688](https://doi.org/10.3390/bioengineering9110688).
- Ali W, Tian W, Din SU, Iradukunda D, Khan AA. 2021.** Classical and modern face recognition approaches: a complete review. *Multimedia Tools and Applications* **80**:4825–4880 DOI [10.1007/s11042-020-09850-1](https://doi.org/10.1007/s11042-020-09850-1).
- Andrejevic M, Selwyn N. 2022.** *Facial recognition*. Cambridge UK: Polity Press, 217.
- Ansari L, Ji S, Chen Q, Cambria E. 2022.** Ensemble hybrid learning methods for automated depression detection. *IEEE Transactions on Computational Social Systems* **10**(1):211–219.
- Appasaheb Borgalli R, Surve S. 2023.** Learning framework for compound facial emotion recognition. *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)* **16**(6):664–676.
- Bagchi T, Mahapatra A, Yadav D, Mishra D, Pandey A, Chandrasekhar P, Kumar A. 2022.** Intelligent security system based on face recognition and IoT. *Materials Today: Proceedings* **62**:2133–2137.
- Barata C, Ruela M, Francisco M, Mendonça T, Marques JS. 2014.** Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Systems* **8**(3):965–979 DOI [10.1109/JSYST.2013.2271540](https://doi.org/10.1109/JSYST.2013.2271540).
- Barsoum E, Zhang C, Ferrer CC, Zhang Z. 2016.** Training deep networks for facial expression recognition with crowd-sourced label distribution. In: *Proceedings of the 18th ACM international conference on multimodal interaction*. New York: ACM, 279–283.
- Bartlett MS, Movellan JR, Sejnowski TJ. 2002.** Face recognition by independent component analysis. *IEEE Transactions on Neural Networks and Learning Systems* **13**(6):1450–1464 DOI [10.1109/TNN.2002.804287](https://doi.org/10.1109/TNN.2002.804287).
- Benitez-Quiroz CF, Srinivasan R, Martinez AM. 2016.** Emotionet: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In: *2016 IEEE conference on Computer Vision and Pattern Recognition (CVPR)*. 5562–5570.
- Benitez-Quiroz CF, Srinivasan R, Feng Q, Wang Y, Martinez AM. 2017.** Emotionet challenge: recognition of facial expressions of emotion in the wild. *ArXiv* [arXiv:1703.01210](https://arxiv.org/abs/1703.01210).
- Blum A, Luedtke H, Ellwanger U, Schwabe R, Rassner G, Garbe C. 2004.** Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. *British Journal of Dermatology* **151**(5):1029–1038 DOI [10.1111/j.1365-2133.2004.06210.x](https://doi.org/10.1111/j.1365-2133.2004.06210.x).

- Brooks JA, Tzirakis P, Baird A, Kim L, Opara M, Fang X, Keltner D, Monroy M, Corona R, Metrick J. 2023.** Deep learning reveals what vocal bursts express in different cultures. *Nature Human Behaviour* 7(2):240–250.
- Cabanac M. 2002.** What is emotion? *Behavioural Processes* 60(2):69–83
DOI 10.1016/S0376-6357(02)00078-5.
- Chen Z, Feng X, Zhang S. 2022.** Emotion detection and face recognition of drivers in autonomous vehicles in IoT platform. *Image and Vision Computing* 128:104569
DOI 10.1016/j.imavis.2022.104569.
- Chen L, Li S, Bai Q, Yang J, Jiang S, Miao Y. 2021.** Review of image classification algorithms based on convolutional neural networks. *Remote Sensing* 13(22):4712
DOI 10.3390/rs13224712.
- Cheng Z, Zhu X, Gong S. 2019.** Low-resolution face recognition. In: *Computer vision—ACCV 2018: 14th Asian conference on computer vision, Perth, Australia, December. 2018. 2–6, Revised Selected Papers, Part III* 14. Cham: Springer International Publishing, 605–621.
- Cherbonnier A, Michinov N. 2021.** The recognition of emotions beyond facial expressions: comparing emoticons specifically designed to convey basic emotions with other modes of expression. *Computers in Human Behavior* 118:106689
DOI 10.1016/j.chb.2021.106689.
