Alexandria Engineering Journal (2023) 68, 817–840

HOSTED BY

ELSEVIER

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

REVIEW

A Comprehensive Survey on Deep Facial Expression Recognition: Challenges, Applications, and Future Guidelines

Muhammad Sajjad [a,*], Fath U Min Ullah [b], Mohib Ullah [a], Georgia Christodoulou [c], Faouzi Alaya Cheikh [a,*], Mohammad Hijji [d], Khan Muhammad [e,*], Joel J.P.C. Rodrigues [f,g]

[a] *The Software, Data and Digital Ecosystems (SDDE) Research Group, Department of Computer Science (IDI), Norwegian University of Science and Technology (NTNU), 2815 Gjøvik, Norway*
[b] *Sejong University, Seoul 143-747, Republic of Korea*
[c] *Catalink Limited, Charistinis Sakkada 5, Nicosia 1040, Cyprus*
[d] *Faculty of Computers and Information Technology (FCIT), University of Tabuk, Tabuk 47711, Saudi Arabia*
[e] *Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), Department of Applied Artificial Intelligence, School of Convergence, College of Computing and Informatics, Sungkyunkwan University, Seoul 03063, Republic of Korea*
[f] *College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266555, China*
[g] *Instituto de Telecomunicações, 6201-001 Covilhã, Portugal*

Received 12 July 2022; revised 31 December 2022; accepted 9 January 2023
Available online 10 February 2023

KEYWORDS

Facial expression recognition;
Edge vision;
Deep learning;
Machine learning;
Health care;
Security;
Artificial intelligence

Abstract Facial expression recognition (FER) is an emerging and multifaceted research topic. Applications of FER in healthcare, security, safe driving, and so forth have contributed to the credibility of these methods and their adoption in human-computer interaction for intelligent outcomes. Computational FER mimics human facial expression coding skills and conveys important cues that complement speech to assist listeners. Similarly, FER methods based on deep learning and artificial intelligence (AI) techniques have been developed with edge modules to ensure efficiency and real-time processing. To this end, numerous studies have explored different aspects of FER. Surveys of FER have focused on the literature on hand-crafted techniques, with a focus on general methods for local servers but largely neglecting edge vision-inspired deep learning and AI-based FER technologies. To consider these missing aspects, in this study, the existing literature on FER is thoroughly analyzed and surveyed, and the working flow of FER methods, their integral and intermediate steps, and pattern structures are highlighted. Further, the limitations in existing FER surveys are discussed. Next, FER datasets are investigated in depth, and the associated challenges and problems are discussed. In contrast to existing surveys, FER methods are considered for edge vision (on e.g., smartphone or Raspberry Pi, devices, etc.), and different measures to evaluate the performance of FER methods are comprehensively discussed. Finally, recommendations and

* Corresponding authors.
E-mail addresses: muhammad.sajjad@ntnu.no (M. Sajjad), faouzi.cheikh@ntnu.no (F. Alaya Cheikh), khan.muhammad@ieee.org (K. Muhammad).
Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2023.01.017

some avenues for future research are suggested to facilitate further development and implementation of FER technologies.

Contents

1. Introduction

The exponential growth of the facial expression recognition (FER) methods performed using computer vision, deep learning, and AI has been observed over the last few owing to its well-known applications in security [1,2], lecturing [3,4], medical rehabilitation [5], FER in the wild [6,7], and safe driving [8]. Facial expressions are remarkably essential in human communication and are produced through the movement of facial muscles and communicate with a range of signal types, from a state of deep-rooted survival to subtle communicative signals, such as raising the eyebrow in a conversational context [9]. Most psychological studies have reported that half of the information in a given speech is conveyed through emotions. Patients with Parkinson's disease develop symptoms of stiffness in facial muscle movements and reduced facial expression, known as hypomania [10], which may contribute to perceptions of FER. Similarly, individuals suffering from a stroke may have an impairment of the left anterior or posterior insula cortex, pallidus, and putamen, which can render the recognition of some of emotions more difficult. Generally, these emotions are complex and, in some cases, may be confused because their manifestation varies considerably among different people owing to differences in age, personal characteristics, gender, methods of communication, and so forth.

FER is significantly affected by the illumination, pose, background, and camera viewpoint of a source image, as well as occlusion or misalignment. Efficient FER relies on both the computations that occur in the visual-perceptual system, supported by perceptual processes, and extrapolated information from the perceptual system [11]. Three main representations of visual FER are sufficient and necessary, namely 1) a series of visual-perceptual representations of postures and the movement of observed expressions, 2) storage of the structural description of features characterizing the known expressions, and 3) semantic representations characterizing expressions.

An extensive body of literature has emerged on FER in the form of articles, surveys, literature reviews, and proposals. However, the works thus far have focused primarily only on working flow and feature extraction. To address the missing aspects, this article first analyzed and considered FER in detail by presenting a taxonomy of and statistics on prior works. To collect FER literature, a yearly search strategy that helped cover a wide range of articles from each year sequentially was applied; subsequently, the articles were correspondingly categorized. The search for articles involved several search engines, including Google, Google Scholar, ScienceDirect, and IEEE Explore. The search revealed an increasing interest in FER in the form of more published articles, and the latest methods tended to be inspired by neural networks and end-to-end network models. Additionally, newly developed emotion recognition datasets were constructed as the field was developed further over subsequent years. The number of articles published each year is shown in Fig. 1 (a). Next, the quality of research on FER was investigated. Highly cited articles have a greater impact on the research community. High citation scores indicate the influence of leading research directions. Hence, article citation scores for each year were considered herein. Some statistics on these citations are shown in Fig. 1 (b). Additionally, the working strategy and contributions of FER from 2015 onward were studied, and its coverage by different sources such as journals, publishers, ArXiv, and conferences are shown in Fig. 1 (c). Similarly, FER must be examined in terms of baseline strategies used to recognize expressions in video or image content. These strategies were broadly divided into three parts, and their visual representations are given in Fig. 1 (d). A list of abbreviations used in this work with their expansions are provided in Table 1.

FER has been considered from both academic and industrial perspectives, and can provide a window to the temperament, cognitive ability, personality, and psychopathology of individuals. For example, an increase in the use of FER technology in the clinical investigation of the effects of neuropsychiatric disorders on expression and perception has been shown to be tractable for quantitative research. Growth in the field of FER has been achieved owing to their wide range of applications in real-life scenarios, science fields, and medical services. Some applications of FER include gauging consumer's emotions regarding products or identifying suspicious activity. Automotive companies applying FER technology aim to make cars safer and more personalized for individual customers.

Human emotional expressions profoundly enrich our interactions with one another [12]. FER technology has been

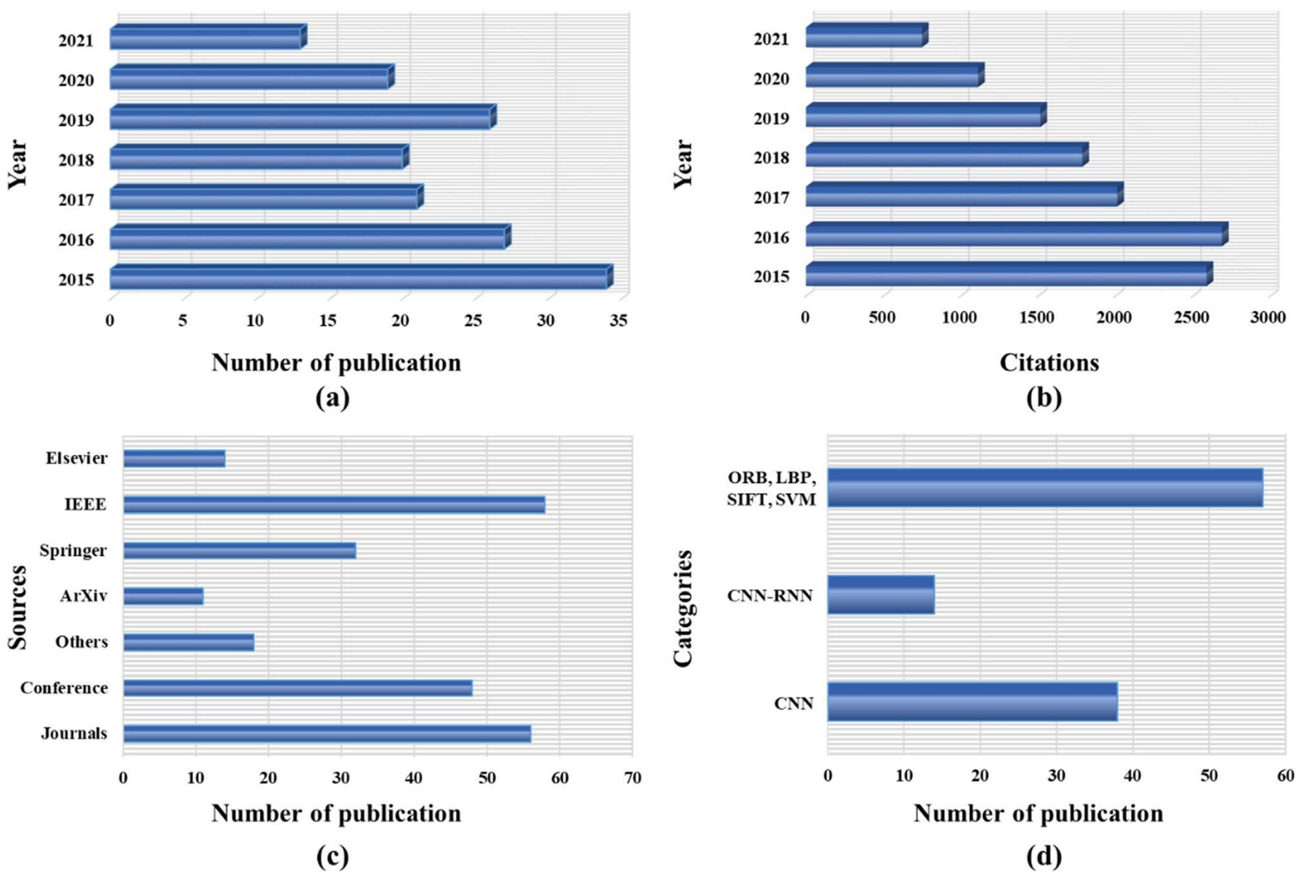

Fig. 1 Statistics of FER publications across search engines in terms of their citations score and publisher-wise distribution of FER methods. **(a)** Number of publications in each year ranging from 2015 to 2021. **(b)** Citations achieved by FER research in each year, where 2016 is the most cited year as the FER methods of this year have been well explored in the later research. **(c)** Division of publications in each portal. **(d)** Categorization of FER methods based on their baseline strategy.

Table 1 Abbreviations used throughout the survey.

Word	Description	Word	Description
ACNN	Attention mechanism CNN	KTN	Knowledgeable teacher network
AI	Artificial intelligence	KDEF	Karolinska directed emotional faces
AFEW	Acted facial expressions in the wild	LBP	Local binary patterns
BOVW	Bag-of-words	LGIN	Learnable graph inception network
BDLSTM	Bi-directional LSTM	LFW	Labelled faces in the wild
BIGRU	Bi-directional GRU	LSTM	Long short-term memory
BU-4DFE	Binghamton University 4d Facial Expression	LPDP	Local prominent directional pattern
BU-3DFE	Binghamton University 3d Facial Expression	MTCNN	Multitask cascaded convolutional networks
BAUM	Bahçeşehir University Multimodal Affective Database	MDSTFN	Multichannel deep spatial–temporal feature fusion neural network
CNN	Convolutional neural network	MLP	Multi-layer perceptron
CLAHE	Contrast limited adaptive histogram equalization	MSCNN	Multi-signal convolutional neural network
DCNN	Difference CNN	MUG	Multimedia understanding group
DWT	Discrete wavelet transform	NB	Naïve Bayesian
DAM-CNN	Deep attentive multi-path CNN	NCUFE	Nanchang University Facial Expression
DNN	Deep neural network	ORB	Oriented FAST and rotated BRIEF
DISFA	Denver intensity of spontaneous facial action	PNN	Parallel neural network
EEM	Emotional education mechanism	PHRNN	Part-based hierarchical bidirectional RNN
EDLM	Ensemble deep learning model	RAFD	Radboud face database
ELM	Extreme learning machine	RF	Random forest
FER	Facial expression recognition	R-CNN	Region-based CNN
FACS	Facial action coding system	RNN	Recurrent neural network
FNN	Feedforward neural network	RML	Ryerson multimedia research lab
FMPN	Facial motion prior networks	RAVDESS	Ryerson audio–visual database of emotional speech and song
FERET	Facial recognition technology	SSD	Single-shot detection
FEW	Facial expressions in the wild databases	SVM	Support vector machine
GRU	Gated recurrent unit	SFEW	Static facial expressions in the wild
GRNN	General regression neural network	SIFT	Scale-invariant feature transform
GFT	Group formation task	SURF	Speeded up robust features
HOG	Histogram of oriented gradients	SCN	Self-cure network
IWFER	Iranian wild facial expression recognition	SSN	Self-taught student network
IoT	Internet of Things	SWE	Stationary wavelet entropy
KNN	K-nearest neighbor	TFEID	Taiwanese facial expression image database
JAFFE	Japanese female facial expression	TFD	Toronto faces dataset

applied in healthcare with AI-empowered recognition to recognize patients' needs for medications or assist physicians in inquiring as to which patient may require more attention. Methods to exploring patients' emotions for better health system outcomes are being developed owing to their observed positive impacts in several medical fields. Automatic FER can assist doctors in operating smart centers to detect stress and depression among patients for health purposes. This approach may also help patients recognize psychological problems related to existing or previous medications [13]. Hospitals worldwide have begun to incorporate AI to handle patients' medication schedules as researchers have focused on applying neural networks to perform FER on patients.

