Multidisciplinary Reviews

BIBLIOMETRIC RESEARCH
Published Online: April 5, 2025
https://doi.org/10.31893/multirev.2025328

Real-time emotion recognition based on facial expressions using Artificial Intelligence techniques: A review and future directions

Cheng Qian[a] ✉ | **João Alexandre Lobo Marques**[b] | **Auzuir Ripardo de Alexandria**[c]

[a]Institute for Data Engineering and Science, University of Saint Joseph, Macao, China.
[b]Laboratory of Applied Neurosciences, University of Saint Joseph, Macao, China.
[c]Federal Institute of Education, Science and Technology of Ceará, Fortaleza/CE, Brazil.

Abstract In recent years, the real-time facial expression recognition system based on artificial intelligence technology has garnered significant attention from academia and industry. This paper presents a systematic literature review and bibliometric analysis to examine the latest publications in this field, summarizing the development and research significance of facial expression recognition technology and emphasizing its vital role in human-computer interaction and affective computing. The study used PRISMA to review 386 articles published from January 2019 to December 2023 in Web of Science, Scopus, IEEE Xplore, and ACM Digital Library. It encompasses covering various research methodologies, datasets, and application areas, as well as artificial intelligence technology, algorithms, and models. This review highlights advancements in Facial Expression Recognition, particularly the predominant use of databases such as FER2013 and CK+ while identifying Convolutional Neural Networks as the primary technique for real-time emotion classification. A quantitative analysis of research trends over the past five years indicates a shift toward keywords like transfer learning and applications in domains such as healthcare and the Internet of Things. Contemporary deep learning models, including CNNs, ResNet, and VGG, demonstrate impressive accuracy in classifying seven basic emotions, facilitating real-time applications across multiple fields. However, challenges such as overfitting, sensitivity to environmental factors, and the necessity for high-performance computing resources impede the broader deployment of these systems. These findings underscore the urgent need for further research to address these limitations and enhance the ethical application of FER technologies. Finally, based on the review and analysis results, this paper outlines future research directions for this technology, including multimodal information fusion, computational modelling, personalized emotion recognition, and interdisciplinary cooperation, thereby providing valuable references and inspiration for future works.

Keywords: emoton recognition, deep learning, machine learning, human-computer interaction, databases, applications

1. Introduction

The identification of emotions has become a prominent area of study within computer vision. Since Picard proposed the concept of affective computing in 1977 (Picard, 1997), it has gained increasing attention and development in the field of many researchers, thus providing computers with the ability to detect human emotions. For example, electroencephalography (EEG) is one of the methods used for emotion detection. By placing electrodes on the surface of the scalp, EEG records brain activity to investigate the brain's reactions to emotional and cognitive triggers. This technique provides a direct approach independent of facial expressions or auditory cues (Fernandes et al., 2024). In 1972, Ekman and Friesen (Russell, 1991) proposed that the rules governing expressions are a fundamental component in the creation of facial expressions related to emotions, with the understanding of these expressions differing across different cultures (Ekman, 1971). The human face is the most complex signalling system (Samal & Iyengar, 1992) and a highly differentiated part of the human body. Even identical twins have different faces in some ways. The uniqueness of the human face (Ko, 2018) is one of the essential reasons why an increasing number of researchers favor facial expression recognition technology. However, the interaction and emotion recognition between computers and humans are complex and diverse, and researchers still face difficulties and challenges in the face of real-time and accurate identification of human emotions (Lopes et al., 2017). Emotion recognition in facial expression analysis, which enables machines to understand human emotions, represents an essential application scenario, and several studies (Muhammad & Hossain, 2021; Kim et al., 2018; Metgud et al., 2022; Rastgoo et al., 2019) can prove it. As artificial intelligence technology has evolved and gained widespread acceptance, encompassing both machine learning and deep learning techniques, the construction of intelligent systems with real-time and accurate recognition of human emotions is gradually approaching reality.

 Received: February 9, 2025 | Accepted: March 17, 2025

Within the realm of affective computing, emotion detection has many applications in computer vision research (Awais et al., 2021; Filipovic et al., 2019; Kim et al., 2018; Werner et al., 2022), including medical areas (Dhope & Neelagar, 2022), the Internet of Things (IOT) (Liakopoulos et al., 2021; Zarif et al., 2021), the field of education (Hussain & Salim Abdallah Al Balushi, 2020), intelligent learning systems (Barron-Estrada et al., 2019; Chaiyarak et al., 2021), neuromarketing (Marques et al., 2024), soft robotics (Lin et al., 2021), virtual reality (Zhang et al., 2023), and recommendation systems (Thomas & S, 2022). A trusted facial emotion detection system with high recognition accuracy allows computers to recognize and understand human emotions and provide feedback in real time, which is conducive to empowering all walks of life and can be applied to different specific areas. The establishment of an emotion recognition system is based mainly on the combination of artificial intelligence technology and models and has been used in various fields; for example, in the field of education (Savchenko et al., 2022), teachers can obtain timely feedback and understand the engagement of students in class so that they can mobilize the enthusiasm of students; in the field of IOT, Raspberry Pi 4 and other development boards can be used for facial emotion recognition (K & Thripurala, 2023), which can be applied to mobile phone recognition of users' emotions or group emotion recognition; in the medical field, an intelligent system can understand the feelings of patients with mental illness immediately (Turcian & Stoicu-Tivadar, 2023); emotion analysis can identify depression or other similar diseases but may also be utilized in assessing and managing autistic children (Arabian et al., 2021).

Over the years, in the facial expression recognition literature, a few reviews have been published on facial expression recognition for different fields or specific facial emotion recognition steps, enabling researchers to create comprehensive and accurate pictures. For example, (Pinto et al., 2023) proposed a systematic review focused on convolution neural networks based on deep learning, focusing on the recognition of expressions and microexpressions; (Samadiani et al., 2019) mainly discussed the multimodal emotional expression recognition method, especially its application in the wild; (Alisawi & Yalçin, 2023) focused only on the implementation of deep learning techniques for their review; and (Canedo & Neves, 2019) summarized and reviewed the literature on facial expression recognition via computer vision technology, focusing on the most frequently employed techniques and algorithms for the main steps of this process.

During the past few years, scientific publications on real-time facial expression recognition have accelerated. Consequently, this systematic review differs from existing studies by providing a detailed summary of the methods and main findings of the review articles through bibliographic quantitative analysis. In addition, this systematic review includes examining journal articles, conference proceedings, etc. This study aims to provide a comprehensive summary and critical evaluation of real-time emotion recognition based on artificial intelligence technology, with the primary goal of identifying the progress in this research field, research gaps, trends, mainstream technologies, and avenues for future research that will provide a framework for subsequent projects.

1.1. Research questions

This review provides a comprehensive synthesis of the literature on real-time facial expression recognition literature based on artificial intelligence technology through a systematic review and bibliometric analysis. Four main research questions drive the review:

RQ1. Which databases and methods are most commonly used in emotion classification system studies?

RQ2. Which artificial intelligence techniques show the most promise for enhancing the accuracy and efficiency of facial expression recognition?

RQ3. What are the most commonly used keywords and research trends in facial expression recognition?

RQ4. How does the current facial expression classification system perform in real-world applications?

RQ5. What are the limitations of current real-time facial expression recognition system studies?

RQ6. Which emerging technologies possess the potential to significantly enhance performance in facial expression recognition tasks yet remain underutilized in contemporary research?

This study aims to provide a comprehensive framework of core concepts in the area of real-time facial expression analysis on the basis of a bibliometric and systematic literature review. An initial contribution of this study is the execution of a quantitative analysis to determine the variety and volume of publications within the domain of facial expression recognition (FER) over the last five years. This includes the identification of important keywords, research trends, the most commonly used FER databases, and the most highly cited articles. Additionally, as a second contribution, the frequency of different artificial intelligence techniques employed in the literature highlights works that achieve high accuracy and provides a comprehensive summary of FER applications in different areas. Finally, the work presents challenges, limitations, and key future directions in the FER domain, discussing unresolved issues and outlining potential avenues for advancement in this domain.

The structure of this paper is outlined as follows. In Section 2, the methodological approach for the reviews is presented. The third section analyses the selected articles and explains the results (including those of the systematic review and bibliometric analysis). Section four expounds on the theoretical and practical significance of the review, and finally, section five summarizes the limitations of this paper and directions for future research.

2. Data and Methodology

To conduct this research, a systematic review and screening of existing scientific papers on real-time emotion recognition using artificial intelligence technology were performed. High-impact, state-of-the-art literature databases were used to search for and extract subject-related publications. An SLR has a higher quality (Siddaway et al., 2019) than other types of literature reviews because it is an independent academic method (Feak & Swales, 2009) designed to identify and evaluate all relevant literature to facilitate conclusions about the questions considered in a particular research topic. SLR can reveal the status of research on a topic (Jesson et al., 2011) while identifying research gaps and areas requiring further research, and it involves a series of steps: search strings, selection and evaluation of articles, synthesis (Hart, 2011), and analysis. To evaluate the relevant literature, the PRISMA protocol, which stands for Preferred Reporting Items for Systematic Review and Meta-Analysis, is used as a structured methodology that considers sources, numbers, and fates of all identified and screened records in this review (Haddaway et al., 2022).

2.1. Literature identification

All the literature in this review was searched in international journals and conference proceedings in computer science and artificial intelligence, with databases such as Web of Science, Scopus, IEEE Xplore, and ACM Digital Library used to access peer-reviewed articles and relevant publications in these fields. As the review topic of this article, artificial intelligence and facial expression recognition are given priority when setting search rules. To consider facial expression recognition publications as comprehensively as possible, the search rules for the literature search are set to include elements that cover facial expression recognition and emotion recognition.

The search string rule of this review is TS = (“Facial expression recognition” OR “Emotion recognition”) AND (“Artificial intelligence” OR “Machine learning” OR “Deep learning”) AND (“Real-time” OR “Real-time”), which is applied across four major scientific databases. The TS represents the subject search criteria, including title and abstract, with specific variations for each database: Scopus, which includes title, abstract, author keywords, and keyword plus; ACM Digital Library, which includes title, abstract, and author keywords; and IEEE Xplore, which includes title, abstract, and author keywords. This search resulted in over 10000 records. A publication period of nearly five years was then defined, with a timespan from 2019/01/01 to 2023/12/31. The language was English, and the type of literature was peer-reviewed, including journal articles, conference proceedings, and a limited number of book chapters. This refined search identified 2666 relevant records.

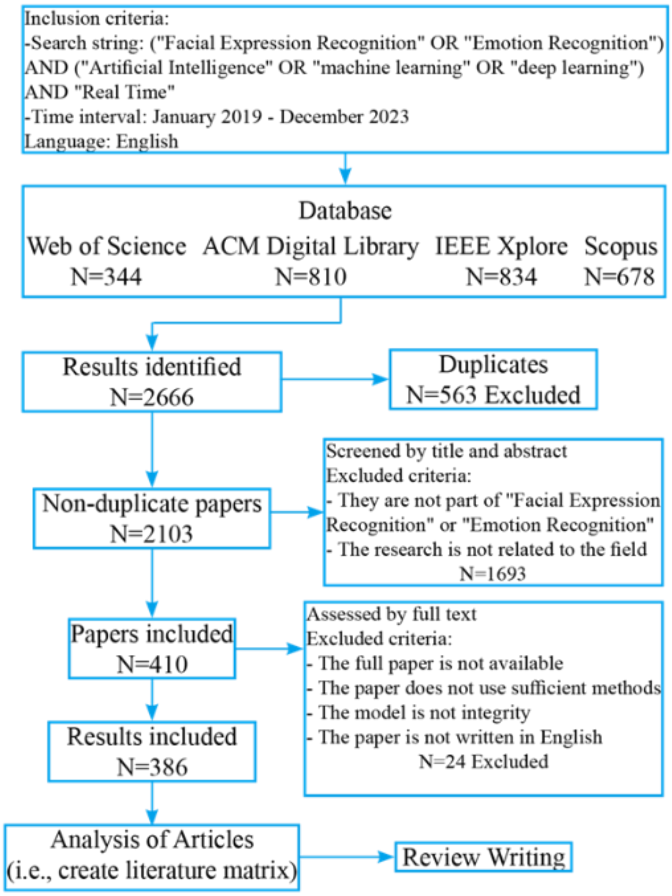


Figure 1 Flowchart of the systematic review according to the PRISMA standard.

2.2. Screening for inclusion

After these literature records from the four databases were imported into Zotero literature management software, duplicate studies were identified, and 563 records were duplicated. The initial screening round for the 2103 articles was conducted by reviewing their titles and abstracts. To further evaluate their relevance to the research topic, real-time emotion recognition based on artificial intelligence technology was used. Two main types of articles were excluded during the screening process. The first type is unrelated to “facial expression recognition” and “emotion recognition”. It does not involve emotion recognition; the other type is a separate study of emotion recognition. Nevertheless, it does not mention facial expression recognition, such as only using biological signals for human emotion recognition, but does not involve facial expression recognition (Zhu et al., 2019; Patil et al., 2023).

In addition, this review aims to provide a technical summary for researchers in artificial intelligence and facial expression recognition; some articles that do not mention specific algorithms or AI techniques were excluded. After that, 410 articles that fit the topic of this review were selected.

2.3. Quality and eligibility assessment

The quality and eligibility of the 410 articles were assessed by reviewing their full texts. The criteria for screening the articles were full paper availability, the use of sufficient methods, the integrity of the proposed methodology, and the use of only English-language papers. Finally, 24 articles were excluded, and 386 articles that fit the topic of this review were included, as shown in Figure 1.

2.4. Analysis and reporting

At this stage, employing a literature matrix is crucial for effective data extraction and rigorous synthesis in the review process. A literature matrix based on the methodologies recommended by (Baltrusaitis et al., 2019; Kabakus, 2020). This matrix serves to systematically analyse the content of the literature and evaluate the results reported in the included articles.

AI Technology: Types of artificial intelligence technologies referenced in the literature.

Algorithms and Model Selection: The algorithms and models employed in each study.

Identification speed: The speed at which models can identify facial expressions.

Databases: The datasets utilized for training and validation.

Field of application: Review of the application fields of facial expression recognition in the literature.

This structured approach enables a comprehensive understanding of how different factors influence the effectiveness of facial expression recognition technologies.

3. Evaluation and Key Findings

3.1. Descriptive analysis

This section provides a comprehensive examination of the distribution and analysis of publications related to facial expression recognition via artificial intelligence technology. It begins with an overview of the types and frequency of publications, highlighting significant trends and preferences within the research community. A detailed analysis of the journals and conference proceedings in which these studies have been published will subsequently be presented, offering insights into the most influential platforms for disseminating research findings in this rapidly evolving field.

3.1.1. Publications distribution

As demonstrated in Figure 2, among the 386 articles included in the systematic review, 76% (293 articles) were conference papers, 24% (91 articles) were journal articles, and two were book chapters. Figure 2 indicates that within the domain of facial expression recognition via artificial intelligence technology, researchers tend to favour the dissemination of discoveries at international conferences.

Figure 3 illustrates the temporal distribution of publications concerning the subject of this review from January 2019 to December 2023. Publication trends can be identified by analyzing the publication years of the selected literature. The data reveal a steady and substantial rise in publications between 2019 and 2023, reaching its highest volume in 2023. There is growing interest and concern among researchers in the realm of real-time facial expression recognition via artificial intelligence technology (Baltrusaitis et al., 2019).

3.1.2. Publications analysis

A review of the selected literature, including both conference and journal papers, identified a total of 59 distinct journals. Figure 4 clearly illustrates that IEEE Access and Multimedia Tools and Applications are the top two journals, characterized by their prolific output of scholarly articles, with IEEE Access featuring 15 papers (Kabakus, 2020; Wang et al., 2021; Yang et al., 2021; Shi et al., 2020) on related topics. Additionally, three journals have three articles each: ACM Transactions on Multimedia Computing, Communications, and Applications, Applied Sciences (Switzerland), and Sensors (Switzerland). Furthermore, eight journals published two related articles, whereas the remaining journals each published only one paper on the topic under review. Notably, among the 59 journals, 24 are IEEE series journals, collectively contributing 40 papers, which represents 44% of the total number of journal articles.