- Chitty-Venkata KT, Emani M, Vishwanath V, Somani AK. 2022.** Neural architecture search for transformers: a survey. *IEEE Access* 10:108374–108412
DOI 10.1109/ACCESS.2022.3212767.
- Cîrneanu AL, Popescu D, Iordache D. 2023.** New trends in emotion recognition using image analysis by neural networks. A systematic review. *Sensors* 23(16):7092
DOI 10.3390/s23167092.
- Costantini G, Parada-Cabaleiro E, Casali D, Cesarini V. 2022.** The emotion probe: on the universality of cross-linguistic and cross-gender speech emotion recognition via machine learning. *Sensors* 22(7):2461 DOI 10.3390/s22072461.
- Dargan S, Kumar M, Ayyagari MR, Kumar G. 2020.** A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering* 27:1071–1092 DOI 10.1007/s11831-019-09344-w.
- Ding C, Tao D. 2016.** A comprehensive survey on pose-invariant face recognition. *ACM Transactions on Intelligent Systems and Technology (TIST)* 7(3):1–42.
- Diniz DN, Rezende MT, Bianchi AGC, Carneiro CM, Luz EJS, Moreira GJP, Ushizima DM, de Medeiros FNS, Souza MJF. 2021.** A deep learning ensemble method to assist cytopathologists in pap test image classification. *Journal of Imaging* 7(7):111
DOI 10.3390/jimaging7070111.
- Du S, Martinez AM. 2015.** Compound facial expressions of emotion: from basic research to clinical applications. *Dialogues in Clinical Neuroscience* 17(4):443–455
DOI 10.31887/DCNS.2015.17.4/sdu.
- Du S, Tao Y, Martinez AM. 2014.** Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America* 111(15):E1454–E1462.

- Dupond S. 2019. A thorough review on the current advance of neural network structures. *Annual Reviews in Control* 14(14):200–230.
- Ebrahim M, Al-Ayyoub M, Alsmirat MA. 2019. Will transfer learning enhance imagenet classification accuracy using imagenet-pretrained models? In: *2019 10th international conference on information and communication systems (ICICS)*. Piscataway: IEEE, 211–216.
- Ekman P. 1992. An argument for basic emotions. *Cognition & Emotion* 6(3–4):169–200 DOI 10.1080/02699939208411068.
- Ekman P, Sorenson ER, Friesen WV. 1969. Pan-cultural elements in facial displays of emotion. *Science* 164:86–88 DOI 10.1126/science.164.3875.86.
- Ekundayo OS, Viriri S. 2021. Facial expression recognition: a review of trends and techniques. *IEEE Access* 9:136944–136973 DOI 10.1109/ACCESS.2021.3113464.
- Etemad K, Chellappa R. 1997. Discriminant analysis for recognition of human face images. In: *International conference on audio- and video-based biometric person authentication*. 125–142.
- Fu Y, Guo G, Huang TS. 2010. Age synthesis and estimation via faces: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(11):1955–1976 DOI 10.1109/TPAMI.2010.36.
- Gautam G, Mukhopadhyay S. 2019. An adaptive localization of pupil degraded by eyelash occlusion and poor contrast. *Multimedia Tools and Applications* 78(6):6655–6677 DOI 10.1007/s11042-018-6371-0.
- Geng Y, Meng H, Dou J. 2022. FEAIS: facial emotion recognition enabled education aids IoT system for online learning. In: *2022 international conference on advanced learning technologies (ICALT)*. Piscataway: IEEE, 403–407.
- Gill SS, Xu M, Ottaviani C, Patros P, Bahsoon R, Shaghaghi A, Golec M, Stankovski V, Wu H, Abraham A, Singh M, Mehta H, Ghosh SK, Baker T, Parlikad AK, Lutfiyya H, Kanhere SS, Sakellariou R, Dustdar S, Rana O, Uhlig S. 2022. AI for next generation computing: emerging trends and future directions. *Internet of Things* 19:100514 DOI 10.1016/j.iot.2022.100514.
- Google Trends. 2024. Internet of things, deep learning, facial expression, facial recognition systems, machine learning. Available at <https://trends.google.com/trends/> (accessed on 20 March 2024).