1.1. Managerial and social implications of FER

Human expressions can show or conceal a variety of complex cognitive processes. Facial expressions elicit a rapid response and often imitate emotions. These effects occur on peoples' faces in a natural way and can be easily observed. By contrast, people recognize the expressions performed by robots but understand that they exhibit programmed behavior rather than the experience of a sentient being. The expressions shown by robots' faces are not reflexive but rather comprise a communication interface. In managerial or social human interaction, expressions can deliver a vast amount of information quite rapidly through the contraction of facial muscles in response to a particular action or question [14]. For instance, if an individual asks a certain question or asks for permission to perform some action, a response can be delivered through the movement of the eye muscles or head pose. Similarly, a person's state can be easily understood and discovered by observing only their facial appearance and muscle movements in response to a particular action. Thus, automatic FER methods are needed to enable computational systems to accurately gauge a person's mood. Regarding this, the proposed survey covers the aspects of FER systems and their challenges in detail as a step toward the development of improved expression recognition systems.

1.2. Applications of FER

In this section, FER applications are discussed in detail.

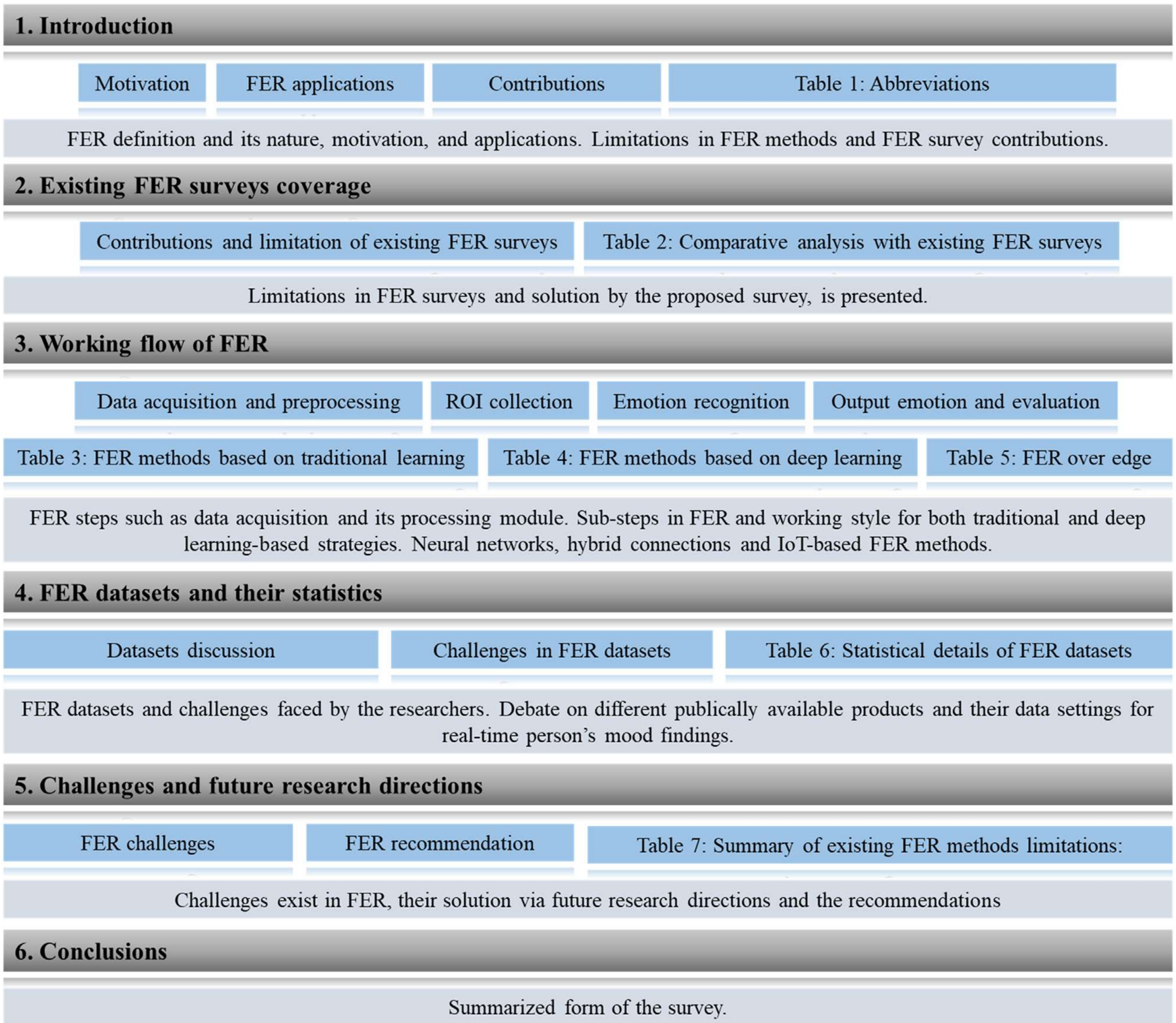

Fig. 2 Work flow of this survey.

1.2.1. FER for the prognosis and diagnosis of neurological disorders

FER is widely utilized in rehabilitation to help and monitor the patients; herein, the emotions of the patients are analyzed to help and provide medical care. Similarly, the doctors or a leading counsel can judge their patients or clients' emotional states from their appearance and body movements to note damaged or affected parts of their body. Patients inpatient care can be treated on a priority basis by capturing data on their state and moods through FER. Similarly, FER has been incorporated to facilitate the prognosis and diagnosis of neurological disorders (i.e., brain conditions or diseases), such as stroke, multiple sclerosis, and Parkinson's disease [10,15]. This enables clinicians to evaluate the mood of patients with neurological disorders. For example, a patient may express inappropriate or excessive emotions to express their state of mind or conditions. Therefore, recognizing these emotions is of value in monitoring patients via smartphone cameras [16].

1.2.2. FER in security

FER also plays an important role in security, where the malicious intentions of criminal suspects or perpetrators may be recognized by analyzing their expressions [17]. At present, ubiquitous surveillance has been implemented using security cameras installed in various locations, such as subways, markets, and stores. These camera feeds can be used to detect and analyze individual's facial emotions. These systems can identify suspicious activity, which can thus be prevented beforehand [18].

1.2.3. FER for learning

Educators can adjust their style of presentation according to learners by understanding learner's emotional expressions of their internal states. Students' enthusiasm may be improved by understanding their feelings in classroom or laboratory work [19].

Numerous groups are rapidly working on developing FER technology to improve performance and ensure real-time processing capability in various potential applications. Researchers must confront several issues and challenges due to the sensitive nature of changes in facial expressions. This survey provides information on the development of platforms for FER methods to show how they can be generalized to deliver a compact representation and learning terminology. Studies on FER are limited and largely describe only particular methods, with little or no focus on the deployment of such models on mobile platforms, such as edge devices and smart phones. Further, to the best of our knowledge, no detailed overview of deep learning and AI-based methods applied to this task has been conducted.

To overcome the existing challenges faced by current surveys, this study provides a comprehensive survey of the development and implementation of FER technologies, as shown in Fig. 2. The main contributions of this study are summarized as follows.

Contributions

1. To the best of our knowledge, this survey is the first to provide a thorough taxonomy of recent literature on FER that considers deep learning, conventional learning, hybrid approaches, and edge vision by analyzing the patterns of these works. In addition, the manner in which the FER has been considered is described from a medical perspective, such as for monitoring patients with Parkinson's disease, stroke, or dementia.
2. Existing surveys are largely limited to methods deployed to cloud computing or PC setups. However, this study covers both edge- and cloud-based FER methods. In addition, different platforms and products are investigated for this purpose. Further, an extensive set of information on debates on FER methods targeting the diagnosis of various diseases, as well as the corresponding journal details, their impact, and the number of citations are provided.
3. A general framework followed by the FER methods is presented. The datasets and challenges faced by researchers in this field are discussed comprehensively. Furthermore, these challenges are addressed by suggesting some promising directions for future research.

The remainder of this paper is organized as follows. Section 2 focuses on existing surveys and their downsides. Section 3 covers the working flow of FER systems in detail while considering of deep learning and conventional learning methods. Section 4 discusses existing FER datasets and some associated challenges. Section 5 sheds light on FER challenges and research guidelines. Finally, Section 6 concludes the paper with some final remarks and suggests possible avenues for future research.

2. Overview of the existing FER literature

This section explains recently published articles that have surveyed FER technology. This survey discusses the contributions and disadvantages of these previous articles and compares the proposed article with state-of-the-art FER surveys. First, the work presented by Zhang et al. [20] explained the advancements made in the creation of FER datasets and technique development. They focused primarily on occlusion problems and studied their effects on FER systems. Moreover, they represented FER in two ways, namely message-based and facial-component movement-based methods. They further categorized message-based methods into discrete and continuous dimensional methods. According to their review, the discrete categorical method is a long-standing method that has been widely adopted by psychologists to describe emotions. Similarly, the continuous method was adopted from psychology; it describes emotions in terms of continuous axes of a multidimensional space. By contrast, movement-based components use the movement of facial muscles for expression encoding. Similarly, Rajan et al. [21] covered FER techniques, the conventional classifiers used for FER classification, and FER datasets. Another recently published survey [22] considered an in-depth study of FER datasets and their creation, and subsequently properly aligned all the steps of conventional FER processes. Further, they overviewed the deep networks, sequential learning mechanisms, issues related to FER, and challenges faced by the researchers during FER; next, they highlighted some possible directions for future research on FER. These details are given Table 2.

Finally, this study presents the main contributions of this survey. The proposed survey presents a thorough FER taxonomy and the most recent FER literature developed for medical applications targeting patients with Parkinson's disease, stroke, multiple sclerosis, and CFS. Similarly, the preprocessing, main architecture steps, and evaluation metrics used to evaluate the performance of FER methods are extensively discussed. Furthermore, the current challenges and issues in FER, and the directions for future research on FER are presented.

3. Working flow of FER

This section describes the stepwise working flow of FER for real-time processing of the generic pipeline of FER, as shown in Fig. 3, and the details of the working procedure of the FER are given in Fig. 4. A comprehensive discussion of each step of the pipeline are provided below.

3.1. Data acquisition and preprocessing

Data collection through vision sensors and preprocessing are essential steps. The data are typically acquired from different sources such as Pi Cam devices, mobile phones, or surveillance cameras. Different data variations, such as illumination, head poses, and background, are common in uncertain scenarios. Therefore, before training a recognition model, preprocessing is applied to normalize and align the visual semantic information of the faces. Several face alignment techniques, such as holistic [26], part-based [27,28], DL-based [29–31], and cascaded alignment [32–34], have been widely applied for this purpose.

State-of-the-art AI-models contain a considerable number of parameters, typically in the order of millions. A sufficient amount of training data is required to ensure the generalizability of such models. However, most existing datasets available for training are insufficient for this purpose. To overcome this challenge, FER methods must apply data augmentation techniques.

Data augmentation methods are designed to expand the size of a dataset and its diversity by applying random perturbations, such as image shifting, skew, rotation, adding noise,

Table 2 Comparative analysis of the present work with existing recent surveys in terms of their categorization as considering deep learning (DL), conventional learning (CL), and hybrid approaches (HA).