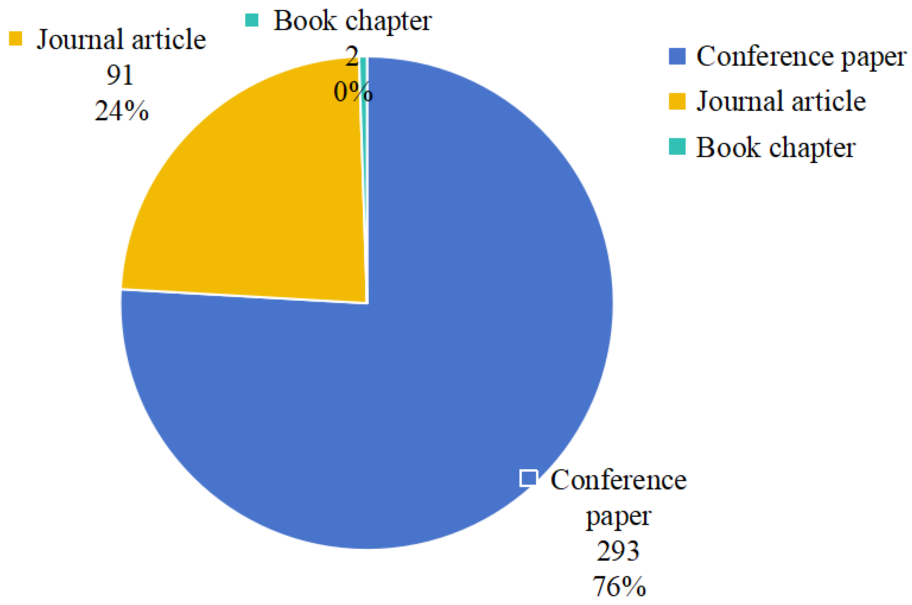


Figure 2 Types of papers.

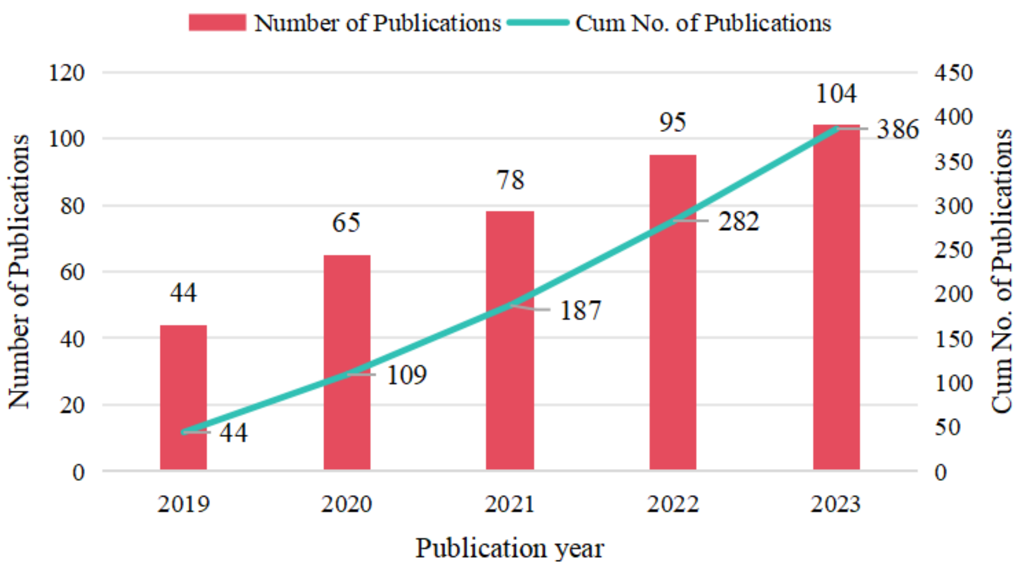


Figure 3 Number and trend of publications (2019–2023).

The statistics of the conference papers revealed that 293 conference papers were included. As illustrated in Figure 5, the names of the conference proceedings with three or more publications are listed separately, totaling six conferences. The International Conference on Multimedia Interaction (ICMI) has five papers (Dhall, 2019; Salman & Busso, 2022; Tsioutas et al., 2021; Withanage Don et al., 2023; Zhu et al., 2020), which is the largest number of publications. The remaining 24 conference proceedings published two articles, and 76% (224) of the conference proceedings had only one article related to the topic of this review.

Figure 6 illustrates the distribution of country-based publications focusing on AI techniques within facial expression recognition. Leading countries in Asia contribute 68% of the total scientific production in this field, whereas leading countries in Europe contribute 26%. Most publications in this area originated from India, with 147 articles (Jaiswal & Nandi, 2020; Karnati et al., 2022) representing 38% of the total publications. This substantial number underscores India's significant emphasis on the application of AI technology in the realm of facial expression recognition. China emerged as the second most prolific country, contributing 86 articles (Zhao et al., 2020; Zhou et al., 2021), accounting for 22% of the total publications. The United

States published 18 articles on this topic, and the remaining countries had fewer than 10 publications (Figure 6 lists only countries with two or more publications).

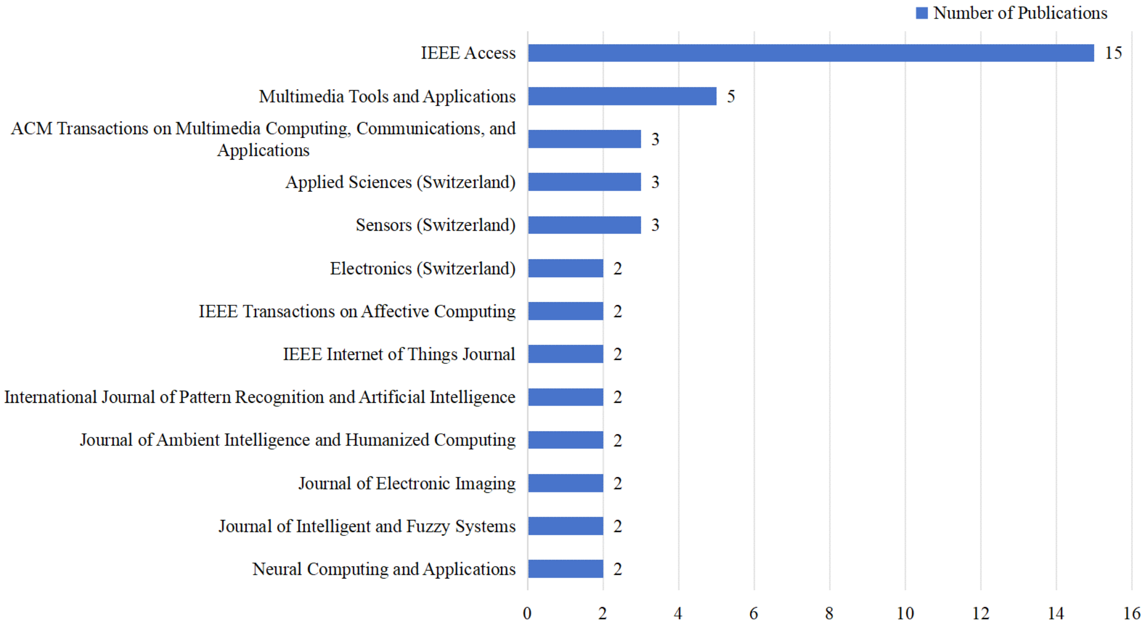


Figure 4 Journals with the most publications.

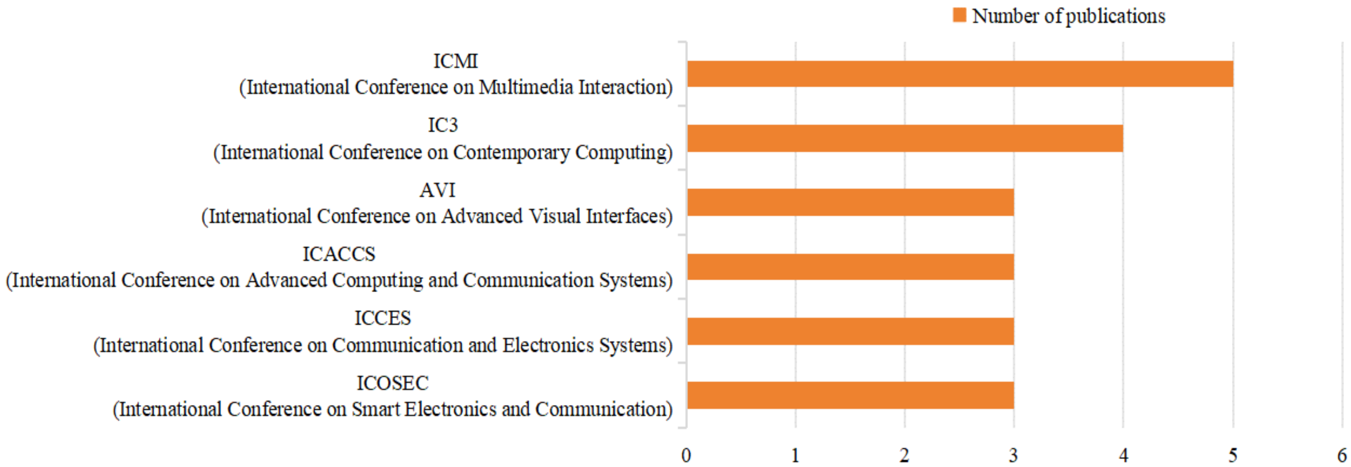


Figure 5 Conference with the most publications.

Notably, while Russia (Savchenko, 2022; Mikhaylevskiy et al., 2021) and Saudi Arabia (Alreshidi & Ullah, 2020; Khan et al., 2022) have only six publications, their citations are 119 and 93, respectively, indicating strong recognition by their peers. Such a high citation count suggests that the research produced in these countries may be of notably high quality, potentially published in well-regarded journals with significant impact factors. Moreover, many countries, such as Indonesia (Dwijayanti et al., 2022), Korea (Jeong et al., 2020), Egypt (Siam et al., 2022), Romania (Melinte & Vladareanu, 2020), and Turkey (Ozdemir et al., 2019), have approximately three publications each; however, they all present high citation rates. This indicates that despite their lower quantity, their research contributions are influential and well regarded, often appearing in impactful journals. This observation emphasizes the importance of not only the publication quantity but also the quality of the research, as reflected in the citation metrics. Recent studies (Zahara et al., 2020; Melinte & Vladareanu, 2020; Siam et al., 2022) published within the last three years further highlight the ongoing contributions from these regions, suggesting a growing engagement in high-quality research in facial expression recognition via AI technologies.

3.2. Keyword co-occurrence analysis and literature review

This section describes bibliometric analysis via VOSviewer software. The purpose of bibliometric analysis is to meticulously scrutinize the scholarly literature within a defined research domain and extract and examine patterns, trends, and impacts over a specified timeframe (Lievrouw, 1989; Van Eck & Waltman, 2010). This section uses VOSviewer for keyword co-occurrence analysis, with a focus on clusters and a systematic review around each cluster.

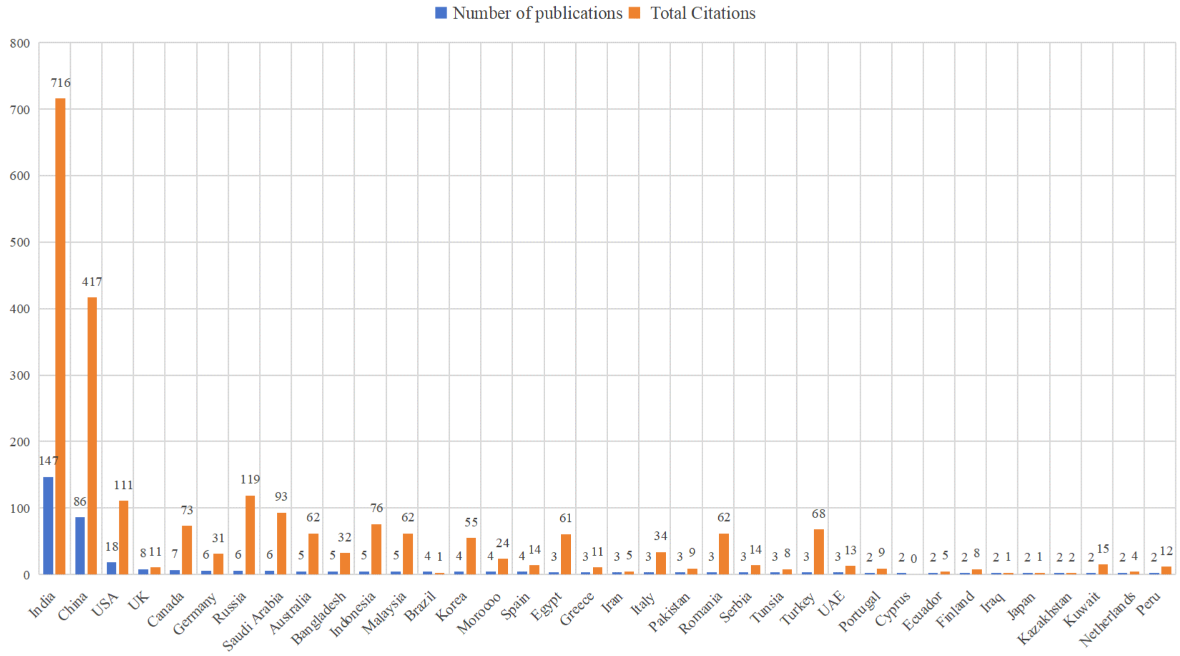


Figure 6 Number of country publications and citations.

The analysis of keyword co-occurrence is a crucial feature of VOSviewer software. It creates clusters of keywords on the basis of their correlation strength and uses natural language processing algorithms (Eck & Waltman, 2009) to calculate them. The software identifies the frequency of different keywords in a text collection and categorizes them with distinct colors (Waltman et al., 2010). Keywords within the same cluster signify similar or closely related research directions or themes, and the distance between the circles reflects the correlation between the keywords.

Using the foundational details of these clusters, a keyword co-occurrence analysis was conducted on the 386 articles included in this systematic review. The visual map incorporated a threshold criterion for terms that appeared more than twice. As depicted in Figure 7, this criterion resulted in the inclusion of 90 keywords, which were categorized into 12 distinct clusters, each represented by a unique color. Among these, the quintet of keywords with the most robust interconnections was found to be centered on “neural networks” (465, including similar terms such as “convolutional neural network” and “recurrent neural network”), deep learning (284, including similar terms “transfer learning”), facial expression recognition (281, including similar terms “facial emotion recognition” and facial expression), humanities (253, including similar terms “humanism” and “humans”), and feature extraction (208). Figure 7 indicates that the most influential keywords within this domain, characterized by the strongest interconnections, are predominantly associated with the themes of neural networks, deep learning, facial expression recognition, humanities, and feature extraction.

Figure 7 shows five main clusters, each represented by a different color. The main keywords in cluster 1 (red) included, for instance, deep learning, facial expression recognition, feature extraction, “convolutional neural networks”, edge computing, real-time systems, and computational modelling. Cluster 2 (designated in green) encompasses a spectrum of pivotal terms, with 'neural networks', 'emotion', 'classification', 'dataset', and 'expression' emerging as the most salient within this thematic grouping. In Cluster 3, denoted in blue, the core terms include machine learning, sentiment analysis, attention mechanism, student feedback, and emotion classification, which are instrumental in defining the thematic scope of this cluster. Cluster 4, identified by the hue of yellow, is characterized by a set of central concepts, namely, 'computer vision', 'image processing', 'human–computer interaction', 'state of mind', and the seminal 'AlexNet', which collectively define the focal points of this research cluster. The main keywords in cluster 5 (purple) are face detection, “recurrence”, “cascade classifier”, “LSTM”, “random forest”, and “internet of things”. Detailed information regarding the keywords is also presented in Table 1.

Figure 8 shows an integrated keyword interaction map, delineating the chronological distribution and progression in the frequency of keyword usage across the specified years. The variation in color intensity within the network corresponds to the average publication year of a given keyword. As indicated in the lower right quadrant of the figure, increased darkness of the hue is indicative of earlier publication dates, whereas lighter shades correspond to more recent publications. A visual inspection of Figure 8 reveals that, in 2020, the most frequently occurring keywords were "neural networks", "support vector machines", "real-time", "object detection", and "feature extraction", indicating a concentrated research interest in neural networks, feature extraction, and object detection methodologies. By 2022, the prevalent keywords shifted to include "transfer learning", "ASD children", "real-time systems", "recurrence", "Raspberry Pi", "the Internet of Things", "computational modelling", "online learning", and "engagement detection". This shift underscores a gradual reorientation of research focus toward applying (e.g., IOT, education, and medical) facial expression recognition across diverse contexts and developing real-time systems.