- Grm K, Struc V, Artiges A, Caron M, Ekenel HK. 2018. Strengths and weaknesses of deep learning models for face recognition against image degradations. *IET Biometrics* 7:81–89 DOI 10.1049/iet-bmt.2017.0083.
- Guo J, Lei Z, Wan J, Avots E, Hajarolasvadi N, Knyazev B, Kuharenko A, Silveira Jacques JC, Baró X, Demirel H, Escalera S, Allik J, Anbarjafari G. 2018. Dominant and complementary emotion recognition from still images of faces. *IEEE Access* 6:26391–26403 DOI 10.1109/ACCESS.2018.2831927.
- Guo G, Li S, Chan KL. 2000. Face recognition by support vector machines. In: *4Th IEEE international conference on automatic face and gesture recognition (FG 2000)*, Grenoble. Piscataway: IEEE, 196–201.

- Guo S, Tan G, Pan H, Chen L, Gao C. 2017.** Face alignment under occlusion based on local and global feature regression. *Multimedia Tools and Applications* 76(6):8677–8694 DOI 10.1007/s11042-016-3470-7.
- Gutta S, Wechsler H. 1997.** Face recognition using hybrid classifiers. *Pattern Recognition* 30(4):539–553 DOI 10.1016/S0031-3203(96)00111-2.
- Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y, Yang Z, Zhang Y, Tao D. 2022.** A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45(1):87–110.
- Haryanto H, Agastya W. 2022.** Synthesis of compound facial expressions based on indonesian sentences using multinomial Naïve Bayes model and dominance threshold equations. *Engineering Letters* 30(1).
- Hashemi VH, Gharahbagh AA. 2015.** Article: a novel hybrid method for face recognition based on 2d wavelet and singular value decomposition. *American Journal of Networks and Communications* 4(4):90–94 DOI 10.11648/j.ajnc.20150404.12.
- He K, Zhang X, Ren S, Sun J. 2016.** Deep residual learning for image recognition. In: *IEEE conference on computer vision and pattern recognition, CVPR 2016, Las Vegas, NV, USA, 27–30*. 770–778.
- Helgeson VS. 2020.** *Psychology of gender*. New York: Routledge.
- Hou C, Ai J, Lin Y, Guan C, Li J, Zhu W. 2022.** Evaluation of online teaching quality based on facial expression recognition. *Future Internet* 14(6):177 DOI 10.3390/fi14060177.
- Hu P, Ning H, Qiu T, Zhang Y, Luo X. 2017.** Fog computing based face identification and resolution scheme in internet of things. *IEEE Transactions on Industrial Informatics* 13(4):1910–1920 DOI 10.1109/TII.2016.2607178.
- Huang Y, Chen F, Lv S, Wang X. 2019.** Facial expression recognition: a survey. *Symmetry* 11(10):1189 DOI 10.3390/sym11101189.
- Huo H, Yu Y, Liu Z. 2023.** Facial expression recognition based on improved depth-wise separable convolutional network. *Multimedia Tools and Applications* 82(12):18635–18652 DOI 10.1007/s11042-022-14066-6.
- Ibrahim Khalil MW, Kamel MAE. 2024.** Towards smart sustainable cities vision and challenges. *International Journal of Nonlinear Analysis and Applications* 15(3):261–274.
- Jack RE, Garrod OG, Yu H, Caldara R, Schyns PG. 2012.** Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America* 109(19):7241–7244.
- Jarraya SK, Masmoudi M, Hammami M. 2020.** Compound emotion recognition of autistic children during meltdown crisis based on deep spatiotemporal analysis of facial geometric features. *IEEE Access* 8:69311–69326 DOI 10.1109/ACCESS.2020.2986654.
- Jiao L, Zhao J. 2019.** A survey on the new generation of deep learning in image processing. *IEEE Access* 7:172231–172263 DOI 10.1109/ACCESS.2019.2956508.
- Kanade T. 1973.** Picture processing system by computer complex and recognition of human faces. Ph.D. thesis, Kyoto University, Japan.

- Karnati M, Seal A, Bhattacharjee D, Yazidi A, Krejcar O. 2023.** Understanding deep learning techniques for recognition of human emotions using facial expressions: a comprehensive survey. *IEEE Transactions on Instrumentation and Measurement* 72:1–31.