Ref	Year	Platform		Categorization			Contributions	Remarks
		PC	Edge	DL	CL	HA		
[20]	2018	✓	✗	✓	✓	✗	-Data creation, technique development, and occlusion problem are investigated for FER systems and associated challenges are discussed.	-Only partial occlusion is widely considered. –No workflow mechanism is provided to describe FER steps. –No comparative study of mainstream FER surveys.
[21]	2019	✓	✗	✗	✓	✗	-FER techniques, classifiers, and datasets are surveyed. Some discussion on face detection methods and features extraction is provided.	-Most traditional FER techniques are covered. -A concrete and easily understandable framework is missing.
[23]	2020	✓	✗	✗	✗	✗	-Three aspects regarding to 3D FER such as face structure and its preprocessing and classification are investigated.	-The entire paper is based only on the occlusion problem under conditions of real-time emotion recognition.
[24]	2021	✓	✗	✓	✓	✗	-FER methods based on CNN are widely focused on with applications of FER.	–No coverage of challenges in FER. Methods are limited to CNN techniques only.
[25]	2022	✓	✓	✗	✓	✗	Major steps including preprocessing, features extraction, and classification are explained.	-Most popular challenges are not covered. Further, directions and recommendations for future research are not provided.
Our	2023	✓	✓	✓	✓	✓	-A thorough taxonomy of FER and the most recent FER literature is covered. Next, both edge- and cloud-based FER methods are highlighted. An extensive set of discussions on journals, citations, and FER applications is performed.	-Widely focused on FER literature and properly categorizing the FER algorithms as DL, CL, and HL techniques. Open challenges in FER are discussed, along with recommendations for future work.

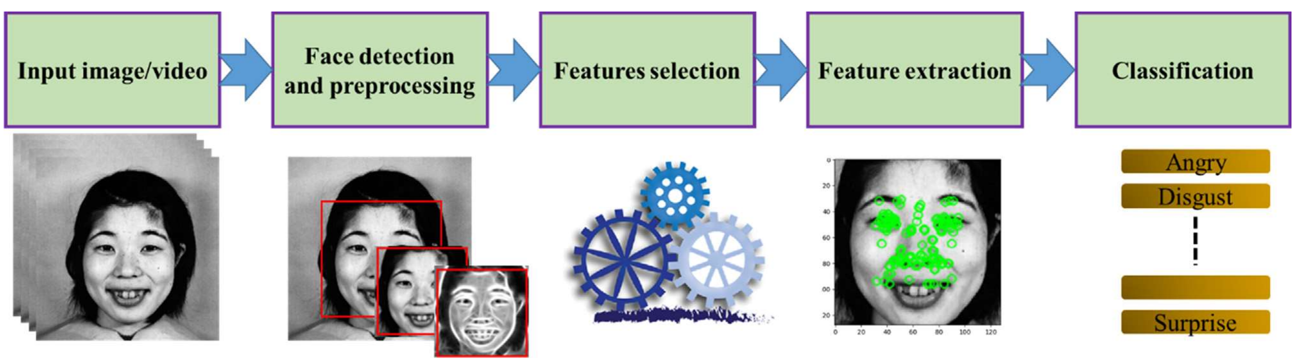

Fig. 3 Generic pipeline of FER (in the case of conventional learning).

and image scaling. More unseen training samples [35] can be generated through combinations of multiple operations that ensure a model's robustness to rotated and deviated faces [36].

3.2. ROI detection

Region of interest (ROI) detection (in this study, the face) is also referred to as facial detection. ROI detection is performed by AI-based techniques to identify and locate faces in images. These methods have been widely adopted in several applications, such as security [37], law enforcement [38], entertainment [39], and personal safety [40] which involve tracking or surveillance. They have advanced considerably from rudimentary vision techniques to enhanced machine learning and artificial neural networks (ANN) [41]. Facial detection is performed using conventional machine learning or deep learning approaches. Several techniques have been studied for face detection, including feature-based [42], knowledge-based [43], and appearance-based methods [44] as well as template matching [45]. In knowledge- or rule-based methods, the human face is described via defined rules and the representation depends entirely on how the rules are proposed. Similarly, feature invariant methods use different types of features, such as human eyes or nose, for face detection. However, this technique can be negatively affected by light and noise. In template matching, an image is compared with features that were previously stored or compared with standard face patterns and correlated for face detection. Furthermore, appearance-based techniques apply machine learning or statistical analysis to identify important face characteristics and have been widely applied to perform emotion recognition.

A major improvement in face detection occurred in 2001 when Viola and Jones proposed a face detection framework with high accuracy [46]. They proposed the use of Haar-like

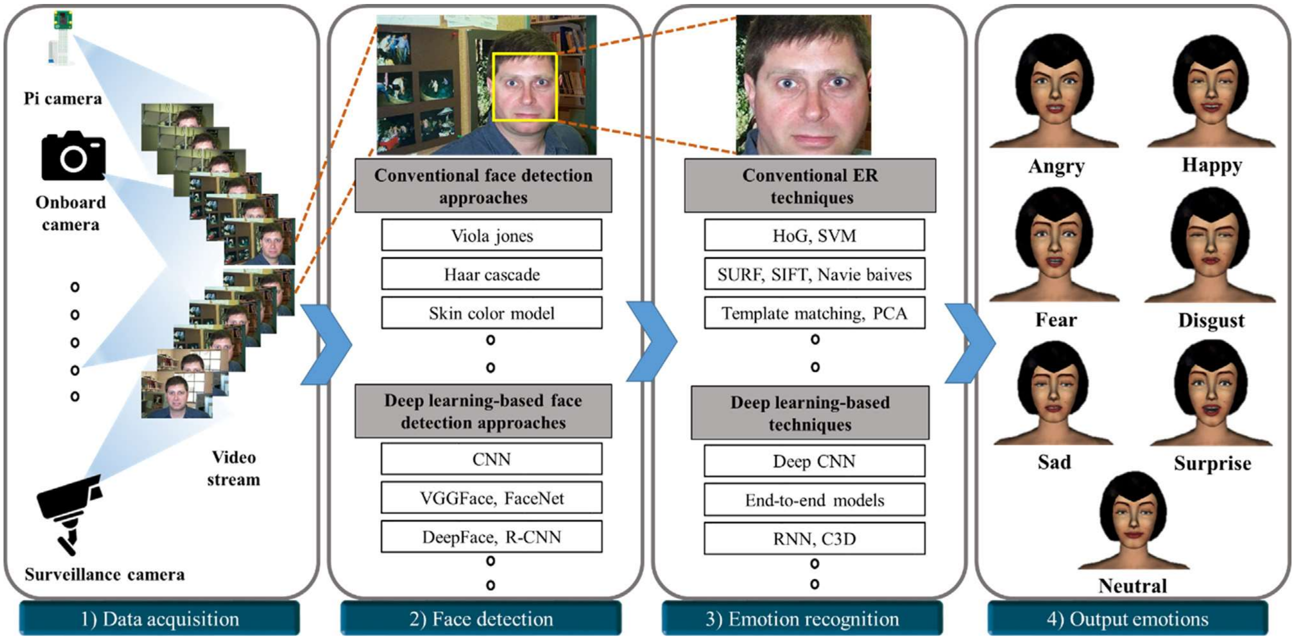

Fig. 4 Working flow of FER techniques using conventional and deep learning techniques. First, the data acquired from any source, such as Raspberry Pi, onboard camera or mobile phone camera devices, is fed into the face detection step. The second step performs face detection. The detected face is forwarded to the emotion recognition step.

features to detect faces. The algorithm observes numerous small subregions and attempts to determine a face by looking for specific features in each subregion. It passes through numerous different positions and scales because an image may contain several faces of various sizes.

The Viola–Jones algorithm remains popular for the detection of faces in real time but fails when a face is masked or covered by a scarf, or may be limited when a face is not oriented or aligned properly. Therefore, to avoid such problems in conventional techniques and improve face detection algorithms, deep learning algorithms, such as R-CNN [47], SSD [48], VGG-Face [49], FaceNet [50], have been developed. Among these, R-CNN was initially introduced for object detection and is significant for its capability of achieving high CNN accuracy on classification task in face detection tasks.

3.3. Emotion recognition

After face detection and ROI extraction, the flow proceeds to the FER stage. Numerous techniques, including conventional and deep-learning methods, are available for this. In conventional approaches, to conduct feature extraction, FER methods use hand-crafted feature engineering techniques, and the extracted features are subsequently fed into the classifier. By contrast, deep learning approaches can automatically extract features and perform classification in an end-to-end manner, where a loss layer is substituted to the end of the network to regulate the backpropagation error.

3.3.1. Conventional learning-based FER techniques

Conventional learning approaches include HOG [51], SVM [52], SURF [53], SIFT [54], and Naive Bayes [55]. Conventional practices use hand-crafted feature engineering techniques, such as preprocessing and data augmentation, prior to feature extraction. A mapped LBP feature was proposed in [56] for illumination-invariant FER. SIFT [57] features that are robust against image rotation and scaling are employed for multiview FER tasks. Combining several descriptors of texture, orientation, and color and using them as inputs helps enhance the performance of network [58,59].

Similarly, part-based representation extracts features by removing noncritical parts from the image and exploiting the key parts that are sensitive to the task. The authors in [60] reported that three regions of interest (ROIs), including the eyes, mouth, and eyebrows, are predominantly related to variations in emotion. Table 3 highlights recently published conventional machine learning FER methods.

3.3.2. Deep learning-based FER

Recently, deep learning has attracted considerable attention for research interest, and has achieved state-of-the-art performance in numerous applications in a wide variety of fields [78] such as computer vision [79,80], and time-series analysis and prediction [81]. Deep learning attempts to capture high-level abstractions via hierarchical networks comprising numerous nonlinear representations and transformations. Unlike conventional learning for FER, where the feature extraction and classification steps are independent, deep networks perform FER in an end-to-end manner. In particular, a loss layer is inserted at the network end to control the generated back-propagation error. Thus, the prediction probability obtained for each sample is directly produced as an output by the network. Typically, in a CNN, the SoftMax loss function is used. In particular, these models aim to minimize the cross-entropy of the model across the entire training dataset. This is achieved by calculating the average cross-entropy loss across all training examples and then back-propagating the loss through the network to optimize the defined loss function by tuning the

Table 3 FER methods based on conventional machine learning techniques with their contributions and corresponding training datasets.

Ref	Technique	Contributions	Dataset
[17]	ORB, SVM	-ORB features were extracted and fed into an SVM.	MMI, JAFFE
[61]	CNN, BoVW, SVM	-Features from a CNN were combined with handcrafted features computed using BOVW. -SVM is applied for final classification.	FER-2013, FER+, AFFECTNET
[62]	LPDP	-An edge descriptor LPDP was developed which considered statistical details of pixel neighborhoods to collect meaningful and reliable information.	CK+, MMI, FACES, ISED, GEMEP-FERA, BU-3DFE
[63]	FERAtt	-An end-to-end architecture which focused on human faces was proposed. -The model applied a Gaussian space representation to recognize an expression.	CK+, BU-3DFE
[64]	CNN	-Four-staged deep learning architectures were proposed. -The first three networks segmented the essential facial components, whereas the fourth combined the holistic facial information for better robustness.	RAFD
[65]	CNN, C4.5 classifier	-Features from CNN are combined with C4.5.	JAFFE, CK+, FER2013, RAFD
[66]	SCN	-SCN is proposed to efficiently suppresses uncertainties to prevent the network from overfitting. -This suppression enabled a self-attention mechanism and careful relabeling to perform well.	RAFD, AFFECTNET, FERPLUS
[67]	FACS	-FACS was developed to measure human facial behavior based on muscle movement.	N/A
[68]	N/A	-Bias and fairness were systematically investigated through three approaches such as attribute-aware, baseline, and disentangled approaches.	RAFD, CELEBA
[69]	3D CNN	-Deep spatiotemporal features were extracted based on deep appearance and neural network.	CK+, MMI, FERA
[70]	CNN	-An activation function was proposed for CNN models, and a piecewise activation technique was proposed for the procedure of FER tasks.	JAFFE, FER-2013
[71]	LBP	-An end-to-end network using an attention mechanism was proposed. -The network comprised features extraction, attention module, reconstruction module, and classification module components.	JAFFE, OULU-CASIA, NCUFE, CK+
[72]	N/A	An FER system validation study was performed for a school in this method.	NA
[73]	LBP, MSAU-Net	-Fine-grained FER in the wild was primarily considered and FG-Emotion was proposed. -FG-Emotions provided several features such as LBP and dense trajectories that facilitated the research.	FG-EMOTIONS, CK+, MMI, FER-2013, RAFD-BASIC, RAFD-COMPOUND
[74]	Channel State Information Processing	-A system based on Wi-Fi signals known as WiFace was developed for FER. -Series of algorithms were developed to process the channel state information signal to extract the most representative waveform patterns.	CSI (PRIVATE DATA)
[75]	KNN, NB, SVM, RF	-A system for FER based on multi-channel, electro-encephalogram, and multi-modal physiological signals was developed.	N/A
[76]	HOG, SVM	-TV-series were considered for human behavior analysis using facial expressions. -The authors detected and tracked faces using the Viola-Jones and Kanade-Lucas-Tomasi (KLT) algorithms -They extracted HOG features and classified the expression using an SVM model.	KDEF
[77]	EMM, KTN, SSN	-A supervised objective AdaReg loss and a re-weighting category was proposed to address class imbalance and increase discrimination expression power.	RAFD, AFFECTNET, FERPLUS

parameters of the network. In addition to end-to-end networks, DNN models can be used to extract features. Subsequently, a traditional classifier, such as an SVM or RF model, is applied to the extracted feature descriptor [82,83]. Furthermore, the works [84,85] presented a covariance descriptor computed via deep CNN features, and its classification was performed by Gaussian kernels on a symmetric positive definition. Table 4 highlights recently published conventional

Table 4 FER methods-based on deep learning mechanism with their contributions and data usage.