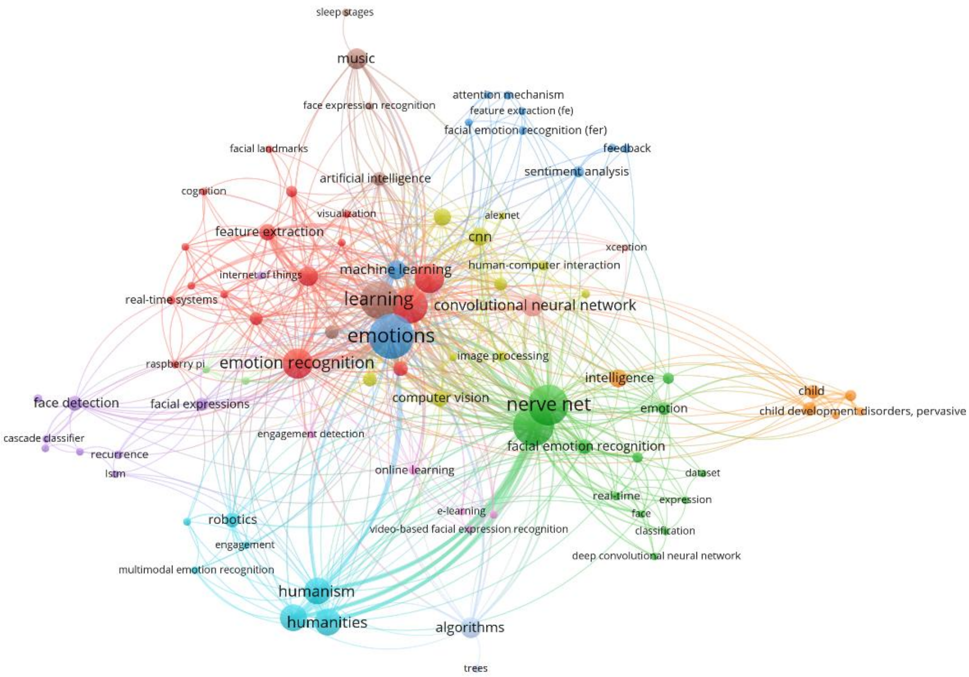


Figure 7 Keyword co-occurrence network analysis spanning 2019 - 2023.

Table 1 Top 5 clusters of the co-occurrence network of keywords.

Clusters	Keywords
Cluster 1 (red)	Deep learning, Facial expression recognition, Feature extraction, Convolutional neural networks, Edge computing, Real-time systems, Computational modelling
Cluster 2 (green)	Neural networks, Emotion, classification, Dataset, Expression
Cluster 3 (blue)	Machine learning, Sentiment analysis, Attention mechanism, Student feedback, Emotion classification
Cluster 4 (yellow)	Computer vision, Image processing, Human-computer interaction, State of mind, AlexNet
Cluster 5 (purple)	Face detection, Recurrence, Cascade classifier, LSTM, random forest, Internet of things

3.3. Most highly cited scholarly articles

A corpus of literature has emerged within the scholarly domain of facial expression recognition based on artificial intelligence technology. It is characterized by depth, breadth, and exploration of multiple fields of application. Within the scope of this systematic review, which included 386 scholarly articles, a few were distinguished by their significant contributions to the field, as evidenced by their high citation count. Following the methodological framework proposed by (Chen, 2017), a citation threshold of 30 was applied. Consequently, this review includes only those studies cited in at least 30 other papers (Table 2). Notably, older articles tend to accumulate a greater number of citations over time, whereas newly published articles often receive few or no citations. The citation counts analyzed in this study reflect data as of January 2024.

Table 2 reveals that the article by (Ozdemir et al., 2019) stands out as the most frequently referenced, highlighting its prominence within the scholarly discourse of the period, which was cited 96 times at the time when the dataset was collected. Notably, the aggregate citation frequency for the work of (Savchenko et al., 2022) is less than that of the most highly cited article within the domain, but it was published in 2022, and the annual average citations are higher than those reported by Prana et al. (2020). Prana et al. (2020) emphasized advancements in algorithmic methodologies by introducing a two-layer convolutional network model that leverages the Adam optimizer to minimize the loss function; this framework offers a foundation for real-time video sequence analysis to discern shifts in emotional states. Moreover, Andrey et al. (2022) examined student conduct within digital learning contexts and proposed an innovative framework to integrate seamlessly with existing e-learning infrastructures. This proposed model aims to facilitate the swift and precise evaluation of classroom sentiment and student engagement levels, enhancing educational experience and outcomes.

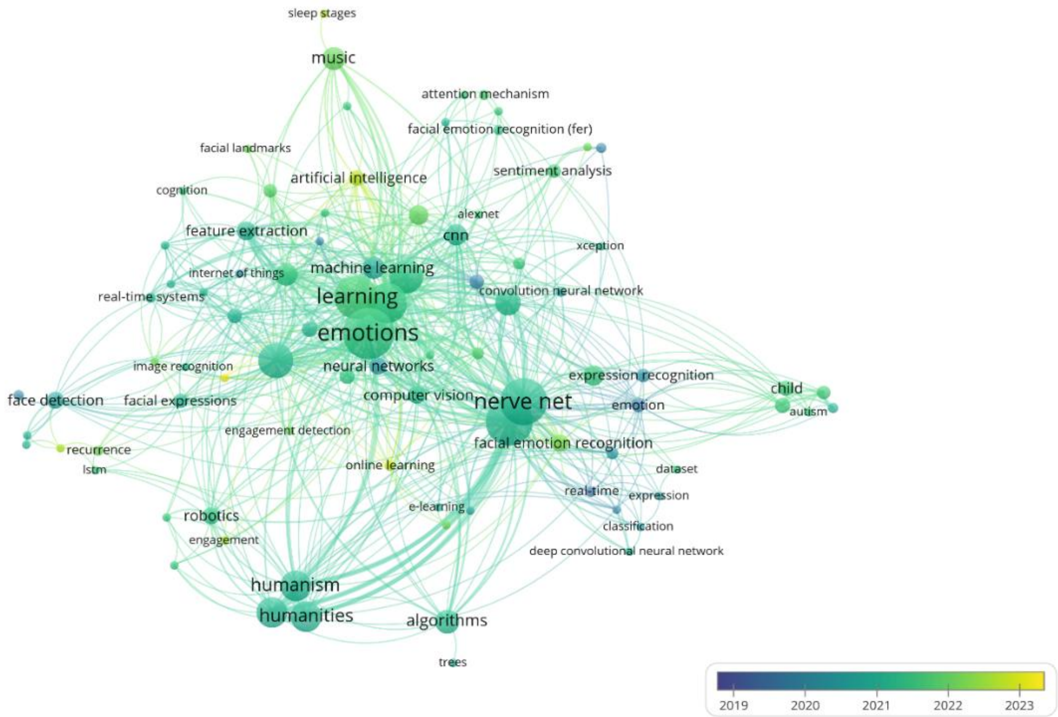


Figure 8 Integrated keyword co-occurrence visualization for the period from 2019–2023.

Table 2 Top 18 most cited papers.

Authors	Title	Citations	Applications
(E. Pranav et al., 2020)	Facial Emotion Recognition Using Deep Convolutional Neural Network	96	Computer Science
(A. V. Savchenko et al., 2022)	Classifying Emotions and Engagement in Online Learning Based on a Single Facial Expression Recognition Neural Network	89	Education
(Melinte & Vladareanu, 2020)	Facial expressions recognition for human–robot interaction using deep convolutional neural networks with rectified Adam optimizer	61	Robot
(L. Zahara et al., 2020)	The Facial Emotion Recognition Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi	57	Computer Science
(Dhall, 2019)	EmotiW 2019: Automatic Emotion, Engagement and Cohesion Prediction Tasks	53	Computer Science
(M. A. Ozdemir et al., 2019)	Real Time Emotion Recognition from Facial Expressions Using CNN Architecture	52	Computer Science
(Bhardwaj et al., 2021)	Application of Deep Learning on Student Engagement in e-learning environments	49	Education
(Ab Wahab et al., 2021)	Efficientnet-Lite and Hybrid CNN-KNN Implementation for Facial Expression Recognition on Raspberry Pi	49	IOT
(Riaz et al., 2020)	eXnet: An efficient approach for emotion recognition in the wild	49	Computer Science
(G. Muhammad & M. S. Hossain, 2021)	Emotion Recognition for Cognitive Edge Computing Using Deep Learning	45	IOT
(N. Zhou et al., 2021)	A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection	44	Computer Science
(Rajan et al., 2020)	Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM	42	Computer Science
(Siam et al., 2022)	Deploying Machine Learning Techniques for Human Emotion Detection	42	Computer Science
(M. Jeong et al., 2020)	Lightweight Multilayer Random Forests for Monitoring Driver Emotional Status	41	Driver status
(Alreshidi & Ullah, 2020)	Facial emotion recognition using hybrid features	39	Computer Science

(Jaiswal & Nandi, 2020)	Robust real-time emotion detection system using CNN architecture	37	Computer Science
(Miao et al., 2019)	A Deep Learning System for Recognizing Facial Expression in Real-Time	30	Computer Science
(G. Zhao et al., 2020)	Expression Recognition Method Based on a Lightweight Convolutional Neural Network	30	Computer Science

As illustrated in Table 2, there has been increasing scholarly interest in the deployment of facial expression recognition (FER) across various domains, including education (Savchenko et al., 2022; Bhardwaj et al., 2021; Mukhopadhyay et al., 2020), the Internet of Things (IoT) (Ab Wahab et al., 2021; Muhammad & Hossain, 2021), driver emotion detection (Jeong et al., 2020), and human–robot interaction (Melinte & Vladareanu, 2020). As discussed in Section 3.2, the period from 2019--2023 witnessed a notable evolution in research focus, with keywords progressively aligning with applications within education, healthcare, and the IoT. This trend further validates the findings presented in the review.

3.4. Datasets.

Facial expression recognition datasets are crucial in advancing research and applications in computer vision, psychology, and human–computer interaction (Canedo & Neves, 2019). These datasets typically consist of images or videos that capture various facial expressions across different individuals and are annotated with labels indicating the corresponding emotions, and some of the databases have biosignals (e.g., EEG, EDA, or EMG). Researchers utilize these datasets to train and evaluate algorithms for facial expression recognition, emotion detection, and synthesis (Kanade et al., 2000). Access to diverse and well-annotated facial expression datasets is essential for developing robust and accurate models capable of effectively understanding and interpreting human emotions.

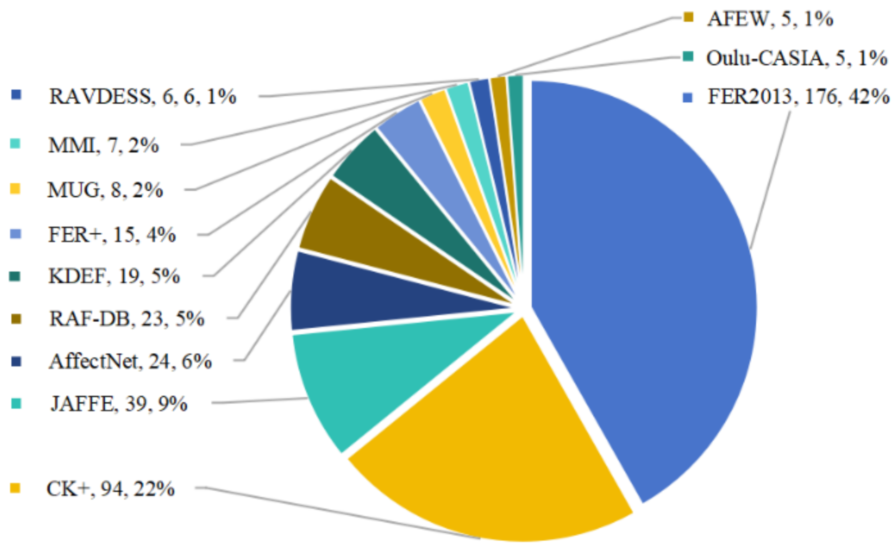


Figure 9 Distribution of the top 12 most utilized databases in FER.

A total of 81 facial expression datasets were identified in this systematic review. This section highlights those that have garnered significant attention within the research community. Figure 9 illustrates that 12 datasets have been utilized on more than five occasions throughout the literature under review. Notably, the dataset identified as FER2013 (Goel et al., 2022; Afsar et al., 2023; Alhajlah, 2023) emerges as the most frequently employed, with a total of 176 times, followed by the CK+ dataset (De Ocampo, 2023; Hatem & Al-Bakry, 2022; Lee et al., 2021; Zhang et al., 2022; Zhang & Tang, 2021), which has been utilized 94 times. Table 3 provides a comprehensive overview of the principal characteristics of these datasets, including their year of publication, the type of sample, and their overall data volume. Additionally, this section delves into a detailed analysis of the mentioned datasets:

Facial expression recognition 2013 (FER2013): The dataset FER2013 is widely recognized as the researchers' most extensively utilized collection of facial expressions. In this review, it emerged as the most frequently employed dataset, having utilized 176 instances. It comprises 35,887 facial expression images, with each image having a resolution of 48×48 pixels, and is organized into seven distinct expression classes (Angry, Disgust, Fear, Happiness, Surprise, Neutral, Sadness). Notably, the dataset contains only 600 images of disgust expressions, whereas the remaining labels each feature approximately 5,000 images. The compilation of the FER2013 dataset is attributed to the collaborative efforts of Pierre-Luc Carrier and Aaron Courville, who initially provided a version of the dataset for the Kaggle competition (Goodfellow et al., 2015).

Table 3 Key features of the 12 leading databases in FER.

Databases	Year	Type of sample	Capacity	Environment	Number of Emotions	Identities
FER2013	2013	Images	35886	Uncontrolled	7	N/A
CK+	2010	Images	593	Laboratory	8	137
JAFFE	1998	Images	213	Laboratory	7	10
AffectNet	2017	Images	291651	Wild	8	N/A
RAF-DB	2015	Images	29672	Wild	7 and 12 compound emotions	N/A
KDFE	1998	Images	4900	Laboratory	7	70
FER+	2018	Images	35887	Wild	8	N/A
MUG	2011	Multimedia-Images, Videos	2283	Laboratory	7	86
MMI	2002	Multi Videos, Images	29000	Laboratory	6	75
RAVDESS	2018	Multi-Audio only, Audio-Video, Video only	7356	Wild	8	24
AFEW-VA	2012	Videos	600	Wild	7	N/A
Oulu-CASIA	2005	Images	2400	Laboratory	6	80

Cohn-Kanade (CK+): The CK+ AU-Coded Facial Expression Database presents a comprehensive platform for the advancement of automated facial image analysis research, enriched with detailed metadata that encompasses annotations of all facial action coding system (FACS) action units alongside emotionally specified expressions (Tian et al., 2001). This repository contains 593 expression image sequences, each with dimensions of 640×480 pixels, and categorizes facial expressions into eight distinct classifications: anger, contempt, fear, happy, sad, disgust, neutral, and surprise (Lucey et al., 2010). This dataset is a pivotal resource for developing and testing facial recognition algorithms and methodologies within computer vision and affective computing, especially in video-based facial expression research.

Japanese Female Facial Expression (JAFFE): The JAFFE dataset consists of 213 images of facial expressions, each with a resolution of 256×256 pixels. These images encapsulate seven distinct emotional states: anger, happy, sad, surprise, disgust, fear, and neutral. The dataset was constructed via the expressive portrayals of 10 Japanese female individuals, each depicting emotions. Furthermore, every image within the dataset underwent a semantic evaluation, where 60 Japanese reviewers assigned ratings across the seven facial expression categories, thereby enriching the dataset's utility for academic and research purposes (Lyons et al., 1998).

AffectNet: The AffectNet collection represents a comprehensive collection of facial expressions captured in the wild, distinguished by its method of collation and annotation of facial images. This database categorizes images into eight distinct emotion categories, namely, neutral, fear, happiness, sadness, disgust, anger, surprise, and contempt, on the basis of a classification model. Additionally, it includes evaluations of emotional intensity in terms of valence and arousal, following a dimensional model (Mollahosseini et al., 2019). With a repository of 291,651 images, AffectNet is the most extensive dataset of facial expressions currently available. Although not publicly accessible, AffectNet invites students and researchers to access the database through a formal agreement process.

Real-World Affective Faces Database (RAF-DB): RAF-DB is an extensive and high-quality collection of facial expression images designed to capture a wide range of real-life emotional displays. It comprises 29,672 images, each assessed by approximately 40 evaluators, covering seven primary emotions: joy, fear, surprise, disgust, sadness, anger, and natural, along with two subsets of labels encompassing 12 categories of compound emotions, five precise landmark locations, and 37 automatic landmark locations (Li et al., 2017). This database is segregated into training and testing sets, with the training set being substantially larger than the testing set, specifically five times the size.

Karolinska Directed Emotional Faces (KDEF): This database, developed by researchers at the Karolinska Institute in Sweden, is a widely used and respected resource in the fields of psychology and neuroscience. This collection is particularly valued for its contributions to emotion research, providing a robust tool for studying facial expressions and their psychological implications. Compared with 4,900 standardized facial expression images containing seven distinct emotions—sadness, fear, anger, disgust, surprise, happiness, and neutrality—the database is carefully controlled to maintain uniformity across images, with careful attention to factors such as lighting, background, and facial orientation (Calvo & Lundqvist, 2008).

Facial expression recognition plus (FER+): FER+ builds upon the FER2013 dataset, which is designed to overcome its limitations, including issues with noise labels and an unbalanced distribution of training samples. FER+ consists of 35,887 images of facial expressions categorized into seven emotional classifications, mirroring the original FER dataset. Additionally, FER+ includes a cover (e.g., sunglasses and masks) and indicators of emotional intensity.