- Kasar MM, Bhattacharyya D, Kim TH. 2016.** Face recognition using neural network: a review. *International Journal of Security and Its Applications* 10(3):81–100.
- Kitchenham B, Brereton OP, Budgen D, Turner M, Bailey J, Linkman S. 2009.** Systematic literature reviews in software engineering—a systematic literature review. *Information and Software Technology* 51(1):7–15 DOI 10.1016/j.infsof.2008.09.009.
- Ko BC. 2018.** A brief review of facial emotion recognition based on visual information. *Sensors* 18(2):401.
- Kosti R, Alvarez JM, Recasens A, Lapedriza A. 2019.** Context based emotion recognition using emotic dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42(11):2755–2766.
- Koutroumbas K, Theodoridis S. 2008.** *Pattern recognition*. Academic Press.
- Krizhevsky A, Sutskever I, Hinton GE. 2012.** Imagenet classification with deep convolutional neural networks. In: *Conf. on neural information processing systems, NIPS*. 1097–1105.
- Kunapuli G. 2023.** *Ensemble methods for machine learning*. New York: Simon and Schuster.
- Kuruvayil S, Palaniswamy S. 2022.** Emotion recognition from facial images with simultaneous occlusion, pose and illumination variations using meta-learning. *Journal of King Saud University-Computer and Information Sciences* 34(9):7271–7282 DOI 10.1016/j.jksuci.2021.06.012.
- Lavanya B, Hannah Inbarani H. 2018.** A novel hybrid approach based on principal component analysis and tolerance rough similarity for face identification. *Neural Computing and Applications* 29:289–299.
- Li S, Deng W. 2020.** Deep facial expression recognition: a survey. *IEEE Transactions on Affective Computing* 13(3):1195–1215.
- Li H, Li Q. 2020.** End-to-end training for compound expression recognition. *Sensors* 20(17):4727 DOI 10.3390/s20174727.
- Liang PP, Morency LP. 2023.** Tutorial on multimodal machine learning: principles, challenges, and open questions. In: *International conference on multimodal interaction*. 101–104.
- Liliana DY, Basaruddin C, Widyanto MR. 2017.** Mix emotion recognition from facial expression using SVM-CRF sequence classifier. In: *Proceedings of the international conference on algorithms, computing and systems*. 27–31.
- Liliana DY, Basaruddin T, Oriza IID. 2018.** The Indonesian Mixed Emotion Dataset (IMED) a facial expression dataset for mixed emotion recognition. In: *Proceedings of the 2018 international conference on artificial intelligence and virtual reality*, vol. 82. 56–60.

- Lindquist KA, Jackson JC, Leshin J, Satpute AB, Gendron M. 2022.** The cultural evolution of emotion. *Nature Reviews Psychology* **1**(11):669–681 DOI [10.1038/s44159-022-00105-4](https://doi.org/10.1038/s44159-022-00105-4).
- Lo L, Ruan BK, Shuai HH, Cheng WH. 2023.** Modeling uncertainty for low-resolution facial expression recognition. *IEEE Transactions on Affective Computing* **15**:198–209.
- Lu C, Ke L, Zhang Q, Deng X, Shang W, Zhao X, Li Y, Xie Y, Wang Z. 2024.** Quality of systematic reviews with meta-analyses of resveratrol: a methodological systematic review. *Phytotherapy Research* **38**(1):11–21 DOI [10.1002/ptr.8025](https://doi.org/10.1002/ptr.8025).
- Lu X, Wang Y, Jain AK. 2003.** Combining classifiers for face recognition. In: *Proceedings of the IEEE international conference on multimedia and expo, ICME, Baltimore*. Piscataway: IEEE, 13–16.
- Mahesh B. 2020.** Machine learning algorithms—a review. *International Journal of Science and Research (IJSR)* **9**(1):381–386.
- Martinez A, Du S. 2012.** A model of the perception of facial expressions of emotion by humans: research overview and perspectives. *Journal of Machine Learning Research* **13**(5).
- Massoli FV, Amato G, Falchi F. 2020.** Cross-resolution learning for face recognition. *Image and Vision Computing* **99**:103927 DOI [10.1016/j.imavis.2020.103927](https://doi.org/10.1016/j.imavis.2020.103927).