Ref	Technique	Contributions	Dataset
[99]	CNN, MTCNN	-MTCCN was used for face detection, while features were extracted via ResNet-64 and were classified at a large margin; a softmax loss was used for discriminative learning.	EMOTIW
[100]	CNN	-A method based on the LeNet-5 architecture, comprising five trainable parameter layers, two subsampling, and a fully connected layer, was proposed. -A SoftMax function was used for the final FER classification.	CK +
[101]	PHRNN, MSCNN	-A deep evolutional spatial–temporal network (composed of PHRNN and MSCNN) was used to extract the partial-whole, geometry-appearance, and dynamic-still information, thus effectively improving the performance of FER.	CK + , OULU-CASIA, MMI
[102]	LSTM-CNN	-For the facial label prediction, the authors used LSTM-CNN.	CK + , DISFA
[103]	3D inception-ResNet-LSTM	-A model with layers of an Inception-ResNet model were followed by an LSTM unit was proposed. -This method extracted temporal and spatial relations within facial images between different frames in video	CK + , MMI, FERA, DISFA
[104]	LSTM-CNN	-Using temporal dependencies, the LSTMs were stacked. -Outputs of CNN and LSTM were aggregated into a fusion network for per-frame prediction.	GFT, BP4D
[105]	CNN	-A prepressing step was used to clean and augment the data. -Subsequently, a CNN was used for feature extraction and classification.	CK + , JAFFE, BU-3DFE
[106]	CNN	-Four layers of CNN were used for features extraction and classification.	FER-2013
[107]	CNN, ACNN	-A CNN with ACNN was proposed to perceive occlusion regions in the face and emphasize the most discriminative un-occluded regions.	RAFD, AFFECTNET, SFEW, CK + , MMI, OULU-CASIA
[108]	CNN-RNN	-A hybrid CNN and RNN model was used for FER.	JAFFE, MMI
[109]	GoogLeNet, AlexNet	-The performance of two different models was compared for FER.	FER-2013
[110]	Pre-trained CNN Inception, VGG, VGG-Face	-Pre-trained state-of-the-art models were used for FER.	CK + , JAFFE, FACES
[111]	ConvNet, FaceNet	-Facial parts were focused on based on depth learning in the field of biometrics	LFW FACE
[112]	3D and 2D CNN	−3D FER was developed to accurately extract parts of face.	BU-3DFE
[113]	SWE and FNN	-FER based on Jaya algorithm was performed, using SWE for features extraction and an FNN for classification.	PRIVATE DATA: 700 FER IMAGES.
[114]	AlexNet CNN, FER-CNN, SVM, MLP	-Five different techniques for real-time basic expression recognition from images were compared.	CK + , KDEF
[115]	Hybrid CNN-SVM	-Humanoid robot for real-time FER was proposed based on convolutional self-learning feature extraction and an SVM classifier.	KDEF, CK +
[116]	FMPN	-An FER framework called FMPN was proposed, in which a branch was introduced for facial mask generation to focus on muscle movement regions.	CK + , MMI, AFFECTNET
[117]	NA	-Features extracted from an appearance-based network were fused with geometric features in hierarchical manner.	CK + , JAFFE
[118]	Spatial CNN, Temporal CNN	-A hybrid deep learning model was proposed for FER. -Two CNNs models, including Spatial and Temporal CNNs, were investigated for FER.	BAUM-1, RML, MMI
[119]	Ensembles of CNNs	-Different aspects of ensemble generation and other factors influencing the FER performance were studied.	FER-2013, CK + , SFEW
[120]	CNN	-An FER approach was presented using a CNN.	FER-2013
[121]	SIFT, CNN	-Features were extracted from SIFT and CNN.	CK + , MMI
[122]	Deep CNN	-Different deep learning methods were employed, with a CNN selected as the best algorithm for FER.	JAFFE
[123]	CNN	-A framework that combines the discriminative features learned via CNN and handcrafted features was proposed.	CK +
[124]	CNN, SVM	-SIFT and deep features from CNN for FER were combined and classified by SVM.	CK +
[125]	Light-CNN	-Three CNN models, namely, the light-CNN, dual-branch CNN, and pre-trained CNN models, were used to extract features for FER.	CK + , BU-3DFE, FER-2013
[126]	CNN	-A CNN was employed for FER.	FER-2013
[127]	CNN	-An FER system was developed based on a CNN model with data augmentation	CK + , FER-2013, MUG

Table 4 (*continued*)

Ref	Technique	Contributions	Dataset
[128]	CNN	-The Viola–Jones algorithm was applied for face detection, CLAHE for image enhancement, DWT to extract the features, and CNN for learning.	JAFFE, CK +
[129]	DAM-CNN	-A model called DAM-CNN was introduced for FER to automatically locate expression-based regions.	JAFFE, CK +, TFEID, BAUM-2I, SFEW
[130]	CNN	-Handcrafted features were proposed with a multi-stream structure to improve performance.	CK +, MUG, IWFER
[131]	CNN, LBP	-The abstract facial features learned via a deep CNN were fused with the modified LBP features.	ORL, CMU-PIE, FERET, FACE-SCRUB FACE
[132]	DCNN	-A two-staged framework based on a DCNN was proposed that was inspired by the nonstationary nature of facial expressions.	CK +, BU-4DFE
[133]	MDSTFN	-A multi-channel network was proposed to fuse and learn spatiotemporal features for FER. -An optical flow was extracted from the changes between the neutral and peak expression.	CK +, RAFD, MMI
[134]	CNN, Auto encoder, SVM	-A CNN-based pre-trained model was used in core cloud to extract deep features.	RML, ENTERFACE'05
[135]	CNN, ELM, SVM	-Speech signal was processed to obtain a mel-spectrogram treated as an image. The spectrogram was fed into a CNN. -The most representative frames were provided to a CNN model and were fused with the output obtained from another CNN model.	PRIVATE DATA
[136]	CNN, EDLM	-Based on ensemble learning model, an algorithm was proposed comprising three sub-networks with different depths. -The sub-networks comprised CNN models that were trained separately.	FER-2013, JAFFE, AFFECTNET
[137]	PNN, CNN, Residual Network, Capsule Network	-A PNN model designed to combine texture features was applied for FER. -This network was constructed using CNN, capsule network, and residual network models.	CK +
[138]	CNN	-The impact of CNN parameters, such as kernel size and number of filters, was investigated for FER.	FER-2013
[139]	CNN	-A vectorized CNN model introducing the attention mechanism to extract features in ROI of face was proposed. -ROIs were marked before feeding them into the network.	CK +, FER2013 AFFECT-NET, JAFFE
[93]	CNN, LSTM	-An FER algorithm was proposed based on a multilayer maxout linear activation function to initialize CNN and LSTM models.	JAFFE, CK +
[140]	CNN, LSTM	-A framework based on CNN and LSTM structures was developed. -Images were preprocessed and input to the CNN architecture.	CK +, MMI, SFEW
[141]	Fast R-CNN	-A video-based infant monitoring system was proposed to analyze infant expressions. -The expressions included discomfort, joy, unhappiness, and neutrality. -The system was based on Fast R-CNN.	PRIVATE DATA
[142]	CNN, LBP	-A system for FER was proposed based on CNN and LBP models.	FER-2013
[143]	CNN-BDLSTM	-An enhanced DNN framework was reported for pain intensity detection via facial expression image using four level thresholds.	VGG-FACE
[144]	CNN	-A CNN-based FER system was proposed from facial images considering edge computing. -The authors trained the model in the cloud and tested the trained model on edge devices.	JAFFE, CK +
[145]	LGIN	-A LGIN model proposed that was designed to learn to identify an underlying graph structure to recognize emotions.	RML, ENTERFACE, RAVDESS
[146]	Transfer learning	-A pre-trained CNN was utilized recognize facial emotions.	CK +, JAFFE
[147]	Firefly algorithm	-An FER technique was proposed based on the firefly algorithm, which was mainly used for feature optimization.	CK +, JAFFE, MMI
[148]	HOG, Deep CNN	-A DNN model was proposed for real-time FER. -The model was able to detect, track, and classify the human face with high performance.	KAGGLE FER DATASET
[149]	Fusion Technique	-Facial expressions were localized based on audio and video frames. -A network for audio recognition and facial recognition was proposed. -Both the networks were assembled as fusion network.	RML AUDIO-VISUAL DATABASE
[150]	Hybrid 3D CNN, RNN	-A DNN was proposed for FER based on videos and a network was used for audio as well.	AFEW-6.0, HAPPEI
[151]	VGGNet, ResNet, GoogleNet, AlexNet	-First, the structure of CNN models was studied. Next, four different CNNs models were applied to recognize human emotion.	FER-2013

(*continued on next page*)

Table 4 (*continued*)

Ref	Technique	Contributions	Dataset
[152]	DNN	-A DNN was proposed for the classification of facial expression based on a naturalistic dataset.	JAFFE, CK +
[153]	LBP, ANN	-LBP was implemented for feature extraction from images. -GRNN was implemented for the classification of FER based on frame features.	JAFFE, TFEID, CK +
[154]	LSM-RNN, SVM	-FER was performed based on LSTM-RNN and SVM models.	EMOTIW-2015
[155]	Deep learning methods	-A DNN was proposed based on a webcam for a smart TV environment to recognize human facial expressions.	FER-2013, CK +
[156]	DNNRL	-A deep learning method with relativity learning was proposed. -This model learned a mapping from the original images into a Euclidean space, where relative distances corresponded to a measure of facial expression similarity.	FER-2013, SFEW-2.0
[157]	CNN	-A deep CNN was presented for accurate detection of human face expressions.	FER-2013, JAFFE
[158]	CFS based on landmark and ANN	-An ANN model was presented to classify facial expressions. -A points/landmark technique was applied to enhance the performance of the ANN.	N/A
[159]	DNN	-Multiple DNNs were presented to detect face expressions and combine their performance.	SFEW-2.0, FER-2013, TFD, GENKI

machine learning methods. Table 5 summarizes FER for different edge devices and platforms for different application settings. Some of these methods were developed to be deployed over IoT devices; a detailed explanation of the libraries, training, settings, and other experiments involved is included in the same table.

As discussed above, directly training deep networks on relatively small FER datasets leads to problems of overfitting. To mitigate this problem, several studies have applied pre-training techniques, wherein popular networks such as AlexNet [86], VGG-face [49], and VGG [87] are pre-trained on benchmark datasets (such as ImageNet), and their last layers are fine-tuned to adapt the network to a particular task. The authors of [88] experimented with the VGG-Face model, which was initially trained for face recognition, and then fine-tuned using the FER 2013 dataset. The results of their experiments revealed that the VGG-Face model was more suitable for the FER task, compared with other networks that were pre-trained on the ImageNet dataset, which was developed for object recognition. Similarly, [89] observed that pre-training on large emotion recognition datasets positively affected recognition performance, and found that fine-tuning with more FER data could improve performance.

Existing techniques commonly adopt RNN models and their variants to recognize emotions in sequences of video frames. Hybrid connections with ConvNets models have achieved remarkable performance in several real-world applications. Details of these networks are provided in the following subsections.

3.3.2.1. LSTM and GRU. To capture the temporal dependencies of sequential data, deep recurrent networks, particularly LSTMs, have achieved promising performance. Recurrent neural networks (RNNs) are neural networks that contain cyclic connections (loops). This characteristic enables them to learn the temporal dynamics of sequential data well. RNNs can connect past information to the present task to predict the current output. However, training RNNs is challenging owing to the *vanishing or exploding gradient problem;* this is a situation in which the network is unable to propagate gradients from the output end of the model back to the layers near the input end of the model. A solution to this problem is the long short-term memory (LSTM) networks, a category of RNNs that can learn long-term dependencies. LSTMs have a chain-like structure comprising memory cells, which include four neurons each, designed to interact in a very special way.

Gated recurrent unit (GRU) models are a variation of the LSTM architecture. GRU models use fewer training parameters and, therefore, less memory. GRUs execute computations faster compared with LSTM models, whereas LSTM is more accurate for larger datasets. Existing state-of-the-art results have been obtained using LSTM or GRU networks. Training such networks for FER further improves performance. A sequence of frames is provided to an LSTM [90] or GRU [91] network to learn variations in facial expressions and determine a person's emotional or mental state. Some of these methods are listed in Table 4.