Multimedia Understanding Group (MUG): The MUG dataset, developed by the Multimedia Understanding Group, was established to address the limitations observed in existing databases, particularly with respect to illumination and resolution quality. Compared with facial expression images from 86 subjects, the MUG database encompasses 2,283 images, each with a resolution of 896×896 pixels (Aifanti et al., 2010). These images are categorized into six primary emotional states: anger, sadness, disgust, fear, happiness, and surprise.

MMI Database (MMI): The MMI dataset is set up to offer the facial expression analysis community an extensive collection of visual data capturing comprehensive temporal patterns of facial expressions. It includes six core emotions, as well as expressions associated with individual facial action units (AUs) from the facial action coding system (FACS). The dataset comprises 2,900 video recordings and high-resolution still images representing 75 participants (Pantic et al., 2005) and is made publicly accessible to the scientific research community.

The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): The RAVDESS dataset is a multimodal compilation of facial and voice expressions, containing 7,356 samples across three styles: audio, video-audio, and video (Livingstone & Russo, 2018). The dataset features a range of emotional expressions, including calm, happiness, sadness, anger, fear, surprise, and disgust in speech, as well as calm, happiness, sadness, anger, and fear in song. For each emotion, two levels of intensity, normal and heightened, are represented, in addition to a neutral expression.

Database for Valence and Arousal Estimation in The WILD (AFEW-VA): The AFEW-VA dataset comprises detailed frame-by-frame annotations for emotional valence and arousal, capturing nuanced variations in these dimensions. It includes annotations of 68 facial landmarks for each frame from 600 complex video segments sourced from movies or TV series (Kossaifi et al., 2017). These video clips are annotated by discrete emotion categories, forming the AFEW database.

Oulu-CASIA NIR-VIS Facial Expression Database (Oulu-CASIA): The Oulu-CASIA database comprises facial expressions depicting six emotions—surprise, fear, anger, happy, sad, and disgust—captured from 80 individuals ranging in age from 23--58 years. The imaging system used to record the series of images operates at a capture rate of 25 frames per second, resulting in 2,400 facial expression images, each with a resolution of 320×240 pixels (Taini et al., 2008).

4. Facial Expression Recognition (FER)

The provided articles illustrated the use of a variety of machine learning and deep learning models in the recognition of facial expressions, as depicted in Figure 10. Close to 94% of the studies utilized deep learning frameworks, including transfer learning. Transfer learning is considered a part of deep learning due to its popularity as a method that uses a pre-trained model to expedite the training process for a new computer vision task by leveraging knowledge gained from a related task. Owing to its frequent usage, transfer learning is presented separately. Figure 11 illustrates the distribution of models used in the selected articles, with the convolutional neural network (CNN) algorithm being the most prevalent (267), followed by SVM (24) and VGG (20). In 1986, (Rumelhart et al., 1986) highlighted the susceptibility of the backpropagation algorithm of artificial neural networks (ANNs) to overfitting and prolonged training. In contrast, Yann Lecun (Lecun et al., 1998) of New York University proposed a CNN in 1988. This model, supervised learning under deep learning, demonstrates strong adaptability and proficiency in extracting local data features. Its weight-sharing network structure makes it more akin to biological neural networks and has yielded positive outcomes across various pattern recognition domains.

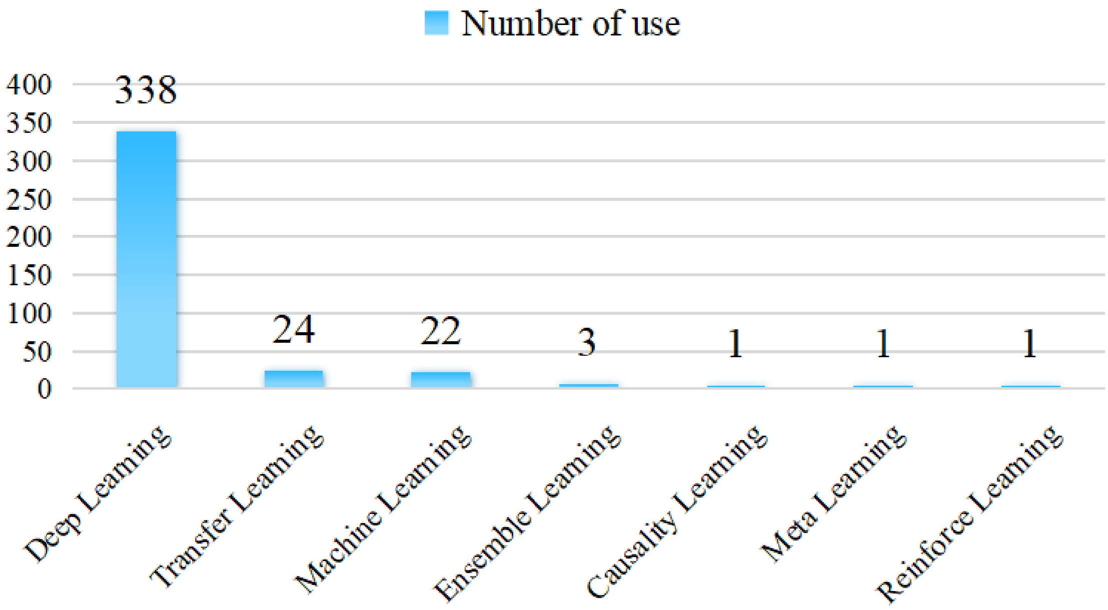


Figure 10 Distribution of AI techniques.

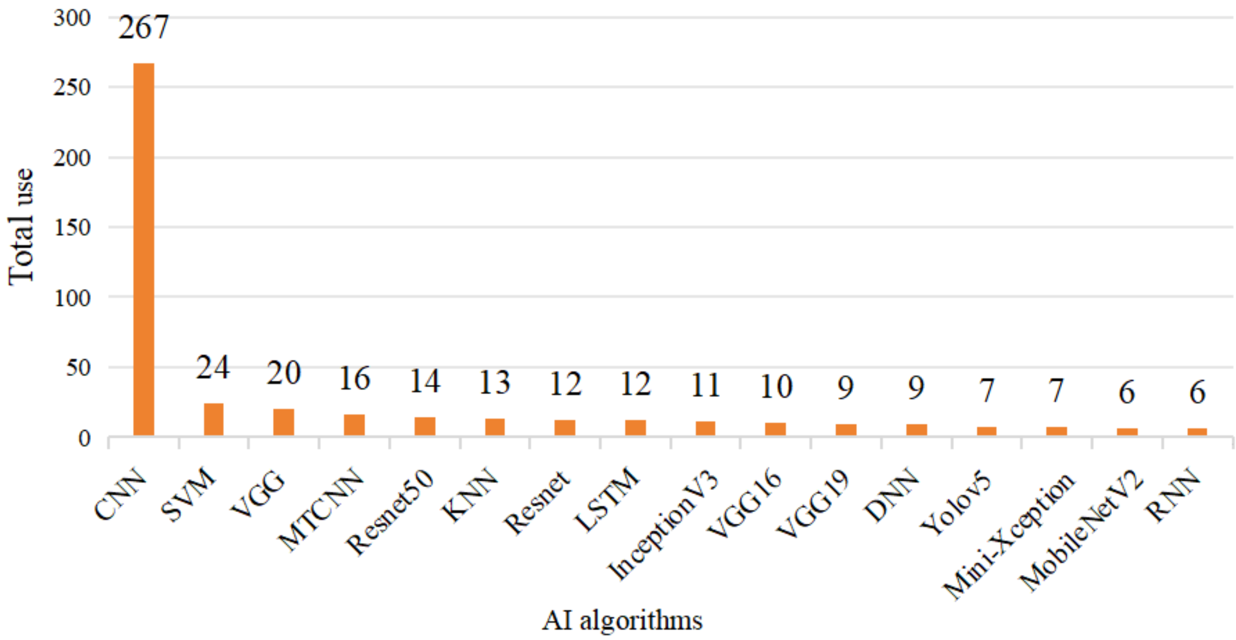


Figure 11 AI algorithm usage time distributions.

4.1. CNN-based FER

Table 4 indicates the selected articles with the highest model recognition accuracy, surpassing 99%, within the convolutional neural network (CNN) architecture among the selected scholarly corpus. The uniform application of CNN structures across these articles emphasizes their advantages in tasks such as object detection and image segmentation, a notion corroborated by the literature. (Talaat, 2023) achieves an exceptional recognition accuracy of 99.99% and introduces an enhanced deep learning (EDL) approach. This innovation is pivotal in developing an emotion detection framework that synergizes with the fog server and the Internet of Things (IoT) to mitigate real-time monitoring latency, enabling swift and accurate identification of emotional expressions in children diagnosed with autism spectrum conditions (ASC). Furthermore, (Jaffar & Abdulbaqi, 2022) employ the CNN architecture, which explicitly incorporates the LeNet model, to devise a facial expression recognition (FER) classification system. This system is applied to help parents and therapists discern the facial expressions of ASD-affected children to foster their well-being and societal integration. Similarly, (Dhope & Neelagar, 2022; Talaat, 2023) leverage IoT technology in concert with a Raspberry Pi 3B+ board, a PI camera, and the PyQt5 framework to create a graphical user interface (GUI) for an AI-driven facial emotion recognition system that achieves a commendable accuracy of 99.88%. (Ashraf et al., 2023) proposed a novel fusion network that integrates an end-to-end attention module for extracting facial expression features combined with YOLOv4, which culminates in establishing a facial emotion recognition system with an accuracy of 99.68%.

Additionally, (Kong et al., 2021) introduced the application of an attention mechanism and key region fusion (WFA-CNN) to accentuate the learning of pivotal facial features, utilizing the Mini_Xception model to curtail computational complexity, thereby achieving promising experimental outcomes (99.6%). Notably, (Ashraf et al., 2023; Wang et al., 2022) pivot towards video-based facial expression recognition models; (Hatem & Al-Bakry, 2022) elucidate an efficient motion pattern through the incorporation of a Diff-based Canny operator (DCO) across various CNN layers to facilitate interframe aggregation and engender novel feature modes. Conversely, (Ashraf et al., 2023) harnesses the ReLU function to significantly streamline the training process of neural networks while also incorporating dropout, batch normalization, and L2 regularization strategies to ameliorate overfitting concerns. (Ashraf et al., 2023) introduced a unique framework that combines real-time 2D body posture analysis with emotion detection from facial expressions. They adopt the same ReLU function and dropout layer as (Ashraf et al., 2023) to forestall model overfitting. (Webb et al., 2020) presented a deep convolutional neural network (DCNN) model that was pretrained as a stacked convolutional encoder (SCAE) and subjected to unsupervised training. (Ramirez Rios & María Reyes Duke, 2023) explored the use of the YOLOv8 framework for object detection, implementing a flexible and iterative approach to the development of CNNs.

In conclusion, the CNN performs well in facial expression recognition. Compared with CNNs without optimization methods, CNNs with optimization methods have shorter processing times and higher accuracy. Typically, researchers incorporate attention mechanism models, integrate additional modules across different CNN layers, employ various activation functions, and utilize L2 regularization and other techniques to prevent overfitting. The selected articles feature commonly used terms such as "image processing," "feature extraction," “data augmentation,” and "face cropping," which highlight the diverse methods employed by researchers to increase model performance.

Table 4 Top 10 most accurate articles on the basis of the CNN.

Authors	Algorithms	Accuracy	Database	Types of Emotions	Applications
(Talaat, 2023)	CNN,GAs	99.99%	AUTISTIC	anger, fear, joy, natural, sadness, surprise	IoT, Medical
(Dhope & Neelagar, 2022)	CNN	99.88%	Own Collection	anger, disgust, fear, happiness, sadness, surprise, and neutral	IoT
(Jaffar & Abdulbaqi, 2022)	CNN	99.78%	KDEF, Alnoor Specialized Center	anger, disgust, fear, happiness, sadness, surprise, and neutral	Medical
(Wang et al., 2022)	CNN, ResNet	99.69%	CK+, AFEW	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Shit et al., 2023)	CNN, Yolov4	99.68%	FER+, KDEF, CK+, JAFFE	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Adhikary & Bhandari, 2021)	CNN	99.60%	CK+, MPII	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Kong et al., 2021)	CNN, Xception	99.60%	JAFFE, CK+	anger, disgust, fear, happiness, sadness, surprise, contempt	Computer Science
(Webb et al., 2020)	CNN	99.52%	JAFFE, FEEDTUM, KDEF	anger, disgust, fear, happiness, sadness, surprise, and neutral	Robot
(Ashraf et al., 2023)	CNN	99.38%	WSEFEP, CK+	anger, sad, happy, neutral, surprised	Computer Science
(Ramirez Rios & María Reyes Duke, 2023)	CNN, Yolov8	99.00%	Roboflow for own dataset	neutral, anxiety, tension, frustration, fatigue, happiness	Computer Science

4.2. VGG-based FER

The VGG network is a convolutional neural network (CNN) architecture designed by the Visual Geometry Group at the University of Oxford in collaboration with Google DeepMind. VGG has demonstrated that the depth of the CNN and the use of small convolutional kernels significantly impact image recognition and classification (Simonyan & Zisserman, 2014). The most commonly used VGG structures include those with 16 layers and 19 layers, which consist of convolutional and fully connected layers but do not consist of pooling and activation layers. The VGG network employs a ReLU activation function after each convolutional layer (Sengupta et al., 2019; Vedaldi & Zisserman, 2016), accelerating the training process and mitigating the vanishing gradient problem. Notably, VGG networks are characterized by their deep architecture, small convolutional kernels (all 3×3 or 1×1), and small pooled kernels (all 2×2) (Majib et al., 2021). To enhance weight regularization and reduce the risk of overfitting, VGG utilizes dropout regularization in the fully connected layer.

Table 5 delineates the paramount articles using the VGG network model within the selected articles, exhibiting an accuracy threshold surpassing 94%. Specifically, (Woodward et al., 2023) emerge as the epitome of precision, achieving the highest accuracy rate of 98.5% among seven scrutinized pre-trained models, including but not limited to Xception, VGG19, ResNet, and MobileNet, with VGG19 being notably distinguished for its accuracy. (Kaur et al., 2022; Winyangkun et al., 2023) employ the DeepFace model, a lightweight yet potent facial recognition framework that packages several models, such as VGG-Face, Google FaceNet, and OpenFace, for facial analysis applications. (Kaur et al., 2022) pioneered the creation of a real-time system for detection from facial images, leveraging the DeepFace model, which is proficient in discerning the facial expressions of individuals or groups. (Dudekula & Purnachand, 2023) adopt a dual-model approach, utilizing Xception and VGG19 to extract facial features and emotion recognition, and implement these methods on the NVIDIA Jeston Nano development board. This board demonstrates superior performance metrics, including uptime and actual performance, significantly outperforming the Raspberry Pi 3, allowing lower hibernation and higher throughput in learning applications.
(Bhargavi et al., 2023) integrates three distinct transfer learning architectures—VGG16, MobileNet, and ResNet50—into therapeutic robots. This integration facilitates advanced feature extraction and classification techniques, enabling these robots to utilize computer vision to identify individuals with autism spectrum disorder (ASD), primarily children. This is instrumental in formulating and establishing tailored treatment strategies, with VGG16 achieving an accuracy of 97.66%.

Table 5 delineates the paramount articles using the VGG network model within the selected articles, exhibiting an accuracy threshold surpassing 94%. Specifically, (Woodward et al., 2023) emerges as the epitome of precision, achieving the highest accuracy rate of 98.5% among seven scrutinized pretrained models, including but not limited to Xception, VGG19, ResNet, and MobileNet, with VGG19 being notably distinguished for its accuracy. (Kaur et al., 2022; Winyangkun et al., 2023) employ the DeepFace model, a lightweight yet potent facial recognition framework that packages several models, such as VGG-Face, Google FaceNet, and OpenFace, for facial analysis applications. (Kaur et al., 2022) pioneered the creation of a real-time system for detection from facial images, leveraging the DeepFace model, which is proficient in discerning the facial expressions of individuals or groups. (Dudekula & Purnachand, 2023) adopt a dual-model approach, utilizing Xception and VGG19 to extract facial features and emotion recognition, and implement these methods on the NVIDIA Jeston Nano development board. This board demonstrates superior performance metrics, including uptime and actual performance, significantly outperforming the Raspberry Pi 3, allowing lower hibernation and higher throughput in learning applications.

Table 5 Top 10 most accurate articles based on VGG.