- Masud M, Muhammad G, Alhumyani H, Alshamrani SS, Cheikhrouhou O, Ibrahim S, Hossain MS. 2020.** Deep learning-based intelligent face recognition in IoT-cloud environment. *Computer Communications* **152**:215–222 DOI [10.1016/j.comcom.2020.01.050](https://doi.org/10.1016/j.comcom.2020.01.050).
- Mathew A, Amudha P, Sivakumari S. 2021.** Deep learning techniques: an overview. *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*:599–608.
- Mazzia V, Angarano S, Salvetti F, Angelini F, Chiaberge M. 2022.** Action transformer: a self-attention model for short-time pose-based human action recognition. *Pattern Recognition* **124**:108487 DOI [10.1016/j.patcog.2021.108487](https://doi.org/10.1016/j.patcog.2021.108487).
- Medapati PK, Murthy PHST, Sridhar KP. 2019.** LAMSTAR: for IoT-based face recognition system to manage the safety factor in smart cities. *Transactions on Emerging Telecommunications Technologies* **2019**:1–15.
- Mohammadi SO, Kalhor A. 2021.** Smart fashion: a review of AI applications in the Fashion & Apparel Industry. ArXiv [arXiv:2111.00905](https://arxiv.org/abs/2111.00905).
- Mohapatra J, Chen PY, Liu S, Daniel L. 2019.** Towards verifying robustness of neural networks against semantic perturbations. ArXiv [arXiv:1912.09533](https://arxiv.org/abs/1912.09533).
- Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. 2019.** Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences of the United States of America* **116**(44):22071–22080.
- Mufti N, Ullah S, Sharif A, Nawaz MW, Alkhayyat A, Imran MA, Abbasi QH. 2023.** IoT enabled vehicle recognition system using inkjet-printed windshield tag and 5G cloud network. *Internet of Things* **23**:100873 DOI [10.1016/j.iot.2023.100873](https://doi.org/10.1016/j.iot.2023.100873).

- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International Journal of Surgery* 88:105906 DOI 10.1016/j.ijssu.2021.105906.
- Pala G, Erdem CE. 2019. Performance comparison of deep learning based face identification methods for video under adverse conditions. In: *2019 15th international conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. Piscataway: IEEE, 90–97.
- Parent A. 2005. Duchenne de boulogne: a pioneer in neurology and medical photography. *Canadian Journal of Neurological Sciences* 32(3):369–377 DOI 10.1017/S0317167100004315.
- Pendhari H, Nagdeote S, Rathod S, Khan L, Vishwakarma S. 2022. Compound emotions: a mixed emotion detection. In: *Proceedings of the international conference on Innovative Computing & Communication (ICICC)*.
- Rao MS, Reddy BE. 2011. Comparative analysis of pattern recognition methods: an overview. *Indian Journal of Computer Science and Engineering (IJCSE)* 2(3):385–390.
- Rastgoo M, Morel O, Marzani F, Garcia R. 2015. Ensemble approach for differentiation of malignant melanoma. In: *Twelfth international conference on quality control by artificial vision 2015 (Vol. 9534)*. SPIE, 304–312.
- Ribani R, Marengoni M. 2019. A survey of transfer learning for convolutional neural networks. In: *2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)*. Piscataway: IEEE, 47–57.
- Rincy TN, Gupta R. 2020. Ensemble learning techniques and its efficiency in machine learning: a survey. In: *2nd international conference on data, engineering and applications (IDEA)*. IEEE, 1–6.
- Roberts LG. 1963. Machine perception of three-dimensional solids. Ph.D. thesis, Massachusetts Institute of Technology.
- Saganowski S. 2022. Bringing emotion recognition out of the lab into real life: recent advances in sensors and machine learning. *Electronics* 11(3):496 DOI 10.3390/electronics11030496.
- Sajjad M, Ullah FUM, Ullah M, Christodoulou G, Cheikh FA, Hijji M, Muhammad K, Rodrigues JJ. 2023. A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines. *Alexandria Engineering Journal* 68:817–840 DOI 10.1016/j.aej.2023.01.017.
- Shahid AR. 2023. A survey on facial expression recognition: modality, methodologies, challenges and emerging topics. Authorea Preprints.