3.3.2.2. CNN-LSTM and CNN-GRU. Several pre-trained models based on CNN architectures and other related variants have been developed and trained for FER. These networks include self-encoder and CNN models as well as confidence networks. They typically exhibit a strong capability for automated feature learning but have no ability to capture contextual time information. For this purpose, several variants RNN models have been combined with CNNs to improve their performance on FER takes such as CNN-LSTM [92–94], CNN-GRU [95]. Such networks obtain richer and more discriminative expression information from facial expression sequences by eliminating the influence of differences and the external environment to improve recognition accuracy. In these networks, the CNN extracts deep visual information, and the LSTM learns to synthesize and identify the temporal dynamic sequence details. These networks focus on the influ-

Table 5 FER over different edge and IoT platforms along with recent products.

Ref/Paper	Description	Platform
[77]	-Training was performed on an NVIDIA TITAN Xp GPUs and deployed on a phone.	Smartphone
[144]	-Three prototypes were used. -The first prototype was an end device implemented on Android version 10, and the second was an edge component implemented using CUDA 10.0-enabled NVIDIA GeForce RTX 2070 8 GB GPU drivers with cuDNN v7.6 for deep learning models. The final result was a communication component with two parts, one running on a smartphone using Apache HttpClient to communicate with server and the other is running in the server with Django.	
[145]	-PyTorch was used with an NVIDIA RTX-2080Ti GPU for experiments.	
[160]	-An algorithm implemented in Python with PyTorch and OpenCV was used for the preprocessing operations on the images. The training of the CNN took approximately one hour with a single NVIDIA Titan X GPU. -To run the trained model on mobile device, it was converted into ONNX format and used ONNX-CoreML to obtain a CoreML model for use on iOS v1 1 or higher.	
[161]	-A smartphone app was used to analyze facial expressions and to construct a classifier to predicts emotional states in mobile settings. -In a testing phase, the feasibility of the approach was demonstrated for certain emotions using a person-dependent classifier.	
[74]	-The proposed model was easily deployable to smartphone devices.	
[105]	N/A	
[162]	N/A	
[144]	N/A	Raspberry Pi
[163]	N/A	Samsung S3
[137]	-The Python programming language on a GTX1070 GPU was used to train the model. -A model was proposed for IoT; however, the device was not defined.	IoT devices
[134]	-The model was proposed for IoT; however, the device was not defined.	
[135]	-The model was proposed for edge devices; however, the device was not defined.	Edge devices
[136]	-The model was proposed for IoT devices; however, the device was not defined.	
Different Products		
Product Name	**Link**	**Platform**
AffdexMe	[AffdexMe on the App Store (apple.com)]	IPhone, IPad
MorphCasto	[MorphCast - Facial Expression and Emotion Recognition AI | Face Emotion Analysis]	Mac/Apple
Emotient	[20 + Emotion Recognition APIs That Will Leave You Impressed, and Concerned | Nordic APIs |]	Apple
Affectiva		Smart phone

ence of micro-expression recognition. Some of these methods are listed in Table 4.

3.3.2.3. CNN-BDLSTM and CNN-BIGRU. BDLSTM and Bidirectional GRU (BIGRU) are extensions of traditional LSTM and GRU architectures, respectively; they improve the performance of learning models for more effective FER. BDLSTM trains two LSTM, and the sequence is processed in both the forward and backward directions. Thus, an additional context is provided to the network, which results in faster learning of the sequence of an expression. Therefore, for FER, a CNN is inserted at the end as a hybrid connection to help the model to deeply process the changes evident in facial expressions. These hybrid connection models include CNN-BDLSTM [96,97] and CNN-BIGRU [98]. Table 4 lists some of the hybrid methods.

3.4. Output emotion and evaluation

Once an FER model is competent in distinguishing different expressions in real-time, it is deployed on edge devices or clouds. The output emotion is generally one of seven emotions: happy, angry, fear, disgust, sad, surprise, or neutral. Performance is evaluated using several metrics, including precision, accuracy, recall, specificity, and F1-score. (Eq. (1)–(6)) Moreover, the method uses a confusion matrix that consists of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) rates. Similarly, models are analyzed in terms of their real-time deployment and sentiment analysis. The time complexity and FER model size were investigated for real-time deployment on edge devices.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

$$AccuracyB = \frac{TPR + TNR}{2} \tag{4}$$

$$Specificity = \frac{TN}{(TN + FP)} \quad (5)$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6)$$

4. FER datasets and associated statistics

For the effective and efficient design of deep ER, FER methods require a large, labeled training dataset that includes numerous variations of the surrounding environment and face structures. In this section, the public benchmark FER databases that include basic facial expressions, which are widely used in the studied papers, are discussed. Table 6 provides several FER datasets that have been widely used in different applications, such as security and law-enforcement, and Fig. 5 depicts frame samples from each expression of the FER datasets.

KDEF* [164]*: KDEF consists of 4900 total set of human expression images, where the averaged KDEF (AKDEF) is an image set proposed from the original KDEF. This dataset was announced in 1998, and since then, it has been publicly available. KDEF has been applied in more than 1500 research articles.

CK+* [165]*: CK+ is a widely used laboratory-controlled dataset for evaluating FER methods. This dataset consists of sequences that convert a neutral expression into a peak facial appearance. For assessment purposes, data selection is performed by selecting the latter one or two frames that contain peak information.

4DFAB* [166]*: This is a large-scale expression dataset with 1,800,000 high-resolution 3D facial images recorded from 180 subjects captured in distinct sessions. Videos of subjects are presented in 4D dynamics, displaying both posed and spontaneous facial variations of six basic emotions.

MMI* [167]*: This is another laboratory-controlled dataset. However, unlike CK+, the sequences of frames present in MMI are labeled in terms of onset and offset. For example, the frame sequence may start with a neutral expression, and reach a peak between the first and last neutral expressions.

JAFFE* [168]*: This contains approximately 213 samples of facial expressions taken from ten Japanese women, with 3–4 images corresponding to each individual showing six basic expressions with a neutral expression.

ExpW* [169]*: The Expression in the Wild dataset consists of 91,793 faces that are downloaded from Google where each face image was manually annotated with a category from among seven expressions. Images with no clear facial appearance were removed.

Table 6 Overview of the FER dataset with some statistical details.

Dataset	Classes	Samples	Individuals	Link
KDEF/AKDEF [164]	7	4900	272	[https://www.kdef.se/]
CK+ [165]		Sequence: 593	123	[https://www.consortium.ri.cmu.edu/ckagree/]
4DFAB [166]		1.8 million 3D faces	180	N/A
GEMEP FERA [183]	5	750,000	10	[https://www.cs.nott.ac.uk/]
MMI [167]	7	Images: 740 Videos: 2900	25	[https://mmifacedb.eu/]
JAFFE [168]		Images: 213	10	[https://www.kasrl.org/jaffe.html]
TFD [184]		Images: 112,234	N/A	[josh@mplab.ucsd.edu]
ExpW [169]		Images: 91,793		[https://mmlab.ie.cuhk.edu.hk/projects/socialrelation/index.html]
FER-2013 [170]		Images: 35,887		[https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge]
AFEW [171]		Videos: 1809		[https://sites.google.com/site/emotiwchallenge/]
SFEW [172]		Images: 1766		[https://cs.anu.edu.au/few/emotiw2015.html]
BP4D+ [185]	8	N/A	140	[https://suny.technologypublisher.com/]
Multi-PIE [173]	6	Images: 755,370	337	[https://www.flintbox.com/public/project/4742/]
EB+ [186]	N/A	—	N/A	[https://www.cs.binghamton.edu/]
BU 3DFE [174]	7	Images: 2500	100	[https://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFEAnalysis.html]
BU 4DFE [175]		3D sequences: 606	101	
Oulu CASIA [176]	6	Image sequence: 2880	80	[https://www.cse.oulu.fi/CMV/Downloads/Oulu-CASIA]
RAF-DB [177]	7	Images: 29,672	N/A	[https://www.whdeng.cn/RAF/model1.html]
KDEF [187]		Images: 4900	70	[https://www.emotionlab.se/kdef/]
EmotioNet [178]	23	Images: 1000,000	N/A	[https://cbcsl.ece.ohio-state.edu/dbformemotionet.html]
CASME II [179]	N/A	N/A	N/A	[https://fu.psych.ac.cn/]
AffectNet [180]	7	Images: 450,000	1 million	[https://mohammadmahoor.com/databases-codes/]
HAPPEI [181]	6	Images: 4886	greater than1	[https://cs.anu.edu.au/few/Group.htm]

The existing FER challenges are comprehensively discussed in detail.

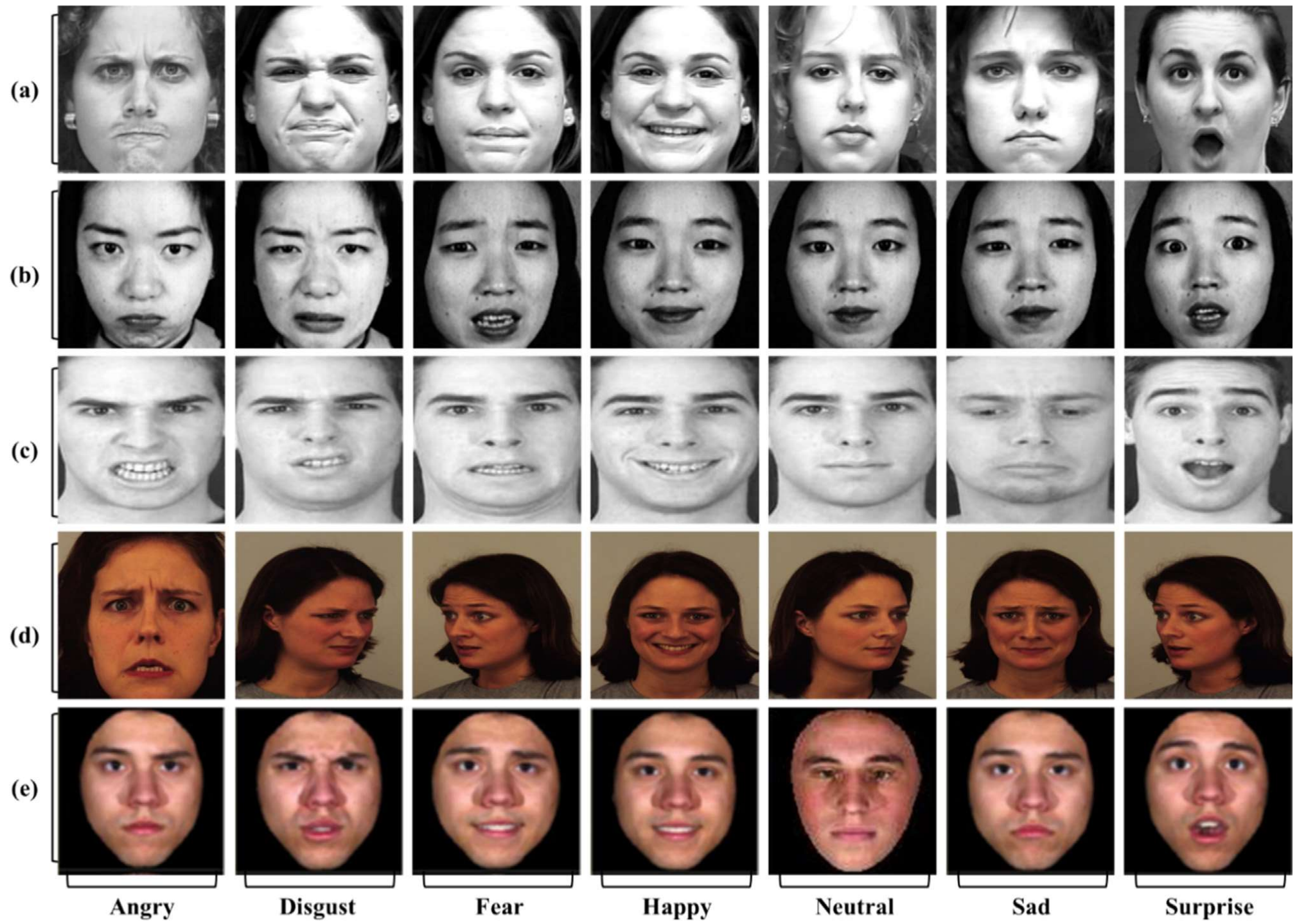

Fig. 5 Visual representation of facial expressions from different well-known datasets: **(a)** Cohn_kanade, **(b)** JAFFE, **(c)** MMI, **(d)** KDEF, and **(e)** BU-3DFE.

FER-2013 [170]***:*** This is an unrestrained large-scale dataset collected from the API of Google image search, wherein the images were registered and resized to 48 × 48 pixels after discarding incorrectly labeled frames. This dataset consists of 35,887 total images with seven emotion labels.

AFEW [171]***:*** This dataset consists of videos clips gathered from movies with impulsive expressions, diverse head poses, illuminations, and occlusions. This is a multimodal dataset that provides a wide range of environmental conditions for video and audio.