Authors	Algorithms	Accuracy	Database	Types of Emotions	Applications
(Woodward et al., 2023)	CNN, VGG19	98.5%	EnvBodySens	N/A	Computer Science
(Dudekula & Purnachand, 2023)	VGG19, Xception	98.4%	CK+	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Kaur et al., 2022)	VGG	97.78%	LFW	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Bhargavi et al., 2023)	VGG16, MobileNet, ResNet50	97.66%	KAGGLE AUTISM	anger, disgust, fear, happiness, sadness, surprise, and neutral	ASD children, Robot
(Ilyas et al., 2023)	VGG16， ResNet50	96.77%	JAFEE, CK+	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Pu & Zhu, 2021)	CNN, VGG16	95.74%	CK+, FER2013	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Khine et al., 2021)	VGG16, K-means	95.00%	MUG	anger, disgust, fear, happiness, sadness, surprise	Computer Science
(C & Palaniswamy, 2022)	CNN, VGG16	95.00%	CMU MultiPIE	anger, happy, disgust, neutral, surprise	Computer Science
(Winyangkun et al., 2023)	Deepface, VGG-face	94.00%	FER2013	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Das et al., 2019)	CNN, VGG16	94.00%	MMI, MUG	anger, disgust, fear, happiness, sadness, surprise	Computer Science

(Bhargavi et al., 2023) integrates three distinct transfer learning architectures—VGG16, MobileNet, and ResNet50—into therapeutic robots. This integration facilitates advanced feature extraction and classification techniques, enabling these robots to utilize computer vision to identify individuals with autism spectrum disorder (ASD), primarily children. This is instrumental in formulating and establishing tailored treatment strategies, with VGG16 achieving an accuracy of 97.66%.

Furthermore, (Ilyas et al., 2023) seamlessly integrate the Viola–Jones facial detection method with the VGG16 network while ensuring a consistent input dimension of 224×224 pixels. (Pu & Zhu, 2021) leveraged the VGG16 model for a recognition classifier to enhance the design of convolutional network extensions for addressing visual network challenges. However, this study reveals that, in comparison with Inception v3 and ResNet50, VGG16 exhibits mediocre generalization capabilities. (Khine et al., 2021) delve into the fine-tuning of the VGG16 model, a process segmented into four pivotal stages: tuning, feature extraction, clustering, and comparison. This process involves the adjustment of the model's last dense layer FC7, resulting in an average accuracy of 95% when the MUG dataset is utilized. (C & Palaniswamy, 2022) selected VGG16 as the feature embedding network and subsequently employed it as the prototype network for an emotion recognition model, whereas (Das et al., 2019) utilized the VGG16 model to reduce the original 1000 classes in the model to six classes, defining six general expressions (neutral, surprise, angry, happy, sad, and fear) to lower the complexity of the VGG16 model. This modification facilitates increased facial expression recognition rates and speeds, surpassing existing methodologies in terms of accuracy and robustness.

In summary, the VGG network structure is straightforward, and enhancing performance is achievable by increasing the depth of the network. Some studies have indicated that removing the fully connected layer in a VGG network has a minimal impact on performance while significantly reducing the number of parameters. The use of a publicly available VGG for pretraining models, rather than retraining them, saves time and computational resources and eliminates the need for extensive datasets, simplifying its usage.

4.3. SVM-based FER

Vladimir Vapnik and Alexey Chervonenkis proposed the support vector machine (SVM) in 1963 (Chauhan et al., 2019). It is a binary classification model that can map the feature vector in a sample to points in space to draw a line to effectively distinguish between the two types of points (Chandra & Bedi, 2021; Cherkassky & Ma, 2004). This method should also be able to classify new points with high accuracy in the future. SVMs are well suited for solving nonlinear and high-dimensional classification problems with small- and medium-sized samples.

Table 6 delineates the findings from ten scholarly articles that have employed the support vector machine (SVM) algorithm to achieve the highest accuracy within their respective studies. (Vice et al., 2019) introduced an innovative prototype for an autonomous, real-time multimodal emotion state classification system. This system leverages both speech and facial expressions to construct a comprehensive classifier, integrating the Inception V3 neural network alongside the SVM model. Remarkably, the SVM model attained an accuracy of 99.2%, surpassing the Inception schemes, which recorded a 92.78% accuracy rate. (Vijayeeta & Pattnayak, 2022; Singh, Singh, et al., 2023) and explored the use of an SVM as a classifier, both of which focused on developing music recommendation systems. Singh et al. differentiate itself by featuring an automatic song playback functionality, whereas Prachi and Parthasarathi enhance the system's repertoire by incorporating language-specific songs. (Xie et al., 2019) presented a novel approach to early expression detection by implementing an online kernel-based structured output SVM (SOSVM), which demonstrated the ability to process sequence data within a nonlinear feature space, a significant advancement over previous methodologies.

Table 6 Top 10 most accurate articles based on SVM.

Authors	Algorithms	Accuracy	Database	Types of Emotions	Applications
(Vice et al., 2019)	SVM, Inception V3, KNN	99.20%	RAF-DB, CK+	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Singh, Singh, et al., 2023)	SVM, K-means	97.50%	Bollywood	happy, sad, excited	Computer Science
(Xie et al., 2019)	SVM, OKED	97.10%	CK+, CASIA	anger, disgust, fear, happiness, sadness, surprise	Computer Science
(Sathya et al., 2022)	SVM	95.41%	own database	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Martínez & Vega, 2023)	CNN, KNN, SVM, MLP	95.00%	CK+	happy, sadness, anger, surprise	Robot
(Siam et al., 2022)	SVM, KNN, NB, LR	94.00%	CK+, JAFFE , RAF-DB	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Vijayeeta & Pattnayak, 2022)	CNN, SVM	93.32%	N/M	anger, disgust, fear, happiness, sadness, surprise, and neutral	Recommended System
(Xu et al., 2022)	SVM, PLFD	92.00%	WLFW	EAR and Maar	Education
(Kartheek et al., 2022)	SVM	92.00%	MUG, TFEID, JAFFE, CK+, KDEF, RAF, SFEW	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science
(Wei et al., 2021)	LBP, LPQ, SVM	88.41%	JAFFE	anger, disgust, fear, happiness, sadness, surprise, and neutral	Computer Science

(Sathya et al., 2022) compared the SVM and tree classifiers with the SVM radial basis function (RBF) kernel, which exhibited superior accuracy at 95.41%. (Xie et al., 2019; Ozdemir et al., 2019) investigated the application of facial emotion detection systems in robotics. The system proposed in (Martínez & Vega, 2023) achieves real-time facial emotion detection at a rate of 13.18 fps on an embedded device (Raspberry Pi 4) via the robot operating system (ROS) noisy version. Siam et al., employing various classification algorithms, including SVM, KNN, and RF, identified SVM as yielding the highest accuracy (94%) on the CK+ datasets, indicating its exceptional detection accuracy and processing time performance. (Xu et al., 2022) constructed a method to evaluate students' learning status in English classroom settings through facial expression recognition technology based on the SVM classifier. This methodology provides educators with intelligent teaching feedback, enabling timely comprehension of students' learning conditions. (Kartheek et al., 2022) introduced the dimensionally reduced chess model (DRCP), a local texture-based image descriptor tested across different datasets via MATLAB software, and a one-vs-all (OVA) multiclass SVM, which achieved an average accuracy of 92%. Finally, (Wei et al., 2021) discussed the integration of SVM superior classification features with the texture description characteristics of the LBP, resulting in the LBP+LPQ algorithm, which demonstrated a recognition rate of 88.41%, surpassing traditional singular LBP or LPQ algorithms.

In summary, SVMs are widely utilized as classifiers in machine learning and pattern recognition. Despite being a binary classification model, SVMs can be employed for multiple classification tasks by transforming them into "n" binary classification problems. The model trained by the SVM relies entirely on the support vector. Even if all nonsupport vector points are removed from the training set and the process is repeated, the resulting model will remain unchanged.

4.4. Application areas

The application of AI-based real-time facial expression recognition (FER) extends beyond the exploration of artificial intelligence algorithms, as it is frequently used in various fields. As depicted in Figure 12, most publications on this topic (229) focused on computer science (59.3%) without being specific to any industry (Ahmed et al., 2022; Miao et al., 2019; Brintha et al., 2022; Srinivas et al., 2023; Saurav, et al. 2023; Wang et al., 2020). While developing AI algorithms and models is crucial, establishing practical relevance in specific fields or industries is equally important. For example, 51 publications applied FER to education (Fakhar et al., 2022; Zhu et al., 2022; Ruan et al., 2022; Villaroya et al., 2022; He & Qin, 2022), and 36 were dedicated to research on the Internet of Things (IoT) (Cotter, 2020; Pathak & Singh, 2020; Zhang et al., 2023; Wu et al., 2020). The number of publications for recommended systems (Bakariya et al., 2023; Zhao & Zeng, 2020), healthcare (Jain et al., 2022; Silva et al., 2021; Hou Lee & Wong, 2020), and consumer behavior (Shetty et al., 2021; Mayuri et al., 2021; Mahrab et al., 2021; Mokadam et al., 2020) was 17, 15, and 10, respectively. Figure 13 shows the annual number of publications across various disciplines between 2019 and 2023. This reveals a steady growth trend in computer science, education, and the IoT. Moreover, consumer behavior and robotics fields have shown no significant change, with a relatively low number of publications each year. The extensive application of FER in various fields is based on the distinct characteristics of different industries. AI-based facial expression recognition technology is poised to enable more effective decision-making and personalized experiences across different industries by deepening the understanding of human emotions and behaviors to promote the development of different industries

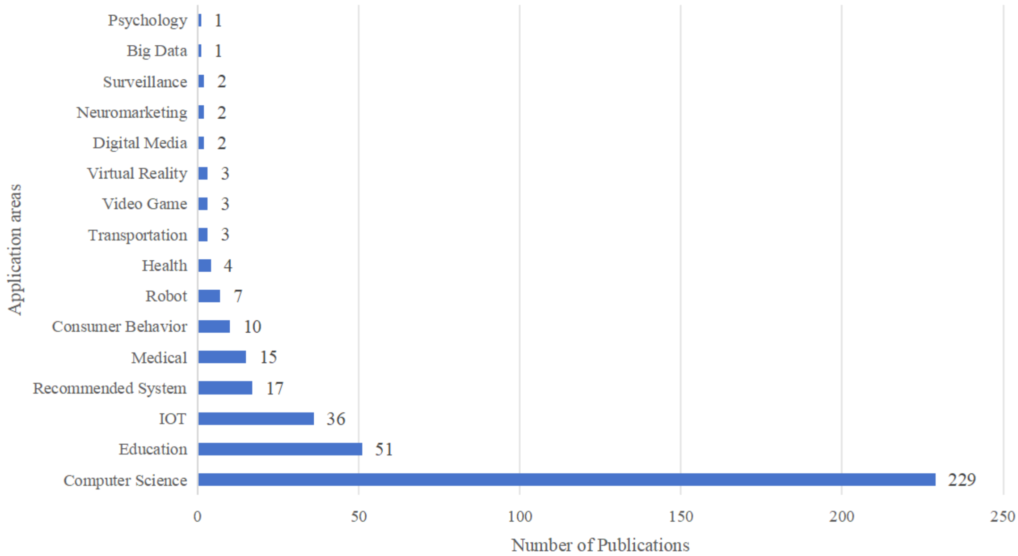
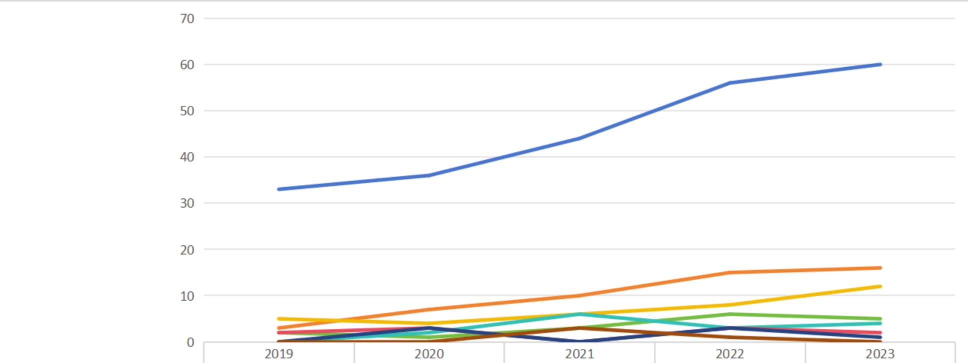


Figure 12 Distribution of publications by application area.



	2019	2020	2021	2022	2023
Computer Science	33	36	44	56	60
Education	3	7	10	15	16
IOT	5	4	6	8	12
Recommended System	2	1	3	6	5
Medical	0	2	6	3	4
Consumer Behavior	2	3	0	3	2
Robot	0	3	0	3	1
Health	0	0	3	1	0

Figure 13 Trend in publications of each application.

4.5. Wild challenge

The integration of facial expression recognition (FER) technology into practical environments has attracted increasing attention from researchers, driven by its wide-ranging potential in areas such as human–computer interaction (HCI) (Han et al., 2020), emotional assessment, and mental health monitoring. While much of the existing research has been conducted in controlled laboratory environments, the aim is to deploy FER systems in real-world scenarios. However, real-world conditions introduce numerous factors, such as lighting, head posture, and individual characteristics, which can hinder accurate facial expression recognition. Addressing these challenges is crucial for applying FER systems outside of laboratory settings.

Facial expression recognition is a data-driven task that requires extensive labelling data to train a deep network capable of capturing subtle changes associated with expressions. Furthermore, variations in facial expressions across different demographic groups, such as race, gender, age, and culture, present additional challenges for accurate recognition. Real-world scenarios involving facial occlusion and head-diverse postures also pose significant obstacles to facial expression recognition. Notably, there is a scarcity of sizeable facial expression datasets that focus on facial partial occlusion and head posture annotation. However, the EmotiNet and RAF-DB datasets are notable exceptions. A lightweight CNN model proposed by (Siddiqui et al., 2022) demonstrated promising real-time recognition performance, achieving a peak test accuracy rate of 55.09% on the AffectNet dataset, approaching human-level efficiency. Notably, many high-accuracy CNN models mentioned in Section 3.5 were developed and tested in controlled laboratory environments, and their real-world applicability may lead to a decline in accuracy. (Han et al., 2020) introduced a low-calorie network CERN, featuring only 1.45 million parameters, which is significantly fewer than the number of leading-edge models used to address this. This model requires only 17 MB of storage space and can operate at a real-time speed of 40 fps. For the real-world facial occlusion dataset FED-RO (Gao & Zhao, 2021), CERN demonstrated a recognition accuracy of 62.25%. Following the ablation experiment, it was more robust in the face of occlusion and postural changes.

In conclusion, despite the existence of several large-scale real-world facial expression databases, they all rely on single-person annotation, which is highly subjective. With advancements in technology, more facial expression datasets from wild environments are anticipated to be established. However, accurately obtaining objective expression labels remains a challenging task. Currently, facial expression recognition in the field presents significant challenges, including occlusion, illumination, head posture changes, concerns regarding the authenticity and reliability of data annotation, and the diversity and complexity of facial expression variations.

5. Discussion and Future Research Directions

Progress in the field of artificial intelligence entails the continuous enhancement of computational processes, paradigms, and frameworks that emulate human intelligence and perform duties historically contingent upon human cognition. Deep learning is the most widely employed AI technology in facial expression recognition, with prevalent algorithm models such as CNN, SVM, VGG, ResNet50, KNN, and others. Among various application fields, computer science predominates in terms of publication volume, followed by education, the Internet of Things (IoT), recommendation systems, healthcare, consumer behavior, and robotics. This insight was obtained through a review of 386 articles published between January 2019 and December 2023.

According to the findings of this review, a notable escalation in scholarly contributions was documented, with a sustained upward trend until December 2023. Recently, the application of artificial intelligence has led to the development of immediate facial expression analysis systems, which predominantly employ sophisticated machine learning approaches such as deep neural networks. The majority of studies focus on CNN (Hashan et al., 2023; Huang et al., 2022), VGG (Yang et al., 2022; Kaur & Kulkarni, 2023), SVM (Nandi et al., 2020; Rathour et al., 2021; Yu & Tapus, 2020), and ResNet50 (Gupta et al., 2023; Hu et al., 2019; Muhammad et al., 2021) algorithms, with the CNN algorithm being the most commonly used model for building real-time facial expression recognition systems. However, researchers tend to choose the VGG model with limited data because of its more straightforward network structure and fewer parameters. The selected articles reveal that there are two primary methods: the classical method and the neural network method based on deep learning. Most researchers use the Viola–Jones facial detection algorithm (Aiswarya et al., 2020; LokeshNaik et al., 2023) combined with recognizers capable of interpreting feature vectors, such as support vector machines (SVMs), naive Bayes, and K-means algorithms. The SVM is the classifier commonly used in traditional methods for constructing facial expression recognition systems, enabling decision-making on the basis of image output classes. On the other hand, research based on the neural network method involves an input layer containing the discretization or specific features of the facial image. The hidden layer is responsible for feature extraction and learning and employs techniques such as principal component analysis (PCA), linear discriminant analysis (LDA), and t-distributed stochastic neighbor embedding (t-SNE). Researchers often enhance performance through model pruning, knowledge distillation, data augmentation, and model compression.