- Shalev-Shwartz S, Ben-David S. 2014. *Understanding machine learning: from theory to algorithms*. Cambridge: Cambridge University Press.
- Shao J, Luo Y. 2021. TAMNet: two attention modules-based network on facial expression recognition under uncertainty. *Journal of Electronic Imaging* 30(3):033021–033021.

- Singh D. 2012.** Human emotion recognition system. *International Journal of Image, Graphics and Signal Processing* **4**(8):50.
- Sirovich L, Kirby M. 1987.** Low-dimensional procedure for the characterization of human faces. *Josa A* **4**(3):519–524.
- Srivastava G, Bag S. 2023.** Modern-day marketing concepts based on face recognition and neuro-marketing: a review and future research directions. *Benchmarking: An International Journal* **31**(2):410–438.
- Suri G, Gross JJ. 2022.** What is an emotion? A connectionist perspective. *Emotion Review* **14**(2):99–110 DOI [10.1177/17540739221082203](https://doi.org/10.1177/17540739221082203).
- Surya T, Selvaperumal S. 2022.** The IoT-based real-time image processing for animal recognition and classification using deep convolutional neural network (DCNN). *Microprocessors and Microsystems* **95**:104693 DOI [10.1016/j.micpro.2022.104693](https://doi.org/10.1016/j.micpro.2022.104693).
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. 2015.** Going deeper with convolutions. In: *IEEE conference on computer vision and pattern recognition, CVPR 2015, Boston, MA, USA, 7–12*. 1–9.
- Tahirkheli AI, Shiraz M, Hayat B, Idrees M, Sajid A, Ullah R, Ayub N, Kim KI. 2021.** A survey on modern cloud computing security over smart city networks: threats, vulnerabilities, consequences, countermeasures, and challenges. *Electronics* **10**(15):1811 DOI [10.3390/electronics10151811](https://doi.org/10.3390/electronics10151811).
- Talaat FM. 2023.** Real-time facial emotion recognition system among children with autism based on deep learning and IoT. *Neural Computing and Applications* **35**(17):12717–12728 DOI [10.1007/s00521-023-08372-9](https://doi.org/10.1007/s00521-023-08372-9).
- Taskiran M, Kahraman N, Erdem CE. 2020.** Face recognition: past, present and future (a review). *Digital Signal Processing* **106**:102809.
- Thuseethan S, Rajasegarar S, Yearwood J. 2020.** Complex emotion profiling: an incremental active learning based approach with sparse annotations. *IEEE Access* **8**:147711–147727.
- Tipirneni S, Leal S. 2022.** Deciphering facial expressions: factors that affect emotion recognition. *Journal of Student Research* **11**(1):1–10.
- Tolba AS. 2000.** A parameter-based combined classifier for invariant face recognition. *Cybernetics and Systems* **31**(8):837–849.
- Tunstall L, Werra LVon, Wolf T. 2022.** *Natural language processing with transformers*. Sebastapol: O'Reilly Media, Inc.
- Turk M, Pentland A. 1991a.** Eigenfaces for face recognition. *Journal of Cognitive Neuroscience* **3**:71–86.
- Turk MA, Pentland AP. 1991b.** Face recognition using eigenfaces. In: *IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE, 586–591.
- Van Thanh N. 2023.** Emotion recognition systems in retail: a detailed analysis of their role in enhancing customer interactions, driving sales, and predicting trends. *Journal of Computational Social Dynamics* **8**(3):1–9.

- Vashishth TK, Kumar B, Panwar R, Kumar S, Chaudhary S. 2023.** Exploring the role of computer vision in human emotion recognition: a systematic review and meta-analysis. In: *2023 second international conference on augmented intelligence and sustainable systems (ICAISS)*. Piscataway: IEEE, 1071–1077.
- Wagner W. 2019.** Recognizing compound expressions and emotion intensity. Ph.D. dissertation, Department of Psychology, University Alabama at Birmingham, Birmingham, AL, USA.
- Wang M, Deng W. 2019.** Deep face recognition: a survey. ArXiv [arXiv:1804.06655](https://arxiv.org/abs/1804.06655).