SFEW [172]***:*** This dataset was gathered from the static frames of the AFEW dataset. The most commonly applied version of SFEW 2.0 comprise three sets: training, testing, and validation. These labels are publicly accessible.

Multi-PIE [173]***:*** This dataset comprises 755,370 images ranging from 337 subjects with 19 illumination conditions up to four recorded sessions and 15 viewpoints, where each face image is labeled as one of six expressions. A multiview FER can be achieved using this dataset.

BU 3DFE [174]***:*** The BU 3DFE consists of 606 emotion sequences captured from 100 individuals. The six expressions were developed from each subject in different manners consisting of multiple intensities. Multi-PIE is also applicable to multiview FER analyses.

BU 4DFE [175]***:*** This dataset is used to analyze facial actions from static 3D space to dynamic 3D space. It contains 606 3D expression sequences in approximately 60,600 frames.

Oulu CASIA [176]***:*** This includes 2880 sequences obtained from 80 individuals, of which each video was recorded and processed by either infrared or visible light systems installed with three distinct illumination settings. The initial frame shows a neutral expression, while the peak expression is given in the last frame. The initial frame with neutral expression and the last three frames from 480 videos delivered by the visible light system under illumination were investigated experimentally.

RAF-DB [177]***:*** The real-world affective face dataset (RAF-DB), contains 29,672 diverse ranges of facial images collected from different sources on the Internet. Seven are basic, and eleven are compound emotion labels that were manually annotated.

EmotionNet [178]***:*** This is a large dataset with one million facial expressions collected from the Internet, of which 950,000 images were annotated using an automatic detection model in [178] and 25,000 images are annotated via 11 automatic detections.

CASME II [179]***:*** This is also laboratory-controlled dataset from which roughly 3000 facial movements, 247 expressions were chosen for the dataset with action units labeled. The samples showed spontaneous and dynamic expressions.

AffectNet [180]***:*** This consists of more than one million images gathered from the Internet by querying different search engines with search terms related to emotions.

HAPPEI [181]***:*** The happy people images is provided to evaluate the intensity of happiness in a group of people. This dataset contains 4886 samples sourced from Flickr using keywords that are associated with groups of people and occasions, such as parties, marriages, reunions, and bars. All collected samples contained more than one individual subject that was annotated with group-level mood.

Synthetic FER Dataset [182]***:*** Existing techniques have various limitations, such as sharpness, translation of distinct images, and preservation of identity. These issues are addressed via the texture deformation-based generative adversarial network, which disentangles the texture from a new image and based on the extracted textures, and transfers the domains.

Challenges in FER Datasets: Several challenges and issues related to FER datasets, such as a lack of large-scale expression data, image quality, and size, widely influence the recognition of emotion in both indoor and outdoor conditions. Numerous solutions have been applied to overcome these challenges. If the images are of very low quality, a diverse range of cleansing and smoothing filters can improve the quality of the frames and thus increase the accuracy of FER. Typically, datasets contain a limited amount of data. However, as deep learning models require large-scale data for training, data augmentation methods have been exploited to improve the diversity of training data and assist in training the network.

5. Challenges and future research directions

This section explains some notable challenges and identifies possible directions for future research.

5.1. FER challenges

Defining an expression as representative of a certain emotion can be difficult even for humans. Studies have shown that different people recognize different types of emotions in the same facial expression. FER involves numerous challenges, such as the fact that diverse training data are required, as well as imagery with diverse backgrounds, different genders, and different nationalities, etc.

5.1.1. Scarcity of FER datasets

Existing publicly available datasets do not suffice for effective FER, nor are they sufficiently diverse. These problems require effective solutions, such as data augmentation, combination of several datasets, modification of existing data, or creating a new dataset [79]. Typically, complex deep learning models are extremely "data-hungry," and require data in different forms for more effective and easier training. This solution avoids the overfitting problem in training the network. Therefore, FER requires data where the expression should be captured from all possible angles for effective outcomes.

5.1.1.1. Illumination. Illumination refers to light variation from different or single angles. A slight change in light conditions is a significant challenge for emotion recognition and significantly affects the results. Changes in illumination can drastically change facial appearance. Hence, the difference between two faces captured under different illuminations is higher than that of two distinct faces captured under the same illumination. This issue makes FER particularly challenging and has attracted attention over the last few decades. Numerous algorithms have been proposed to handle illumination, and they broadly involve three distinctions. The first approach deals with image processing methods that are helpful for the normalization of faces with distinct lighting effects. For this purpose, histogram equalization (HE) [188,189], logarithm transforms [190], or gamma intensity correlation [189] have been considered. Another approach is 3D facial modeling. Researchers in [191,192] suggested that a face viewed form the front with different illumination creates a cone known as the illumination cone. Similarly, in the third approach, the features of the face are extracted where they are illuminated, and the features are subsequently forwarded for recognition.

5.1.1.2. Face pose. Face pose is another major challenge; FER systems are very sensitive to slight changes in pose. The face pose varies with the head movement and changes in viewing angle. The head movement or variation in the camera point of view can cause changes in the facial appearance, thus creating intra-class variations and considerably decreasing the performance of FER methods [193]. However, despite the powerful recognition rate of CNN models to extract features, their recognition rate decreases significantly with the introduction of face poses [194]. The human face is roughly shaped like a convex spheroid, and pose leads to the self-occlusion phenomenon and reduces the FER accuracy. Therefore, performing FER reliably for different head posed remains a significant challenge.

5.1.1.3. Occlusion. Occlusion refers to cases in which a certain part of the face is not visible or is hidden. Occlusions occur because of beards, accessories, moustaches, masks, and so forth. The presence of such components makes the subjects more diverse and can causes recognition systems to fail. Owing to the complex and variable environment in which a face is presented, occlusion may change significantly. Occlusions in FER can be categorized into temporary and systematic [20]. Temporary occlusions occur when the face portion are temporarily obscured by other objects; for example, a hand-covering face, people moving across the face, or different environmental changes, such as lightening and shadows. Sometimes, self-occlusion may occurs owing to variation in head pose. Whereas, systematic occlusion is produced by the occurrence of individual facial components, such as hair, scars, or a moustaches [195].

5.1.1.4. Ageing. Human facial features tend to change with age, such as, lines, shapes, and some other aspects. Recognizing emotions in such cases is a very challenging, and solving this problem requires a considerable amount of training data. Considering the age of the face, the majority of the mainstream research has investigated whether posed facial expressions are decoded less accurately compared to young people faces [196,197], regardless of the expression. This occurs with facial muscle contraction and the actual landmark change [198]. Earlier literature attempted to discuss the decline in the recognition of expressions in several ways. For example, older people are presumed to focus on the lower half of their face during communications. Therefore, they can fail FER, which is expressed primarily in the eye regions.

5.1.1.5. Low resolution. Low-resolution images or videos in FER systems represent another challenge. The minimum resolution for a standard image is 16×16, whereas an image less than 16×16 is considered as low resolution for FER. Images with low resolution lead to the loss of feature information extracted via traditional techniques and the degradation of better recognition. Similarly, the feature distribution changes with a reduction in the resolution. This reduction occurs because of the limitations in the quality of the camera equipment and the distance of the person from the lens; therefore, the captured face image has different resolutions. Image super-resolution technology can recover high-resolution images from low-resolution images with rich information [199–201]. Some studies [202] have used image super-resolution to enhance low-resolution images for better FER.

5.2. Recommendations

A thorough investigation of FER methods throughout the literature reveals numerous drawbacks and limitations that need to be solved and addressed. A summary of these limitations is provided in Table 7, and a detailed discussion of these limitations and future research directions is provided below.

Table 7 Summarized form of limitation/drawbacks in existing FER.

#	Terms	Remarks
1	Bias and imbalanced data distribution	Bias and inconsistency exist in annotations that occur owing to different conditions and subjectivity of annotations. Therefore, the algorithms using intra-datasets lack generalizability on unseen data and exhibit reduced performance.
2	Single modalities	Humans with different behaviors in the real world include an encoding from various perspectives, whereas facial expressions in existing methods are based primarily on single modality.
3	Head motions, illumination, and aging	These variations widely effect the performance of the FER methods, particularly in videos and 2D images, whereas 3D data is somewhat robust to such variations.
4	Dependency	FER algorithms are dependent predominantly on large number of features points.
5	Manual intervention	Although FER methods are automatic, several systems still require intervention.
6	Age	Most methods do not consider the time and effects of age.
7	Dissimilarity in data	Facial data exhibit a high degree of dissimilarity, and FER systems can accurately recognize the expressions only for faces similar learned in training.
8	Action Units- (AU)	Detection of AU or combination of several AUS has not been addressed.

5.2.1. Surveillance-scaled FER datasets

As the focus of FER research shifts toward challenging in-the-wild environmental conditions, several researchers have focused on deep learning technologies designed to handle difficulties such as occlusions, illumination problems, nonfrontal poses, and recognition of lower-intensity emotion. As FER is a data-driven task in which the training of a deep network requires a large amount of training data to capture subtle facial expression-related deformations, the lack of large-scale training data is a major challenge in terms of quality and quantity. Owing to different genders and cultures, emotions are interpreted in different ways. An ideal dataset must include images with precise facial attribute labels, along with other attributes, such as gender, race, ethnicity, and age, thus facilitating related research on different genders, distant age ranges, and distinct cultural FER via deep learning methods, such as transfer learning approaches and deep networks. Similarly, existing FER datasets are widely captured using normal cameras, whereas FER patterns are only recorded in terms of regular patterns. Models trained on such data are less effective in recognizing expressions in surveillance footage or expressions that occur far from the camera viewpoint. The problem of occlusion and face pose has also attracted significant attention to overcome the scarcity of a diverse range of FER datasets covering different head-posing annotations and surveillance-based captured expressions.

5.2.2. FER with lower computational resources

Combining edge computing with the deep learning technologies is expected to further enhance data processing and ensure real-time processing to provide instant decisions. FER over an edge improves connectivity and security, and the data are processed over the edge. Edge intelligence further improves the network control of data and communication management, and helps reduce the time delay. Thus, the FER is performed with less computation, and the decision is made on the same platform where the entire processing is performed. For the FER domain, this may be considered a “missing concept” of performing recognition of emotions at the edge and making real-time decisions. Similarly, several devices can be clustered, thereby forming an IoT-assisted network, where all devices are interconnected and share information [17]. Such methods enable complex applications to be executed on the network edge with limited process power [203].

5.2.3. FER via E2E

Although a various technique that choose the learned features for FER as a prerequisite step can be found in the literature, deep networks or models that obtain a single video image as input and process it directly to generate the type of expression are lacking. The FER literature lacks such end-to-end (E2E) deep CNN models that can directly process frames and provide real-time expressions. Thus, the development of such models is highly recommended in the future for FER with satisfactory accuracy. Such networks are intended to process frames or sequences of frames from the camera through different convolutional layers and pooling layers. These models are expected to be relatively user-friendly, easy to operate, and employed for real-time FER.

5.2.4. Group expression analysis

Recognition of emotions of a single individual is comparatively easy for deep network models. However, a collective and group emotion may positively provide a thorough scenario of the ongoing action to analyze the mood and examine the subject's actions and probable gestures. Therefore, a group FER method, wherein the overall expression of all individuals is computed, is required. AI-based deep models should be proposed and fine-tuned for this purpose. Similarly, deep models are developed for deployment on the network edge to be easily equipped in a learning class or workplace.

5.2.5. FER everywhere

The exposure of FER-based code and implementation resources is a very important consideration in future research owing to its positive impact on real-world applications [76]. Although several techniques either introduce a novel way of learning expressions using hybrid frameworks or modified ER-based systems, such methods are limit to applications in homes, organizations, or other private sectors. Their implementation and related resources are private and unavailable for the development of real-time FER systems. Therefore, publicizing source codes along with all the resources used on different websites, including GitHub and "Papers with Code," is highly recommended for effective usage by the FER researchers.

5.2.6. Federated learning

Federated learning (FL) is a novel concept in machine learning; herein, an algorithm is dispersed among other edge devices or servers storing the data sample locally without exchanging it [204]. This procedure is different from the commonly applied centralized algorithms that require all local datasets to be loaded on a single server [205]. This learning enables the model to gain more experience from a wide range of datasets at different locations. Features are extracted from both audio and images [204] and the collected information recognizes facial expressions.

5.2.7. AML for FER

In adversarial machine learning (AML), adversaries act as malicious inputs designed to ensure that the model fails to predict the correct labels. In recent years, AML has become a crucial part of computer vision tasks, such as FER, object detection, and activity recognition. In [206], an AML approach was proposed that provides anonymity for individual subjects whose expressions have to be recognized by applying convolutional transformation, which degrades the individual relevant data for fully connected layers. The output was passed to two classifiers to recognize the expression.