A total of 293 conference papers identified through publication statistics were included in the selected papers. Among these, the International Conference on Multimedia Interaction (ICMI) has the most significant number of papers, with five publications. Among the 59 journals in the selected papers, the top two most prolific publishers are IEEE Access and Multimedia

Tools and Applications, with IEEE Access leading in publication volume at 15. When the country-based distribution of publications was analyzed, it was found that Asian countries contributed 68% of the publications in this area. India leads with a maximum count of 147 scholarly outputs, succeeded by China, which has registered 86 publications. Despite Russia and Saudi Arabia having only six publications each, their citations are 119 and 93, respectively, indicating recognition of their articles by their peers. In terms of facial expression datasets, 81 databases were found during the review, with the FER2013 database being the most commonly used (176 times). This is likely due to FER2013 having many samples (35,887 images) and being prepared for the Kaggle competition, which enjoyed higher popularity and better training results. Following this is the CK+ dataset (94 times) and the JAFFE dataset (39 times), detailed in Section 3.4 for datasets that are used more than five times. Many studies enhance their datasets by employing various techniques, including geometric transformation classes (flip, crop, deformation, scaling, etc.), color transformation classes (noise, blur, color transformation, erase, fill, etc.), and multiple data augmentation methods, such as synthetic minority oversampling techniques (SMOTE), sample pairing, mixup, etc.). Unsupervised data augmentation is primarily based on generative adversarial networks (GANs). These data augmentation methods effectively address the issue of small data volume or unbalanced data distribution.

The analysis reveals that current research on facial expression recognition via artificial intelligence technology is divided into two key clusters: the initial focus is on enhancing artificial intelligence methodologies within the realm of computer science, with keywords such as "CNN," "edge computing," "computational modelling," and "feature extraction,", etc. The second cluster pertains to the application of FER in diverse fields, with keywords such as "student engagement," "Raspberry Pi," "ASD children," and "customer sentiment,", etc., mainly in the fields of computer science, education, Internet of Things, recommendation systems, medical care, consumer behavior, and robotics. These clusters provide insight for researchers into potential future developments in the field.

Through a comprehensive systematic literature review coupled with bibliometric analysis, an in-depth examination of 386 scholarly articles was conducted. This rigorous analysis has identified several prevailing research gaps within the literature. Consequently, the following future research directions are proposed:

(1) Multimodal fusion: This approach involves blending facial expression analysis with complementary modalities, including body language, vocal cues, and contextual meaning, thereby enhancing the precision and reliability of emotion detection through facial cues. Although few studies have developed multimodal emotion recognition, the potential benefits are significant. However, implementing multimodal emotion recognition typically requires expensive equipment, such as EEG acquisition and analysis systems and high-precision eye tracking systems. Despite the cost, multimodal fusion can significantly enhance the system's ability to understand and interpret human emotions, such as feature-level, decision-level, and hybrid emotions, resulting in a more intelligent and natural interactive experience. Therefore, further exploration of the use of nonvisual sensors, such as physiological signals and sounds, to aid facial expression recognition, particularly in situations involving changes in illumination and partial facial occlusion, is essential. Integrating information from multiple sources can significantly improve facial expression recognition accuracy.

(2) Computational modelling and real-time systems: Emotion recognition relies mostly on computational modelling and real-time systems. With an increasing number of studies focusing on video-based emotion recognition, researchers are improving models to accommodate the processing of long-sequence time series facial dynamics. This is essential for capturing the dynamics and trends of emotional changes. Only a few papers have utilized time series modelling methods such as attention mechanisms, long and short memory networks, and recurrent neural networks for facial expression recognition. Addressing the challenge of processing long-term facial dynamics remains a significant focus for future studies.

(3) Personalized emotion recognition and transfer learning: The current use of artificial intelligence technologies, such as CNNs, involves training and testing specific databases for accurate facial expression recognition. However, personalized models for facial expression recognition are lacking. Considering the impact of individual differences and cultural backgrounds on facial expressions, personalized models can better capture individual emotional characteristics and preferences, leading to more accurate emotional recognition and interactive experiences. Transfer learning and incremental learning methods are recommended to accelerate the learning and adaptation of new tasks by leveraging previously acquired knowledge and experience. These methods can enhance facial expression recognition systems' generalizability and portability, allowing them to adapt better to various individual differences, cultural contexts, and environments.

(4) Interdisciplinary collaboration and social impact: On the basis of the analysis of selected articles, while most of them focus on computer science, numerous articles explore the application of facial expression recognition (FER) in diverse fields, such as education, the Internet of Things, recommendation systems, and healthcare. Integrating cognitive and neural mechanisms from psychology and neuroscience is instrumental in comprehending the correlations between facial expressions and emotion, cognition, and behavior. The incorporation of principles from biomedical engineering can advance biometric detection technology, thereby refining the fidelity and precision of systems designed for facial expression recognition (FER) systems. Collaboration with the medical field can facilitate the utilization of FER systems in medical diagnosis and assessment of mental health, particularly in ASD children; this entails merging clinical data with real cases to validate the effectiveness and reliability of FER systems. Furthermore, integration with human–computer interaction and user experience design, among

other fields, is also feasible. Therefore, fostering interdisciplinary collaboration in facial expression recognition necessitates the amalgamation of expertise and competencies from various domains, which will, in turn, bolster the development of more comprehensive and accurate FER technology.

6. Conclusion and Limitations

With the rise of AI-based technology and its broad applicability across various fields and disciplines, this research aims to map and summarize publications on AI-based facial expression recognition. It also proposes insightful research approaches and directions for future studies. This study employs bibliometric analysis and systematic literature review methods to identify the significantly contributing journals, conferences, countries, and the most influential papers. It describes the trends in the number of publications, the development prospects of research directions, the most commonly used artificial intelligence technologies in the field of FER, and their applications in various fields.

Within the scope of this research, a bibliometric analysis and systematic literature review of the field of real-time emotion recognition based on artificial intelligence technologies were conducted, covering the period from January 2019 to December 2023. The evaluation method illustrated in this paper is founded on an initial pool of 2666 articles, of which 386 were chosen for a systematic review. The objective of this in-depth scholarly assessment is to provide a holistic view of facial expression recognition for professional researchers within the discipline, facilitating the enhancement of their systems or the application of knowledge to various research areas, ultimately benefiting future professionals. The study considered articles from the Web of Science, IEEE Xplore, Scopus, and ACM Digital Library databases, potentially overlooking interesting articles in other databases.

Addressing the research questions posed in this work, the following conclusions can be drawn:

To address the first research question, it is essential to highlight the five most frequently utilized databases most prominently leveraged within the realm of facial expression recognition (FER): FER2013 (176 times), CK+ (94 times), JAFFE (39 times), AffectNet (24 times), and RAF-DB (23 times). In real-time emotion classification systems, the main artificial intelligence technique employed is convolutional neural networks (CNNs), followed by support vector machines (SVMs). Other notable methodologies include YOLO algorithms.

In addressing the second research question, this study reveals that convolutional neural networks (CNNs) within deep learning architectures are particularly favored by researchers in the field of facial expression recognition. This preference is attributed to the remarkable performance of CNNs in image processing, characterized by advantages such as local connectivity, parameter sharing, and hierarchical feature learning. Among them, ResNet50 and VGG16 exhibit outstanding performance in facial expression recognition tasks because of their depth, feature extraction capabilities, and adaptability to complex expressions. Numerous researchers have applied these two algorithms to publicly available datasets, achieving accuracy rates that consistently exceed 80%. Although ResNet50 may not be one of the most frequently cited algorithms in this review, its performance is remarkable, likely because of the use of residual connections within its architecture, which enhances the network's capability to learn more intricate and abstract feature representations. This capability is particularly crucial for effectively capturing the subtle nuances of complex emotional variations in facial expression recognition tasks, resulting in ResNet50 being one of the most promising algorithms in this domain.

Identifying the most frequently used keywords and research trends over the past five years constitutes the third research question proposed in this work. The top five keywords in terms of frequency are neural networks, deep learning, facial expression recognition, humanistic, and feature extraction. Since 2019, the most prevalent keywords have included neural networks, support vector machines, real-time methods, object detection, and feature extraction. By 2023, there has been a noticeable shift toward keywords such as transfer learning, children with autism spectrum disorder (ASD), real-time systems, recursion, Raspberry Pi, Internet of Things (IoT), computational modelling, and online learning. This shift underscores an escalating emphasis on harnessing facial expression recognition (FER) across a spectrum of domains, including the IoT, education, healthcare, and autonomous driving, where real-time systems are being developed to recognize facial expressions in diverse contexts. Concurrently, researchers are increasingly emphasizing the integration of additional physiological signals (such as EGG and heart rate) and contextual information (including speech, text, and environmental factors) to increase the accuracy of emotion recognition and understanding.

Considering the fourth research question, current facial expression classification systems, particularly those utilizing deep learning approaches such as the CNN, ResNet, and VGG models, have demonstrated the ability to achieve high accuracy on several publicly available datasets, effectively recognizing seven basic emotion categories: happiness, surprise, neutrality, sadness, anger, disgust, and fear. Additionally, some classification systems meet the requirements for real-time applications, enabling rapid emotion recognition within video streams. This capability is crucial for deployment in real-time interactive scenarios, such as virtual reality, gaming, classrooms, and mental health monitoring. Moreover, these systems have significant potential for widespread application in many areas, such as sentiment analysis, market research (e.g., consumer emotion analysis), and social media sentiment detection.

To address the fifth research question posed in this paper, it is acknowledged that while current facial expression classification systems can achieve high accuracy, they are prone to overfitting. In practical applications, the classification accuracy of these systems can be easily affected by various factors, including changes in illumination, micro variations in expressions, viewing angles, and facial occlusions. Furthermore, real-time classification systems often rely on high-performance computing devices and resources, such as GPUs, presenting challenges for deployment on certain mobile devices. An important challenge for the advancement of facial expression recognition technology is how to utilize these techniques without the invasion of personal privacy. Additionally, concerning the accuracy of training data, publicly available datasets still present limitations such as imbalanced data distributions or insufficient diversity, resulting in significant discrepancies in model performance across different populations and cultural backgrounds.

Finally, to address the final research question, this study presents the following findings. In terms of recognition speed, YOLO, as a real-time object detection algorithm, has significant advantages in facial expression classification. YOLO boasts an exceptionally high processing speed, enabling rapid image processing within a minimal timeframe. Additionally, YOLO leverages spatial information encoding to make multiple predictions, thus enhancing its detection accuracy. Conversely, with respect to ResNet50, this analysis reveals that papers employing ResNet50 generally report high accuracy rates, indicating its efficacy in facial expression recognition. However, out of the 386 articles reviewed in this study, only 12 utilized the ResNet algorithm, whereas 13 employed various YOLO algorithms. Among these, 7 articles utilized YOLOv5, 3 articles utilized YOLOv3, and only one article referenced each of the earlier versions (YOLOv1, YOLOv4, and YOLOv8).

In summary, both the YOLO and ResNet algorithms have achieved remarkable performance in facial expression recognition in terms of speed and accuracy. Future research can delve deeper into the applications of YOLO and ResNet and explore the potential of integrating these two algorithms. By fully using the rapid detection capabilities of YOLO alongside the deep feature extraction strength of ResNet50, it may be possible to engineer a system for real-time facial expression analysis that combines both rapid processing and high accuracy.

Acknowledgement

We acknowledge Professor João Alexandre Lobo Marques for his meticulous guidance and selfless support throughout the entire research process. Professor João Alexandre Lobo Marques' expertise and invaluable advice have provided a solid foundation for this study. We also acknowledge Professor Auzuir Ripardo de Alexandria for his thorough review and the professional feedback provided, which greatly enhanced the quality of this article.

Ethical Considerations

Not applicable.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Funding

This research did not receive any financial support.

References

Adhikary, A., & Bhandari, N. (2021). PosEmotion—Combining real-time 2D body pose estimation and facial emotion recognition to analyze human behavior. *2021 26th International Conference on Automation and Computing (ICAC)* (pp. 1–6). https://doi.org/10.23919/ICAC50006.2021.9594155

Ahmed, T. U., Hossain, S., Hossain, M. S., Islam, R. U., & Andersson, K. (2022). A deep learning approach with data augmentation to recognize facial expressions in real time. (pp. 487-500). Springer. https://doi.org/10.1007/978-981-16-7597-3_40

Aifanti, N., Papachristou, C., & Delopoulos, A. (2010). The MUG facial expression database. *11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10* (pp. 1-4).

Aiswarya, P., Manish, & Mangalraj, P. (2020). Emotion recognition by inclusion of age and gender parameters with a novel hierarchical approach using deep learning. *2020 Advanced Communication Technologies and Signal Processing (ACTS)* (pp. 1-6). https://doi.org/10.1109/ACTS49415.2020.9350479

Alhajlah, M. (2023). A novel efficient patient monitoring FER system using optimal DL-features. *Computers, Materials and Continua, 74*(3), 6161–6175. https://doi.org/10.32604/cmc.2023.032505

Alisawi, M., & Yalçın, N. (2023). Real-time emotion recognition using deep learning methods: Systematic review. *Intelligent Methods in Engineering Sciences, 2*(1), 5–21. https://doi.org/10.58190/imiens.2023.7

Alreshidi, A., & Ullah, M. (2020). Facial emotion recognition using hybrid features. *Informatics, 7*(1). https://doi.org/10.3390/informatics7010006

Arabian, H., Wagner-Hartl, V., & Moeller, K. (2021). Traditional versus neural network classification methods for facial emotion recognition. *Current Directions in Biomedical Engineering, 7*(2), 203-206. https://doi.org/10.1515/cdbme-2021-2052

Ashraf, A., Gunawan, T. S., Arifin, F., Kartiwi, M., Sophian, A., & Habaebi, M. H. (2023). Enhanced emotion recognition in videos: A convolutional neural network strategy for human facial expression detection and classification. *Indonesian Journal of Electrical Engineering and Informatics, 11*(1), 286–299. https://doi.org/10.52549/ijeei.v11i1.4449

Awais, M., Raza, M., Singh, N., Bashir, K., Manzoor, U., Islam, S. U., & Rodrigues, J. J. P. C. (2021). LSTM-based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19. *IEEE Internet of Things Journal, 8*(23), 16863–16871. https://doi.org/10.1109/JIOT.2020.3044031

Bakariya, B., Singh, A., Singh, H., Raju, P., Rajpoot, R., & Mohbey, K. K. (2023). Facial emotion recognition and music recommendation system using CNN-based deep learning techniques. *Evolving Systems*. https://doi.org/10.1007/s12530-023-09506-z

Baltrusaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Barron-Estrada, M. L., Zatarain-Cabada, R., & Oramas-Bustillos, R. (2019). Emotion recognition for education using sentiment analysis. *Research in Computing Science, 148*(5), 71–80. https://doi.org/10.13053/rcs-148-5-8

Bhardwaj, P., Gupta, P. K., Panwar, H., Siddiqui, M. K., Morales-Menendez, R., & Bhaik, A. (2021). Application of deep learning on student engagement in e-learning environments. *Computers and Electrical Engineering, 93*, 107277. https://doi.org/10.1016/j.compeleceng.2021.107277

Bhargavi, Y., D, B., & Prince, S. (2023). AI-based emotion therapy bot for children with autism spectrum disorder (ASD). *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)* (Vol. 1, pp. 1895-1899). https://doi.org/10.1109/ICACCS57279.2023.10112868

Brintha, N. C., Narayana, J. A., Jaswanth, G. L. V. S., Chandrapal, G. J., & Venkat, D. (2022). Realtime facial emotion detection using machine learning. *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)* (pp. 1-5). https://doi.org/10.1109/ICSES55317.2022.9914318

C, V., & Palaniswamy, S. (2022). Emotion recognition at real-time applications using meta-learning. *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)* (pp. 1-6). https://doi.org/10.1109/GCAT55367.2022.9971841

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods, 40*(1), 109–115. https://doi.org/10.3758/BRM.40.1.109

Canedo, D., & Neves, A. J. R. (2019). Facial expression recognition using computer vision: A systematic review. *Applied Sciences, 9*(21), 4678. https://doi.org/10.3390/app9214678

Chaiyarak, S., Nilsook, P., & Wannapiroon, P. (2021). An empirical study of intelligent virtual universal learning platforms. *2021 Research, Invention, and Innovation Congress: Innovation Electricals and Electronics (RI2C)* (pp. 66-73). https://doi.org/10.1109/RI2C51727.2021.9559785

Chandra, M. A., & Bedi, S. S. (2021). Survey on SVM and their application in image classification. *International Journal of Information Technology, 13*(5), 1-11. https://doi.org/10.1007/s41870-017-0080-1

Chauhan, V. K., Dahiya, K., & Sharma, A. (2019). Problem formulations and solvers in linear SVM: A review. *Artificial Intelligence Review, 52*(2), 803-855. https://doi.org/10.1007/s10462-018-9614-6

Chen, C. (2017). Science mapping: A systematic review of the literature. *Journal of Data and Information Science, 2*(2), 1-40. https://doi.org/10.1515/jdis-2017-0006

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks, 17*(1), 113-126. https://doi.org/10.1016/S0893-6080(03)00169-2

Cotter, S. (2020). Low complexity deep learning for mobile face expression recognition. *Proceedings of the 3rd International Conference on Vision, Image and Signal Processing*. https://doi.org/10.1145/3387168.3387175

Das, P. J., Talukdar, A. K., & Sarma, K. K. (2019). A framework for human behaviour detection using combined analysis of facial expression and eye gaze. *2019 International Conference on Image and Signal Processing for Communication (IESPC)* (pp. 154-160). https://doi.org/10.1109/IESPC.2019.8902367

De Ocampo, A. L. P. (2023). Haar-CNN cascade for facial expression recognition. *2023 International Electrical Engineering Congress (iEECON)* (pp. 89-92). https://doi.org/10.1109/iEECON56657.2023.10126902

Dhall, A. (2019). EmotiW 2019: Automatic emotion, engagement and cohesion prediction tasks. *2019 International Conference on Multimodal Interaction* (pp. 546-550). https://doi.org/10.1145/3340555.3355710

Dhope, P., & Neelagar, M. B. (2022). Real-time emotion recognition from facial expressions using artificial intelligence. *2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP)* (pp. 1-6). https://doi.org/10.1109/AISP53593.2022.9760654

Dudekula, U., & Purnachand, N. (2023). Analysis of facial emotion recognition rate for real-time application using NVIDIA Jetson Nano in deep learning models. *Indonesian Journal of Electrical Engineering and Computer Science, 30*(1), 598-605. https://doi.org/10.11591/ijeecs.v30.i1.pp598-605

Dwijayanti, S., Iqbal, M., & Suprapto, B. Y. (2022). Real-time implementation of face recognition and emotion recognition in a humanoid robot using a convolutional neural network. *IEEE Access, 10*, 89876-89886. https://doi.org/10.1109/ACCESS.2022.3200762

Eck, N. J. V., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology, 60*(8), 1635-1651. https://doi.org/10.1002/asi.21075

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation, 19*, 207-283.