- Wang X, Xu W, Niu KH, Miao X. 2020.** Emotion recognition of students based on facial expressions in online education based on the perspective of computer simulation. *Complexity* 2020:1–9.
- Win SSK, Siritanawan P, Kotani K. 2023.** Compound facial expressions image generation for complex emotions. *Multimedia Tools and Applications* 82(8):11549–11588 DOI [10.1007/s11042-022-14289-7](https://doi.org/10.1007/s11042-022-14289-7).
- Wingenbach TS, Ashwin C, Brosnan M. 2018.** Sex differences in facial emotion recognition across varying expression intensity levels from videos. *PLOS ONE* 13(1):e0190634 DOI [10.1371/journal.pone.0190634](https://doi.org/10.1371/journal.pone.0190634).
- Wright J. 2023.** Suspect AI: vibraimage, emotion recognition technology and algorithmic opacity. *Science, Technology and Society* 28(3):468–487 DOI [10.1177/09717218211003411](https://doi.org/10.1177/09717218211003411).
- Xiao H, Li W, Zeng G, Wu Y, Xue J, Zhang J, Li C, Guo G. 2022.** On-road driver emotion recognition using facial expression. *Applied Sciences* 12(2):807 DOI [10.3390/app12020807](https://doi.org/10.3390/app12020807).
- Xie Y, Tian W, Ma T. 2020.** A transfer learning approach to compound facial expression recognition. In: *Proceedings of the 4th international conference on advances in image processing, vol. 131*. 95–101.
- Xu S, Fang J, Hu X, Ngai E, Wang W, Guo Y, Leung VC. 2022.** Emotion recognition from gait analyses: current research and future directions. *IEEE Transactions on Computational Social Systems* 11(1):363–377.
- Xue-fang L, Tao P. 2013.** Realization of face recognition system based on Gabor wavelet and elastic bunch graph matching. In: *25th Chinese control and decision conference, CCDC*. 3384–3386.
- Yang J, Qian T, Zhang F, Khan SU. 2021.** Real-time facial expression recognition based on edge computing. *IEEE Access* 9:76178–76190 DOI [10.1109/ACCESS.2021.3082641](https://doi.org/10.1109/ACCESS.2021.3082641).
- Yassine A, Singh S, Hossain MS, Muhammad G. 2019.** IoT big data analytics for smart homes with fog and cloud computing. *Future Generation Computer Systems* 91:563–573 DOI [10.1016/j.future.2018.08.040](https://doi.org/10.1016/j.future.2018.08.040).
- Yu F, Xiu X, Li Y. 2022.** A survey on deep transfer learning and beyond. *Mathematics* 10(19):3619 DOI [10.3390/math10193619](https://doi.org/10.3390/math10193619).
- Zhang D, An P, Zhang H. 2018.** Application of robust face recognition in video surveillance systems. *Optoelectronics Letters* 14:152–155 DOI [10.1007/s11801-018-7199-6](https://doi.org/10.1007/s11801-018-7199-6).

- Zhang Z, Fort JM. 2023.** Facial expression recognition in virtual reality environments: challenges and opportunities. *Frontiers in Psychology* **14**:1280136 DOI [10.3389/fpsyg.2023.1280136](https://doi.org/10.3389/fpsyg.2023.1280136).
- Zhang Z, Yi M, Xu J, Zhang R, Shen J. 2020.** Two-stage recognition and beyond for compound facial emotion recognition. In: *2020 15th IEEE international conference on automatic face and gesture recognition (FG 2020)*. Piscataway: IEEE, 900–904.
- Zhang H, Zhang Y, Huang TS. 2013.** Pose-robust face recognition via sparse representation. *Pattern Recognition* **46**(5):1511–1521 DOI [10.1016/j.patcog.2012.10.025](https://doi.org/10.1016/j.patcog.2012.10.025).
- Zhao W, Chellappa R. 1999.** *Robust face recognition using symmetric shape-from shading*. Computer Vision Laboratory, Center for Automation Research, University of Maryland.
- Zhao W, Chellappa R, Phillips PJ, Rosenfeld A. 2003.** Face recognition: a literature survey. *ACM Computing Surveys (CSUR)* **35**(4):399–458 DOI [10.1145/954339.954342](https://doi.org/10.1145/954339.954342).