6. Conclusion

Facial/emotion Recognition in real-time has a wide range of applications in healthcare, security, mood analysis, and safety measurements. Numerous studies have been conducted on this topic in the form of proposals, techniques, networks, and surveys. Computational FER mimics human coding skills and conveys important cues that complement speech to assist listeners. Similarly, the latest progress in FER development considers deep learning and AI using edge modules to ensure efficiency. To this end, numerous studies have contributed to the literature on FER. Most existing FER surveys focus on the features and characteristics of emotions from methods with different application directions. However, they have ignored the challenges of existing datasets and their solutions. Furthermore, most studies do not provide any direction or motivation towards the edge/IoT setup for facial emotion recognition. In this study, the existing FER techniques were surveyed, and the relevant literature was thoroughly analyzed and surveyed, essentially highlighting the FER working flow, integral and intermediate steps of most methods, as well as pattern structures and limitations in existing FER surveys. In contrast to current surveys, the FER for edge vision (that is, on mobile devices such as smartphones or Raspberry Pi computers) has been deliberately examined, and different FER evaluation tactics have been comprehensively discussed. Finally, a discussion on the challenges in FER along with some possible directions for future research were presented.

In the future, we plan to provide a detailed comparative analysis of FER methods applied for different purposes by exploring their implementation resources and algorithms. Our efforts will focus on the investigation and inclusion of FER in security, performance on edge devices, precision, and so forth. Similarly, data from different genders, races, and scenarios are not widely available; therefore, we plan to explore such datasets and evaluate their performance in terms of different aspects considering different modalities.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was funded by the European Union through the Horizon 2020 Research and Innovation Program, in the context of the ALAMEDA (Bridging the Early Diagnosis and Treatment Gap of Brain Diseases via Smart, Connected, Proactive and Evidence-based Technological Interventions) project under grant agreement No GA 101017558. This work is also partially funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the Project UIDB/50008/2020; and by the Brazilian National Council for Scientific and Technological Development-CNPq, via Grant No. 313036/2020-9.

References

[1] Y. Nan, J. Ju, Q. Hua, H. Zhang, B. Wang, A-MobileNet: An approach of facial expression recognition, Alex. Eng. J. 61 (6) (2022) 4435–4444.

[2] Z. Li, T. Zhang, X. Jing, Y. Wang, Facial expression-based analysis on emotion correlations, hotspots, and potential occurrence of urban crimes, Alex. Eng. J. 60 (1) (2021) 1411–1420.

[3] K. Mannepalli, P.N. Sastry, M. Suman, A novel adaptive fractional deep belief networks for speaker emotion recognition, Alex. Eng. J. 56 (4) (2017) 485–497.

[4] G. Tonguç, B.O. Ozkara, Automatic recognition of student emotions from facial expressions during a lecture, Comput. Educ. 148 (2020) 103797.
[5] S.S. Yun, J. Choi, S.K. Park, G.Y. Bong, H. Yoo, Social skills training for children with autism spectrum disorder using a robotic behavioral intervention system, Autism Res. 10 (7) (2017) 1306–1323.
[6] H. Li, M. Sui, F. Zhao, Z. Zha, and F. Wu, "Mvt: Mask vision transformer for facial expression recognition in the wild," *arXiv preprint arXiv:2106.04520,* 2021.
[7] X. Liang, L. Xu, W. Zhang, Y. Zhang, J. Liu, Z. Liu, A convolution-transformer dual branch network for head-pose and occlusion facial expression recognition, Vis. Comput. (2022) 1–14.
[8] M. Jeong, B.C. Ko, Driver's facial expression recognition in real-time for safe driving, Sensors 18 (12) (2018) 4270.
[9] K. Kaulard, D.W. Cunningham, H.H. Bülthoff, C. Wallraven, The MPI facial expression database—a validated database of emotional and conversational facial expressions, PLoS One 7 (3) (2012) e32321.
[10] M.R. Ali, T. Myers, E. Wagner, H. Ratnu, E. Dorsey, E. Hoque, Facial expressions can detect Parkinson's disease: preliminary evidence from videos collected online, npj Digital Med. 4 (1) (2021) 1–4.
[11] Y. Du, F. Zhang, Y. Wang, T. Bi, J. Qiu, Perceptual learning of facial expressions, Vision Res. 128 (2016) 19–29.
[12] A. A. Varghese, J. P. Cherian, and J. J. Kizhakkethottam, "Overview on emotion recognition system," in *2015 International Conference on Soft-Computing and Networks Security (ICSNS)*, 2015: IEEE, pp. 1-5.
[13] M. Egger, M. Ley, S. Hanke, Emotion recognition from physiological signal analysis: A review, Electron. Notes Theor. Comput. Sci. 343 (2019) 35–55.
[14] G. Mattavelli et al, Facial expressions recognition and discrimination in Parkinson's disease, J. Neuropsychol. 15 (1) (2021) 46–68.
[15] B. Sonawane, P. Sharma, Review of automated emotion-based quantification of facial expression in Parkinson's patients, Vis. Comput. 37 (5) (2021) 1151–1167.
[16] Y.-S. Lee, W.-H. Park, Diagnosis of Depressive Disorder Model on Facial Expression Based on Fast R-CNN, Diagnostics 12 (2) (2022) 317.
[17] M. Sajjad, M. Nasir, F.U.M. Ullah, K. Muhammad, A.K. Sangaiah, S.W. Baik, Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services, Inf. Sci. 479 (2019) 416–431.
[18] Y. Huang, X. Li, W. Wang, T. Jiang, and Q. Zhang, "Towards cross-modal forgery detection and localization on live surveillance videos," in *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, 2021: IEEE, pp. 1-10.
[19] K. Wang, X. Peng, J. Yang, D. Meng, Y. Qiao, Region attention networks for pose and occlusion robust facial expression recognition, IEEE Trans. Image Process. 29 (2020) 4057–4069.
[20] L. Zhang, B. Verma, D. Tjondronegoro, V. Chandran, Facial expression analysis under partial occlusion: A survey, ACM Computing Surveys (CSUR) 51 (2) (2018) 1–49.
[21] S. Rajan, P. Chenniappan, S. Devaraj, N. Madian, Facial expression recognition techniques: a comprehensive survey, IET Image Proc. 13 (7) (2019) 1031–1040.
[22] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE transactions on affective computing,* 2020.
[23] G.R. Alexandre, J.M. Soares, G.A.P. Thé, Systematic review of 3D facial expression recognition methods, Pattern Recogn. 100 (2020) 107108.
[24] S.M.S. Abdullah, A.M. Abdulazeez, Facial expression recognition based on deep learning convolution neural network: A review, J. Soft Comput. Data Min. 2 (1) (2021) 53–65.
[25] I.M. Revina, W.S. Emmanuel, A survey on human face expression recognition techniques, J. King Saud Univ.-Comput. Inform. Sci. 33 (6) (2021) 619–628.
[26] T. Cootes, J. Edwards, C. Taylor, Active apperance models. ieee transactions on pattern analysis and machine intelligence, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (1998) 681685.
[27] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *2012 IEEE conference on computer vision and pattern recognition*, 2012: IEEE, pp. 2879-2886.
[28] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3444-3451.
[29] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3476-3483.
[30] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, IEEE Signal Process Lett. 23 (10) (2016) 1499–1503.
[31] F.U.M. Ullah, M.S. Obaidat, A. Ullah, K. Muhammad, M. Hijji, S.W. Baik, A Comprehensive Review on Vision-based Violence Detection in Surveillance Videos, ACM Comput. Surv. (2023) 1–44.
[32] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 532-539.
[33] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 fps via regressing local binary features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1685-1692.
[34] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Incremental face alignment in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1859-1866.
[35] W. Li, M. Li, Z. Su, and Z. Zhu, "A deep-learning approach to facial expression recognition with candid images," in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, 2015: IEEE, pp. 279-282.
[36] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015, pp. 435-442.
[37] J. Tan et al, Face detection and verification using lensless cameras, IEEE Trans. Comput. Imaging 5 (2) (2018) 180–194.
[38] R. Ranjan et al, A fast and accurate system for face detection, identification, and verification, IEEE Trans. Biometrics, Behavior, Identity Sci. 1 (2) (2019) 82–96.
[39] C. Hong, J. Yu, J. Zhang, X. Jin, K.-H. Lee, Multimodal face-pose estimation with multitask manifold deep learning, IEEE Trans. Ind. Inf. 15 (7) (2018) 3952–3961.
[40] G. Sikander, S. Anwar, Driver fatigue detection systems: A review, IEEE Trans. Intell. Transp. Syst. 20 (6) (2018) 2339–2352.
[41] Z.-Q. Zhao, P. Zheng, S.-T. Xu, X. Wu, Object detection with deep learning: A review, IEEE Trans. Neural Networks Learn. Syst. 30 (11) (2019) 3212–3232.
[42] W. Kim, S. Suh, J.-J. Han, Face liveness detection from a single image via diffusion speed model, IEEE Trans. Image Process. 24 (8) (2015) 2456–2465.
[43] S. Zafeiriou, C. Zhang, Z. Zhang, A survey on face detection in the wild: past, present and future, Comput. Vis. Image Underst. 138 (2015) 1–24.
[44] H. Yang, L. Liu, W. Min, X. Yang, X. Xiong, Driver yawning detection based on subtle facial action recognition, IEEE Trans. Multimedia 23 (2020) 572–583.