Fakhar, S., Baber, J., Bazai, S. U., Marjan, S., Jasinski, M., Jasinska, E., Chaudhry, M. U., Leonowicz, Z., & Hussain, S. (2022). Smart classroom monitoring using novel real-time facial expression recognition system. *Applied Sciences, 12*(23), 12134. https://doi.org/10.3390/app122312134

Feak, C., & Swales, J. (2009). *Telling a research story: Writing a literature review*. University of Michigan Press/ELT. https://doi.org/10.3998/mpub.309338

Fernandes, J. V. M. R., Alexandria, A. R. D., Marques, J. A. L., Assis, D. F. D., Motta, P. C., & Silva, B. R. D. S. (2024). Emotion detection from EEG signals using machine deep learning models. *Bioengineering, 11*(8), 782. https://doi.org/10.3390/bioengineering11080782

Filipovic, F., Despotovic-Zrakic, M., Radenkovic, B., Jovanic, B., & Živojinovic, L. (2019). An application of artificial intelligence for detecting emotions in neuromarketing. *2019 International Conference on Artificial Intelligence: Applications and Innovations (IC-AIAI)* (pp. 49-494). https://doi.org/10.1109/IC-AIAI48757.2019.00016

Gao, J., & Zhao, Y. (2021). TFE: A transformer architecture for occlusion aware facial expression recognition. *Frontiers in Neurorobotics, 15*, 763100. https://doi.org/10.3389/fnbot.2021.763100

Goel, A. K., Jain, A., Saini, C., Ashutosh, Das, R., & Deep, A. (2022). Implementation of AI/ML for human emotion detection using facial recognition. *2022 IEEE 4th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)* (pp. 511–515). https://doi.org/10.1109/ICCCMLA56841.2022.9989091

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., ... & Bengio, Y. (2015). Challenges in representation learning: A report on three machine learning contests. *Neural Networks, 64*, 59-63. https://doi.org/10.1016/j.neunet.2014.09.005

Gupta, S., Kumar, P., & Tekchandani, R. K. (2023). Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. *Multimedia Tools and Applications, 82*(8), 11365-11394. https://doi.org/10.1007/s11042-022-13558-9

H. Zhu, P. Hu, X. Tang, & D. Xia. (2022). NAGNet: A Convolutional Neural Network for Real-Time Sentiment Analysis of Students. 2022 12th International Conference on Information Technology in Medicine and Education (ITME), 75–80. https://doi.org/10.1109/ITME56794.2022.00027

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell Systematic Reviews, 18(2), e1230. https://doi.org/10.1002/cl2.1230

Han, B., Yun, W.-H., Yoo, J.-H., & Kim, W. H. (2020). Toward Unbiased Facial Expression Recognition in the Wild via Cross-Dataset Adaptation. IEEE Access, 8, 159172–159181. https://doi.org/10.1109/ACCESS.2020.3018738

Hart, C. (2011). Doing a literature review: Releasing the social science research imagination (Repr.). Sage.

Hashan, A. M., Adnan Adhab, K. A.-S., Islam, R. M. R. U., Avinash, K., & Dey, S. (2023). Automated human facial emotion recognition system using depthwise separable convolutional neural network. *2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)* (pp. 113–117). https://doi.org/10.1109/IAICT59002.2023.10205785

Hashan, A. M., Adnan Adhab, K. A.-S., Islam, R. M. R. U., Avinash, K., & Dey, S. (2023). Automated human facial emotion recognition system using depthwise separable convolutional neural network. *2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)* (pp. 113-117). https://doi.org/10.1109/IAICT59002.2023.10205785

Hatem, A. S., & Al-Bakry, A. M. (2022). Real-time human emotion recognition using transfer learning. *2022 International Conference on Data Science and Intelligent Computing (ICDSIC)* (pp. 224–227). https://doi.org/10.1109/ICDSIC56987.2022.10076032

He, Z., & Qin, X. (2022). Analysis of facial expressions in class based on lightweight convolutional neural network. *2022 International Conference on Industrial Automation, Robotics and Control Engineering (IARCE)* (pp. 68-74). https://doi.org/10.1109/IARCE57187.2022.00023

Hu, H., Zhu, Y., Zhang, Y., Zhou, Q., Feng, Y., & Tan, G. (2019). Comprehensive driver state recognition based on deep learning and PERCLOS criterion. *2019 IEEE 19th International Conference on Communication Technology (ICCT)* (pp. 1678-1682). https://doi.org/10.1109/ICCT46805.2019.8947282

Huang, M., Zhang, X., Lan, X., Wang, H., & Tang, Y. (2022). Convolution by multiplication: Accelerated two-stream Fourier domain convolutional neural network for facial expression recognition. *IEEE Transactions on Circuits and Systems for Video Technology, 32*(3), 1431-1442. https://doi.org/10.1109/TCSVT.2021.3073558

Hussain, S. A., & Salim Abdallah Al Balushi, A. (2020). A real time face emotion classification and recognition using deep learning model. Journal of Physics: Conference Series, 1432(1), 012087. https://doi.org/10.1088/1742-6596/1432/1/012087

Ilyas, B. R., Abderrazak, T. A., Sofiane, B. M., Bahidja, B., Imane, H., & Miloud, K. (2023). A robust-facial expressions recognition system using deep learning architectures. *2023 International Conference on Data Analytics for Business and Industry (DASA)* (pp. 541–546). https://doi.org/10.1109/DASA59624.2023.10286798

J. Kaur, J. Saxena, J. Shah, Fahad, & S. P. Yadav. (2022). Facial Emotion Recognition. 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), 528–533. https://doi.org/10.1109/CISES54857.2022.9844366

J. L. Silva, I. Oliveira, Z. Topolniak, & A. B. Alvarez. (2021). A CNN Approach Implemented to Emotional Facial Expression Recognition for the Prevention of Autistic Meltdowns. 2021 2nd Sustainable Cities Latin America Conference (SCLA), 1–6. https://doi.org/10.1109/SCLA53004.2021.9540183

J. R. Hou Lee & A. Wong. (2020). TimeConvNets: A Deep Time Windowed Convolution Neural Network Design for Real-time Video Facial Expression Recognition. 2020 17th Conference on Computer and Robot Vision (CRV), 9–16. https://doi.org/10.1109/CRV50864.2020.00010

J. Vice, M. M. Khan, & S. Yanushkevich. (2019). Multimodal Models for Contextual Affect Assessment in Real-Time. 2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI), 87–92. https://doi.org/10.1109/CogMI48466.2019.00020

J. Yang, T. Qian, F. Zhang, & S. U. Khan. (2021). Real-Time Facial Expression Recognition Based on Edge Computing. IEEE Access, 9, 76178–76190. https://doi.org/10.1109/ACCESS.2021.3082641

J. Zhu, Y. Wang, R. La, J. Zhan, J. Niu, S. Zeng, & X. Hu. (2019). Multimodal Mild Depression Recognition Based on EEG-EM Synchronization Acquisition Network. IEEE Access, 7, 28196–28210. https://doi.org/10.1109/ACCESS.2019.2901950

Jaffar, S. S., & Abdulbaqi, H. A. (2022). Facial expression recognition in static images for autism children using CNN approaches. *2022 Fifth College of Science International Conference of Recent Trends in Information Technology (CSCTIT)* (pp. 202-207). https://doi.org/10.1109/CSCTIT56299.2022.10145689

Jain, A., Verma, R., Khokhar, G. S., & Bhadauria, M. (2022). Akshi: An assistance system for visually challenged using machine learning. *2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST)* (pp. 1–6). https://doi.org/10.1109/AIST55798.2022.10064996

Jaiswal, S., & Nandi, G. C. (2020). Robust real-time emotion detection system using CNN architecture. Neural Computing and Applications, 32(15), 11253–11262. Scopus. https://doi.org/10.1007/s00521-019-04564-4

Jeong, M., Nam, J., & Ko, B. C. (2020). Lightweight multilayer random forests for monitoring driver emotional status. *IEEE Access, 8*, 60344-60354. https://doi.org/10.1109/ACCESS.2020.2983202

Jesson, J., Matheson, L., & Lacey, F. M. (2011). Doing Your Literature Review: Traditional and Systematic Techniques. SAGE Publications.

K, S. V., & Thripurala, S. (2023). Real-time facial emotion detection system using multimodal fusion deep learning architecture. *2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM)* (pp. 1-6). https://doi.org/10.1109/ELEXCOM58812.2023.10370457

K. Mayuri, N. V. Krishna Rao, N. Jayanthi, T. AlakanandaKasam, G. Nalla, & S. Jaggavarapu. (2021). Understanding Customer Reviews using Facial Expression Recognition System. 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 1482–1486. https://doi.org/10.1109/ICECA52323.2021.9676126

Kabakus, A. T. (2020). PyFER: A facial expression recognizer based on convolutional neural networks. *IEEE Access, 8*, 142243–142249. https://doi.org/10.1109/ACCESS.2020.3012703

Kanade, T., Cohn, J. F., & Yingli Tian. (2000). Comprehensive database for facial expression analysis. Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), 46–53. https://doi.org/10.1109/AFGR.2000.840611

Karnati, M., Seal, A., Yazidi, A., & Krejcar, O. (2022). FLEPNet: Feature level ensemble parallel network for facial expression recognition. *IEEE Transactions on Affective Computing, 13*(4), 2058-2070. https://doi.org/10.1109/TAFFC.2022.3208309

Kartheek, M. N., Prasad, M. V. N. K., & Bhukya, R. (2022). DRCP: Dimensionality Reduced Chess Pattern for Person Independent Facial Expression Recognition. In INTERNATIONAL JOURNAL OF PATTERN RECOGNITION AND ARTIFICIAL INTELLIGENCE (Vol. 36, Issue 11). WORLD SCIENTIFIC PUBL CO PTE LTD. https://doi.org/10.1142/S021800142256016X

Kaur, S., & Kulkarni, N. (2023). FERFM: An Enhanced Facial Emotion Recognition System Using Fine-tuned MobileNetV2 Architecture. IETE Journal of Research. Scopus. https://doi.org/10.1080/03772063.2023.2202158

Khan, M., Hariharasitaraman, S., Joshi, S., Jain, V., Ramanan, M., SampathKumar, A., & Elngar, A. A. (2022). A deep learning approach for facial emotions recognition using principal component analysis and neural network techniques. Photogrammetric Record, 37(180), 435–452. Scopus. https://doi.org/10.1111/phor.12426

Khine, W. S. S., Siritanawan, P., & Kotani, K. (2021). Automatic peak frame selection from dynamic facial expressions. *2021 60th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)* (pp. 1088-1093).

Kim, Y., Soyata, T., & Behnagh, R. F. (2018). Towards Emotionally Aware AI Smart Classroom: Current Issues and Directions for Engineering and Education. IEEE Access, 6, 5308–5331. https://doi.org/10.1109/ACCESS.2018.2791861

Ko, B. (2018). A Brief Review of Facial Emotion Recognition Based on Visual Information. Sensors, 18(2), 401. https://doi.org/10.3390/s18020401

Kong, Y., Ren, Z., Zhang, K., Zhang, S., Ni, Q., & Han, J. (2021). Lightweight facial expression recognition method based on attention mechanism and key region fusion. Journal of Electronic Imaging, 30(6). Scopus. https://doi.org/10.1117/1.JEI.30.6.063002

Kossaifi, J., Tzimiropoulos, G., Todorovic, S., & Pantic, M. (2017). AFEW-VA database for valence and arousal estimation in-the-wild. Image and Vision Computing, 65, 23–36. https://doi.org/10.1016/j.imavis.2017.02.001

L. Liakopoulos, N. Stagakis, E. I. Zacharaki, & K. Moustakas. (2021). CNN-based stress and emotion recognition in ambulatory settings. 2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA), 1–8. https://doi.org/10.1109/IISA52424.2021.9555508

L. Xie, J. Zhao, H. Wei, K. Zhang, & G. Pang. (2019). Online Kernel-Based Structured Output SVM for Early Expression Detection. IEEE Signal Processing Letters, 26(9), 1305–1309. https://doi.org/10.1109/LSP.2019.2929713

L. Zahara, P. Musa, E. Prasetyo Wibowo, I. Karim, & S. Bahri Musa. (2020). The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi. 2020 Fifth International Conference on Informatics and Computing (ICIC), 1–9. https://doi.org/10.1109/ICIC50835.2020.9288560

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278-2324. https://doi.org/10.1109/5.726791

Lee, J. R., Wang, L., & Wong, A. (2021). EmotionNet Nano: An efficient deep convolutional neural network design for real-time facial expression recognition. *Frontiers in Artificial Intelligence, 3*, 609673. https://doi.org/10.3389/frai.2020.609673

Li, S., Deng, W., & Du, J. (2017). Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2584-2593). https://doi.org/10.1109/CVPR.2017.277

Lievrouw, L. A. (1989). The invisible college reconsidered: Bibliometrics and the development of scientific communication theory. *Communication Research, 16*(5), 615-628. https://doi.org/10.1177/009365089016005004

Lin, K.-Y., Gamboa-Gonzalez, A., & Wehner, M. (2021). Soft robotic sensing, proprioception via cable and microfluidic transmission. *Electronics, 10*(24), 3166. https://doi.org/10.3390/electronics10243166

Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE, 13*(5), e0196391. https://doi.org/10.1371/journal.pone.0196391

LokeshNaik, S. K., Punitha, A., Vijayakarthik, P., Kiran, A., Dhangar, A. N., Reddy, B. J., & Sudheeksha, M. (2023). Real time facial emotion recognition using deep learning and CNN. *2023 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-5). https://doi.org/10.1109/ICCCI56745.2023.10128259

Lopes, A. T., de Aguiar, E., de Souza, A. F., & Oliveira-Santos, T. (2017). Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order. *Pattern Recognition, 61*, 610-628. https://doi.org/10.1016/j.patcog.2016.07.026

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops* (pp. 94-101). https://doi.org/10.1109/CVPRW.2010.5543262

Lyons, M., Kamachi, M., & Gyoba, J. (1998). *The Japanese Female Facial Expression (JAFFE) Dataset* [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.3451523

Mahrab, N., Salim, S. A., Ali, A. I., Mim, I. J., & Khan, R. (2021). Facial expression based automated restaurant food review system using CNN. *2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)* (pp. 1-4). https://doi.org/10.1109/IICAIET51634.2021.9573899

Majib, M. S., Rahman, M. M., Sazzad, T. M. S., Khan, N. I., & Dey, S. K. (2021). VGG-SCNet: A VGG net-based deep learning framework for brain tumor detection on MRI images. *IEEE Access, 9*, 116942-116952. https://doi.org/10.1109/ACCESS.2021.3105874

Marques, J. A. L., Neto, A. C., Silva, S. C., & Bigne, E. (2024). Predicting consumer ad preferences: Leveraging a machine learning approach for EDA and FEA neurophysiological metrics. *Psychology & Marketing, 41*(3). https://doi.org/10.1002/mar.22118

Martínez, J., & Vega, J. (2023). ROS system facial emotion detection using machine learning for a low-cost robot based on Raspberry Pi. *Electronics, 12*(1), 90. https://doi.org/10.3390/electronics12010090

Melinte, D. O., & Vladareanu, L. (2020). Facial expressions recognition for human-robot interaction using deep convolutional neural networks with rectified Adam optimizer. *Sensors, 20*(8), 2393. https://doi.org/10.3390/s20082393

Metgud, P., Naik, N. D., S, S. M., & Prasad, A. S. (2022). Real-time student emotion and performance analysis. *2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* (pp. 1-5). https://doi.org/10.1109/CONECCT55679.2022.9865114

Miao, Y., Dong, H., Jaam, J. M. A., & Saddik, A. E. (2019). A deep learning system for recognizing facial expression in real-time. *ACM Transactions on Multimedia Computing, Communications, and Applications, 15*(2). https://doi.org/10.1145/3311747

Mikhaylevskiy, S., Chernyavskiy, V., Pavlishen, V., Romanova, I., & Solovyev, R. (2021). Fast emotion recognition neural network for IoT devices. *2021 International Seminar on Electron Devices Design and Production (SED)* (pp. 1-6). https://doi.org/10.1109/SED51197.2021.9444517

Mokadam, P., Kulkarni, T., & Mulla, N. (2020). Customer reaction analysis using convolutional neural network and aspect based LSTM model. *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 696-700). https://doi.org/10.1109/ICIMIA48430.2020.9074939

Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2019). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing, 10*(1), 18-31. https://doi.org/10.1109/TAFFC.2017.2740923

Muhammad, G., & Hossain, M. S. (2021). Emotion recognition for cognitive edge computing using deep learning. *IEEE Internet of Things Journal, 8*(23), 16894-16901. https://doi.org/10.1109/JIOT.2021.3058587

Muhammad, S., Ahmed, S., & Naik, D. (2021). Real time emotion based music player using CNN architectures. *2021 6th International Conference for Convergence in Technology (I2CT)* (pp. 1-5). https://doi.org/10.1109/I2CT51068.2021.9417949

Mukhopadhyay, M., Pal, S., Nayyar, A., Pramanik, P. K. D., Dasgupta, N., & Choudhury, P. (2020). Facial emotion detection to assess learner's state of mind in an online learning system. *Proceedings of the 2020 5th International Conference on Intelligent Information Technology* (pp. 107-115). https://doi.org/10.1145/3385209.3385231

Nandi, A., Xhafa, F., Subirats, L., & Fort, S. (2020). A survey on multimodal data stream mining for e-learner's emotion recognition. *2020 International Conference on Omni-Layer Intelligent Systems (COINS)* (pp. 1–6). https://doi.org/10.1109/COINS49042.2020.9191370

Nandi, A., Xhafa, F., Subirats, L., & Fort, S. (2020). A survey on multimodal data stream mining for e-learner's emotion recognition. *2020 International Conference on Omni-Layer Intelligent Systems (COINS)* (pp. 1-6). https://doi.org/10.1109/COINS49042.2020.9191370

Ozdemir, M. A., Elagoz, B., Alaybeyoglu, A., Sadighzadeh, R., & Akan, A. (2019). Real time emotion recognition from facial expressions using CNN architecture. *2019 Medical Technologies Congress (TIPTEKNO)* (pp. 1-4). https://doi.org/10.1109/TIPTEKNO.2019.8895215

Afsar, P., Malayil, S., Verghese, A., Sithara, A., Pakarath, A., & Ashmin, K. T. (2023, July). Real-Time Student Emotion and Drowsiness Detection Using YOLOv5 and CNN for Enhanced Learning. In *2023 International Conference on Innovations in Engineering and Technology (ICIET)* (pp. 1-8). IEEE.https://doi.org/10.1109/ICIET57285.2023.10220929

Pantic, M., Valstar, M., Rademaker, R., & Maat, L. (2005). Web-based database for facial expression analysis. *2005 IEEE International Conference on Multimedia and Expo* (pp. 317-321). https://doi.org/10.1109/ICME.2005.1521424

Pathak, R., & Singh, Y. (2020). Real time baby facial expression recognition using deep learning and IoT edge computing. *2020 5th International Conference on Computing, Communication and Security (ICCCS)* (pp. 1-6). https://doi.org/10.1109/ICCCS49678.2020.9277428

Patil, V. K., Pawar, V. R., Kulkarni, S. P., Mehta, T. A., & Kharea, N. R. (2023). Real time emotion recognition with AD8232 ECG sensor for classwise performance evaluation of machine learning methods. *International Journal of Engineering, Transactions B: Applications, 36*(6), 1040-1047. https://doi.org/10.5829/ije.2023.36.06c.02

Picard, R. W. (1997). *Affective computing*. MIT Press.

Pinto, L. V. L., Alves, A. V. N., Medeiros, A. M., Costa, S. W. D. S., Pires, Y. P., Costa, F. A. R., & Seruffo, M. C. D. R. (2023). A systematic review of facial expression detection methods. *IEEE Access, 11*, 61881-61891. https://doi.org/10.1109/ACCESS.2023.3287090

Pranav, E., Kamal, S., Chandran, C. S., & Supriya, M. H. (2020). Facial emotion recognition using deep convolutional neural network. *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 317-320). https://doi.org/10.1109/ICACCS48705.2020.9074302

Pu, L., & Zhu, L. (2021). Differential residual learning for facial expression recognition. *Proceedings of the 2021 5th International Conference on Machine Learning and Soft Computing* (pp. 103-108). https://doi.org/10.1145/3453800.3453819

Rajan, S., Chenniappan, P., Devaraj, S., & Madian, N. (2020). Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM. *IET Image Processing, 14*(7), 1227-1232. https://doi.org/10.1049/iet-ipr.2019.1188

Ramirez Rios, F. E., & Reyes Duke, A. M. (2023). Building of a convolutional neuronal network for the prediction of mood states through face recognition based on object detection with YOLOV8 and Python. *2023 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)* (pp. 1-6). https://doi.org/10.1109/ICMLANT59547.2023.10372862

Rastgoo, M. N., Nakisa, B., Maire, F., Rakotonirainy, A., & Chandran, V. (2019). Automatic driver stress level classification using multimodal deep learning. *Expert Systems with Applications, 138*, 112793. https://doi.org/10.1016/j.eswa.2019.07.010

Rathour, N., Khanam, Z., Gehlot, A., Singh, R., Rashid, M., Alghamdi, A. S., & Alshamrani, S. S. (2021). Real-time facial emotion recognition framework for employees of organizations using Raspberry-Pi. *Applied Sciences, 11*(22), 10540. https://doi.org/10.3390/app112210540

Riaz, M. N., Shen, Y., Sohail, M., & Guo, M. (2020). eXnet: An efficient approach for emotion recognition in the wild. *Sensors, 20*(4), 1087. https://doi.org/10.3390/s20041087

Ruan, X., Palansuriya, C., & Constantin, A. (2022). Real-time feedback based on emotion recognition for improving children's metacognitive monitoring skill. *Proceedings of the 21st Annual ACM Interaction Design and Children Conference* (pp. 672-675). https://doi.org/10.1145/3501712.3538831

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533-536. https://doi.org/10.1038/323533a0

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin, 110*(3), 426-450. https://doi.org/10.1037/0033-2909.110.3.426

Salman, A., & Busso, C. (2022). Privacy preserving personalization for video facial expression recognition using federated learning. *Proceedings of the 2022 International Conference on Multimodal Interaction* (pp. 495-503). https://doi.org/10.1145/3536221.3556614

Samadiani, N., Huang, G., Cai, B., Luo, J., Chi, C.-H., Xiang, Y., & He, J. (2019). A review on automatic facial expression recognition systems assisted by multimodal sensor data. *Sensors, 19*(8), 1863. https://doi.org/10.3390/s19081863

Samal, A., & Iyengar, P. A. (1992). Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition, 25*(1), 65-77. https://doi.org/10.1016/0031-3203(92)90007-6

Sathya, R., Manivannan, R., & Vaidehi, K. (2022). Vision-based personal face emotional recognition approach using machine learning and tree-based classifier. In *[Book Title]* (pp. 561-573). Springer. https://doi.org/10.1007/978-981-16-6723-7_42

Savchenko, A. V. (2022). Video-based frame-level facial analysis of affective behavior on mobile devices using EfficientNets. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 2358–2365). https://doi.org/10.1109/CVPRW56347.2022.00263

Savchenko, A. V., Savchenko, L. V., & Makarov, I. (2022). Classifying emotions and engagement in online learning based on a single facial expression recognition neural network. *IEEE Transactions on Affective Computing, 13*(4), 2132–2143. https://doi.org/10.1109/TAFFC.2022.3188390

Sengupta, A., Ye, Y., Wang, R., Liu, C., & Roy, K. (2019). Going deeper in spiking neural networks: VGG and residual architectures. *Frontiers in Neuroscience, 13*, 95. https://doi.org/10.3389/fnins.2019.00095

Shetty, C., Khan, A., Singh, T., & Kharatmol, K. (2021). Movie review prediction system by real time analysis of facial expression. *2021 6th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1109–1113). https://doi.org/10.1109/ICCES51350.2021.9489171

Shi, M., Xu, L., & Chen, X. (2020). A novel facial expression intelligent recognition method using improved convolutional neural network. *IEEE Access, 8*, 57606-57614. https://doi.org/10.1109/ACCESS.2020.2982286

Shit, S., Rana, A., Das, D. K., & Ray, D. N. (2023). Real-time emotion recognition using end-to-end attention-based fusion network. *Journal of Electronic Imaging, 32*(1), 013050. https://doi.org/10.1117/1.JEI.32.1.013050

Siam, A. I., Soliman, N. F., Algarni, A. D., Abd El-Samie, F. E., & Sedik, A. (2022). Deploying machine learning techniques for human emotion detection. *Computational Intelligence and Neuroscience, 2022*, 8032673. https://doi.org/10.1155/2022/8032673

Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology, 70*(1), 747-770. https://doi.org/10.1146/annurev-psych-010418-102803

Siddiqui, N., Reither, T., Dave, R., Black, D., Bauer, T., & Hanson, M. (2022). A robust framework for deep learning approaches to facial emotion recognition and evaluation. *2022 International Conference on Advanced Computing and Machine Learning (CACML)* (pp. 68-73). https://doi.org/10.1109/CACML55074.2022.00020

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv*. https://doi.org/10.48550/arXiv.1409.1556

Singh, R., Saurav, S., Kumar, T., Saini, R., Vohra, A., & Singh, S. (2023). Facial expression recognition in videos using hybrid CNN & ConvLSTM. *International Journal of Information Technology, 15*(4), 1819-1830. https://doi.org/10.1007/s41870-023-01183-0

Singh, R., Singh, V., Verma, P., Rao, G. V. E., & Bakthula, R. (2023). Real-time mood-based music auto-play system from facial expressions. In *[Book Title]* (pp. 363-373). Springer. https://doi.org/10.1007/978-981-99-3734-9_30

Srinivas, P., Khamar, S. N., Borusu, N., Vuyyuru, H., & Raghavendra, K. M. G. (2023). Identification of facial emotions in hitech modern era. *2023 2nd International Conference on Edge Computing and Applications (ICECAA)* (pp. 1202-1208). https://doi.org/10.1109/ICECAA58104.2023.10212285

Taini, M., Zhao, G., Li, S. Z., & Pietikainen, M. (2008). Facial expression recognition from near-infrared video sequences. *2008 19th International Conference on Pattern Recognition* (pp. 1-4). https://doi.org/10.1109/ICPR.2008.4761697

Talaat, F. M. (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. https://doi.org/10.1007/s00521-023-08372-9

Thomas, M., & S., S. H. (2022). ANN based facial emotion detection and music selection. *2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC)* (pp. 927-931). https://doi.org/10.1109/IIHC55949.2022.10060593

Tian, Y.-I., Kanade, T., & Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(2), 97-115. https://doi.org/10.1109/34.908962

Turcian, D., & Stoicu-Tivadar, V. (2023). Real-time detection of emotions based on facial expression for mental health. *Studies in Health Technology and Informatics, 309*, 272-276. https://doi.org/10.3233/SHTI230795

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523-538. https://doi.org/10.1007/s11192-009-0146-3

Vedaldi, A., & Zisserman, A. (2016). *VGG convolutional neural networks practical*. Department of Engineering Science, University of Oxford.

Vijayeeta, P., & Pattnayak, P. (2022). A deep learning approach for emotion based music player. *2022 OITS International Conference on Information Technology (OCIT)* (pp. 278-282). https://doi.org/10.1109/OCIT56763.2022.00060

Villaroya, S. M., Gamboa-Montero, J. J., Bernardino, A., Maroto-Gomez, M., Castillo, J. C., & Salichs, M. A. (2022). Real-time engagement detection from facial features. *2022 IEEE International Conference on Development and Learning (ICDL)* (pp. 231-237). https://doi.org/10.1109/ICDL53763.2022.9962228

Wahab, M. N. A., Nazir, A., Ren, A. T. Z., Noor, M. H. M., Akbar, M. F., & Mohamed, A. S. A. (2021). EfficientNet-Lite and hybrid CNN-KNN implementation for facial expression recognition on Raspberry Pi. *IEEE Access, 9*, 134065–134080. https://doi.org/10.1109/ACCESS.2021.3113337

Wahab, M. N. A., Nazir, A., Ren, A. T. Z., Noor, M. H. M., Akbar, M. F., & Mohamed, A. S. A. (2021). EfficientNet-Lite and hybrid CNN-KNN implementation for facial expression recognition on Raspberry Pi. *IEEE Access, 9*, 134065-134080. https://doi.org/10.1109/ACCESS.2021.3113337

Waltman, L., Van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics, 4*(4), 629-635. https://doi.org/10.1016/j.joi.2010.07.002

Wang, H., Tobón V., D. P., Hossain, M. S., & Saddik, A. E. (2021). Deep learning (DL)-enabled system for emotional big data. *IEEE Access, 9*, 116073-116082. https://doi.org/10.1109/ACCESS.2021.3103501

Webb, N., Ruiz-Garcia, A., Elshaw, M., & Palade, V. (2020). Emotion recognition from face images in an unconstrained environment for usage on social robots. *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). https://doi.org/10.1109/IJCNN48605.2020.9207494

Werner, P., Lopez-Martinez, D., Walter, S., Al-Hamadi, A., Gruss, S., & Picard, R. W. (2022). Automatic recognition methods supporting pain assessment: A survey. *IEEE Transactions on Affective Computing, 13*(1), 530-552. https://doi.org/10.1109/TAFFC.2019.2946774

Winyangkun, T., Vanitchanant, N., Chouvatut, V., & Panyangam, B. (2023). Real-time detection and classification of facial emotions. *2023 15th International Conference on Knowledge and Smart Technology (KST)* (pp. 1-6). https://doi.org/10.1109/KST57286.2023.10086866

Woodward, K., Kanjo, E., & Tsanas, A. (2023). Combining deep learning with signal-image encoding for multi-modal mental wellbeing classification. *ACM Transactions on Computing for Healthcare, 4*(3). https://doi.org/10.1145/3631618

Xu, Y., Wei, L., & Wang, A. (2022). A study on the evaluation of English classroom learning status based on expression recognition technology. *2022 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE)* (pp. 439-442). https://doi.org/10.1109/ISAIEE57420.2022.00097

Yang, B., Zhang, Q., & Liu, Z. (2022). ICANet: A method of short video emotion recognition driven by multimodal data. *2022 2nd International Conference on Networking Systems of AI (INSAI)* (pp. 22–25). https://doi.org/10.1109/INSAI56792.2022.00014

Zarif, N. E., Montazeri, L., Leduc-Primeau, F., & Sawan, M. (2021). Mobile-optimized facial expression recognition techniques. *IEEE Access, 9*, 101172-101185. https://doi.org/10.1109/ACCESS.2021.3095844

Zhang, S., Liu, R., & Sawada, H. (2023). Deep learning-based system for real-time face tracking and expression recognition. *2023 IEEE International Conference on Mechatronics and Automation (ICMA)* (pp. 1598-1603). https://doi.org/10.1109/ICMA57826.2023.10215584

Zhang, Z., Fort, J. M., & Giménez Mateu, L. (2023). Facial expression recognition in virtual reality environments: Challenges and opportunities. *Frontiers in Psychology, 14*, 1280136. https://doi.org/10.3389/fpsyg.2023.1280136

Zhang, Z., Lin, P., Ma, S., & Xu, T. (2022). An improved Yolov5s algorithm for emotion detection. *2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)* (pp. 1002-1006). https://doi.org/10.1109/PRAI55851.2022.9904113

Zhao, G., Yang, H., & Yu, M. (2020). Expression recognition method based on a lightweight convolutional neural network. *IEEE Access, 8*, 38528-38537. https://doi.org/10.1109/ACCESS.2020.2964752

Zhao, Y., & Zeng, J. (2020). Library intelligent book recommendation system using facial expression recognition. *2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 55-58). https://doi.org/10.1109/IIAI-AAI50415.2020.00021

Zhou, N., Liang, R., & Shi, W. (2021). A lightweight convolutional neural network for real-time facial expression detection. *IEEE Access, 9*, 5573-5584. https://doi.org/10.1109/ACCESS.2020.3046715