[45] T. Zhang, J. Li, W. Jia, J. Sun, H. Yang, Fast and robust occluded face detection in ATM surveillance, Pattern Recogn. Lett. 107 (2018) 33–40.
[46] P. Viola, M.J. Jones, Robust real-time face detection, Int. J. Computer Vision 57 (2) (2004) 137–154.
[47] W. Wu, Y. Yin, X. Wang, D. Xu, Face detection with different scales based on faster R-CNN, IEEE Trans. Cybern. 49 (11) (2018) 4017–4028.
[48] R. Ranjan, V.M. Patel, R. Chellappa, Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition, IEEE Trans. Pattern Anal. Mach. Intell. 41 (1) (2017) 121–135.
[49] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," 2015.
[50] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815-823.
[51] P. Carcagni, M. Del Coco, M. Leo, C. Distante, Facial expression recognition and histograms of oriented gradients: a comprehensive study, Springerplus 4 (1) (2015) 1–25.
[52] L. Chen, C. Zhou, L. Shen, Facial expression recognition based on SVM in E-learning, Ieri Procedia 2 (2012) 781–787.
[53] Q. Rao, X. Qu, Q. Mao, and Y. Zhan, "Multi-pose facial expression recognition based on SURF boosting," in *2015 international conference on affective computing and intelligent interaction (ACII)*, 2015: IEEE, pp. 630-635.
[54] H. Soyel and H. Demirel, "Improved SIFT matching for pose robust facial expression recognition," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2011: IEEE, pp. 585-590.
[55] N. Sebe, M.S. Lew, I. Cohen, A. Garg, T.S. Huang, Emotion recognition using a cauchy naive bayes classifier, Object recognition supported by user interaction for service robots, vol. 1, IEEE, 2002, pp. 17–20.
[56] G. Levi and T. Hassner, "Emotion recognition in the wild via convolutional neural networks and mapped binary patterns," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015, pp. 503-510.
[57] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, 1999, vol. 2: Ieee, pp. 1150-1157.
[58] Z. Luo, J. Chen, T. Takiguchi, and Y. Ariki, "Facial Expression Recognition with deep age," in *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2017: IEEE, pp. 657-662.
[59] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, A.M. Dobaie, Facial expression recognition via learning deep sparse autoencoders, Neurocomputing 273 (2018) 643–649.
[60] L. Chen, M. Zhou, W. Su, M. Wu, J. She, K. Hirota, Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction, Inf. Sci. 428 (2018) 49–61.
[61] M.-I. Georgescu, R.T. Ionescu, M. Popescu, Local learning with deep and handcrafted features for facial expression recognition, IEEE Access 7 (2019) 64827–64836.
[62] F. Makhmudkhujaev, M. Abdullah-Al-Wadud, M.T.B. Iqbal, B. Ryu, O. Chae, Facial expression recognition with local prominent directional pattern, Signal Process. Image Commun. 74 (2019) 1–12.
[63] P. D. M. Fernandez, F. A. G. Pena, T. I. Ren, and A. Cunha, "Feratt: Facial expression recognition with attention net," *arXiv preprint arXiv:1902.03284,* vol. 3, 2019.
[64] G. Yolcu et al, Facial expression recognition for monitoring neurological disorders based on convolutional neural network, Multimed. Tools Appl. 78 (22) (2019) 31581–31603.
[65] Y. Wang, Y. Li, Y. Song, X. Rong, Facial expression recognition based on random forest and convolutional neural network, Information 10 (12) (2019) 375.
[66] K. Wang, X. Peng, J. Yang, S. Lu, and Y. Qiao, "Suppressing uncertainties for large-scale facial expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6897-6906.
[67] B. Waller, E. Julle-Daniere, J. Micheletta, Measuring the evolution of facial 'expression'using multi-species FACS, Neurosci. Biobehav. Rev. 113 (2020) 1–11.
[68] T. Xu, J. White, S. Kalkan, H. Gunes, Investigating bias and fairness in facial expression recognition, in: European Conference on Computer Vision, Springer, 2020, pp. 506–523.
[69] D. Jeong, B.-G. Kim, S.-Y. Dong, Deep joint spatiotemporal network (DJSTN) for efficient facial expression recognition, Sensors 20 (7) (2020) 1936.
[70] Y. Wang, Y. Li, Y. Song, X. Rong, The influence of the activation function in a convolution neural network model of facial expression recognition, Appl. Sci. 10 (5) (2020) 1897.
[71] J. Li, K. Jin, D. Zhou, N. Kubota, Z. Ju, Attention mechanism-based CNN for facial expression recognition, Neurocomputing 411 (2020) 340–350.
[72] M. Andrejevic, N. Selwyn, Facial recognition technology in schools: Critical questions and concerns, Learn. Media Technol. 45 (2) (2020) 115–128.
[73] L. Liang, C. Lang, Y. Li, S. Feng, J. Zhao, Fine-grained facial expression recognition in the wild, IEEE Trans. Inf. Forensics Secur. 16 (2020) 482–494.
[74] Y. Chen, R. Ou, Z. Li, K. Wu, WiFace: Facial Expression Recognition Using Wi-Fi Signals, IEEE Trans. Mob. Comput. (2020).
[75] J. Zhang, Z. Yin, P. Chen, S. Nichele, Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review, Information Fusion 59 (2020) 103–126.
[76] M. Sajjad, S. Zahir, A. Ullah, Z. Akhtar, K. Muhammad, Human behavior understanding in big multimedia data using CNN based facial expression recognition, Mobile Networks Appl. 25 (4) (2020) 1611–1621.
[77] H. Li, N. Wang, X. Ding, X. Yang, X. Gao, Adaptively Learning Facial Expression Representation via CF Labels and Distillation, IEEE Trans. Image Process. 30 (2021) 2016–2028.
[78] L. Deng, D. Yu, Deep learning: methods and applications, Found. Trends Signal Processing 7 (3–4) (2014) 197–387.
[79] F.U.M. Ullah, M.S. Obaidat, K. Muhammad, A. Ullah, S.W. Baik, F. Cuzzolin, J.J. Rodrigues, V.H. de Albuquerque, An intelligent system for complex violence pattern analysis and detection, Int. J. Intell. Syst. 37 (12) (2022) 10400–10422.
[80] F.U.M. Ullah, K. Muhammad, I.U. Haq, N. Khan, A.A. Heidari, S.W. Baik, V.H. de Albuquerque, et al, AI assisted Edge Vision for Violence Detection in IoT based Industrial Surveillance Networks, IEEE Trans. Ind. Inf. 18 (8) (2021) 5359–5370.
[81] F.U.M. Ullah, N. Khan, T. Hussain, M.Y. Lee, S.W. Baik, Diving Deep into Short-Term Electricity Load Forecasting: Comparative Analysis and a Novel Framework, Mathematics 9 (6) (2021) 611.
[82] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806-813.
[83] J. Donahue *et al*., "Decaf: A deep convolutional activation feature for generic visual recognition," in *International conference on machine learning*, 2014: PMLR, pp. 647-655.
[84] D. Acharya, Z. Huang, D. Pani Paudel, and L. Van Gool, "Covariance pooling for facial expression recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 367-374.
[85] N. Otberdout, A. Kacem, M. Daoudi, L. Ballihi, and S. Berretti, "Deep covariance descriptors for facial expression recognition," *arXiv preprint arXiv:1805.03869,* 2018.

[86] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Proces. Syst. 25 (2012) 1097–1105.
[87] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.
[88] H. Kaya, F. Gürpınar, A.A. Salah, Video-based emotion recognition in the wild using deep transfer learning and score fusion, Image Vis. Comput. 65 (2017) 66–75.
[89] B. Knyazev, R. Shvetsov, N. Efremova, and A. Kuharenko, "Convolutional neural networks pretrained on large face recognition datasets for emotion classification from video," *arXiv preprint arXiv:1711.04598,* 2017.
[90] Z. Yu, G. Liu, Q. Liu, J. Deng, Spatio-temporal convolutional features with nested LSTM for facial expression recognition, Neurocomputing 317 (2018) 50–57.
[91] R.-H. Huan, J. Shu, S.-L. Bao, R.-H. Liang, P. Chen, K.-K. Chi, Video multimodal emotion recognition based on Bi-GRU and attention fusion, Multimed. Tools Appl. 80 (6) (2021) 8213–8240.
[92] B.T. Hung, L.M. Tien, Facial expression recognition with CNN-LSTM, in: Research in Intelligent and Computing in Engineering, Springer, 2021, pp. 549–560.
[93] F. An, Z. Liu, Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM, Vis. Comput. 36 (3) (2020) 483–498.
[94] W. M. S. Abedi, A. T. Sadiq, and I. Nadher, "Modified CNN-LSTM for Pain Facial Expressions Recognition," 2020.
[95] M. T. Vu, M. Beurton-Aimar, and S. Marchand, "Multitask multi-database emotion recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3637-3644.
[96] Z.-X. Liu, D.-G. Zhang, G.-Z. Luo, M. Lian, B. Liu, A new method of emotional analysis based on CNN–BiLSTM hybrid neural network, Clust. Comput. 23 (4) (2020) 2901–2913.
[97] P. Du, X. Li, and Y. Gao, "Dynamic Music emotion recognition based on CNN-BiLSTM," in *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 2020: IEEE, pp. 1372-1376.
[98] W. Yan, L. Zhou, Z. Qian, L. Xiao, H. Zhu, Sentiment Analysis of Student Texts Using the CNN-BiGRU-AT Model, Sci. Program. 2021 (2021).
[99] L. Tan, K. Zhang, K. Wang, X. Zeng, X. Peng, and Y. Qiao, "Group emotion recognition with individual facial emotion CNNs and global image based CNNs," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017, pp. 549-552.
[100] M. Mohammadpour, H. Khaliliardali, S. M. R. Hashemi, and M. M. AlyanNezhadi, "Facial emotion recognition using deep convolutional networks," in *2017 IEEE 4th international conference on knowledge-based engineering and innovation (KBEI)*, 2017: IEEE, pp. 0017-0021.
[101] K. Zhang, Y. Huang, Y. Du, L. Wang, Facial expression recognition based on deep evolutional spatial-temporal networks, IEEE Trans. Image Process. 26 (9) (2017) 4193–4203.
[102] D.K. Jain, Z. Zhang, K. Huang, Multi angle optimal pattern-based deep learning for automatic facial expression recognition, Pattern Recogn. Lett. (2017).
[103] B. Hasani, M.H. Mahoor, Facial expression recognition using enhanced deep 3D convolutional neural networks, in: in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 30–40.
[104] W.-S. Chu, F. De la Torre, and J. F. Cohn, "Learning spatial and temporal cues for multi-label facial action unit detection," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017: IEEE, pp. 25-32.
[105] A.T. Lopes, E. de Aguiar, A.F. De Souza, T. Oliveira-Santos, Facial expression recognition with convolutional neural networks: coping with few data and the training sample order, Pattern Recogn. 61 (2017) 610–628.
[106] H. Yar, T. Jan, A. Hussain, and S. Din, "Real-Time Facial Emotion Recognition and Gender Classification for Human Robot Interaction Using CNN," ed.
[107] Y. Li, J. Zeng, S. Shan, X. Chen, Occlusion aware facial expression recognition using CNN with attention mechanism, IEEE Trans. Image Process. 28 (5) (2018) 2439–2450.
[108] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, and M. J. P. R. L. Zareapoor, "Hybrid deep neural networks for face emotion recognition," vol. 115, pp. 101-106, 2018.
[109] P. Giannopoulos, I. Perikos, I. Hatzilygeroudis, Deep learning approaches for facial emotion recognition: A case study on FER-2013, in: Advances in hybridization of intelligent methods, Springer, 2018, pp. 1–16.
[110] A. Sajjanhar, Z. Wu, and Q. Wen, "Deep learning models for facial expression recognition," in *2018 digital image computing: Techniques and applications (dicta)*, 2018: IEEE, pp. 1-6.
[111] X. Han, Q. Du, Research on face recognition based on deep learning, in: Sixth International Conference on Digital Information, Networking, and Wireless Communications (DINWC), Beirut, 2018, pp. 53–58, https://doi.org/10.1109/DINWC.2018.8356995.
[112] A. Jan, H. Ding, H. Meng, L. Chen, and H. Li, "Accurate facial parts localization and deep learning for 3D facial expression recognition," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018: IEEE, pp. 466-472.
[113] S.-H. Wang, P. Phillips, Z.-C. Dong, Y.-D. Zhang, Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm, Neurocomputing 272 (2018) 668–676.
[114] A. Kartali, M. Roglić, M. Barjaktarović, M. Đurić-Jovičić, and M. M. Janković, "Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches," in *2018 14th Symposium on Neural Networks and Applications (NEUREL)*, 2018: IEEE, pp. 1-4.
[115] A. Ruiz-Garcia, M. Elshaw, A. Altahhan, V. Palade, A hybrid deep learning neural approach for emotion recognition from facial expressions for socially assistive robots, Neural Comput. & Applic. 29 (7) (2018) 359–373.
[116] Y. Chen, J. Wang, S. Chen, Z. Shi, and J. Cai, "Facial motion prior networks for facial expression recognition," in *2019 IEEE Visual Communications and Image Processing (VCIP)*, 2019: IEEE, pp. 1-4.
[117] J.-H. Kim, B.-G. Kim, P.P. Roy, D.-M. Jeong, Efficient facial expression recognition algorithm based on hierarchical deep neural network structure, IEEE Access 7 (2019) 41273–41285.
[118] S. Zhang, X. Pan, Y. Cui, X. Zhao, L. Liu, Learning affective video features for facial expression recognition via hybrid deep learning, IEEE Access 7 (2019) 32297–32304.
[119] "IEA, Electricity mix in the European Union, January-May 2020, IEA, Paris https://www.iea.org/data-and-statistics/charts/electricity-mix-in-the-european-union-january-may-2020.".
[120] I. Talegaonkar, K. Joshi, S. Valunj, R. Kohok, A. Kulkarni, Real time facial expression recognition using deep learning, in *Proceedings of International Conference on Communication and Information Processing (ICCIP)*, 2019.
[121] X. Sun, M. Lv, Facial expression recognition based on a hybrid model combining deep and shallow features, Cogn. Comput. 11 (4) (2019) 587–597.
[122] C. M. M. Refat and N. Z. Azlan, "Deep learning methods for facial expression recognition," in *2019 7th International Conference on Mechatronics Engineering (ICOM)*, 2019: IEEE, pp. 1-6.

[123] X. Fan, T. Tjahjadi, Fusing dynamic deep learned features and handcrafted features for facial expression recognition, J. Vis. Commun. Image Represent. 65 (2019) 102659.

[124] F. Wang, J. Lv, G. Ying, S. Chen, C. Zhang, Facial expression recognition from image based on hybrid features understanding, J. Vis. Commun. Image Represent. 59 (2019) 84–88.

[125] J. Shao, Y. Qian, Three convolutional neural network models for facial expression recognition in the wild, Neurocomputing 355 (2019) 82–92.

[126] K.-C. Liu, C.-C. Hsu, W.-Y. Wang, and H.-H. Chiang, "Real-Time facial expression recognition based on cnn," in *2019 International Conference on System Science and Engineering (ICSSE)*, 2019: IEEE, pp. 120-123.

[127] T. U. Ahmed, S. Hossain, M. S. Hossain, R. ul Islam, and K. Andersson, "Facial expression recognition using convolutional neural network with data augmentation," in *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 2019: IEEE, pp. 336-341.

[128] R.I. Bendjillali, M. Beladgham, K. Merit, A. Taleb-Ahmed, Improved facial expression recognition based on DWT feature for deep CNN, Electronics 8 (3) (2019) 324.

[129] S. Xie, H. Hu, Y. Wu, Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition, Pattern Recogn. 92 (2019) 177–191.

[130] J.A. Aghamaleki, V. Ashkani Chenarlogh, Multi-stream CNN for facial expression recognition in limited training data, Multimed. Tools Appl. 78 (16) (2019) 22861–22882.

[131] F. Kong, Facial expression recognition method based on deep convolutional neural network combined with improved LBP features, Pers. Ubiquit. Comput. 23 (3) (2019) 531–539.

[132] J. Chen, Y. Lv, R. Xu, C. Xu, Automatic social signal analysis: Facial expression recognition using difference convolution neural network, J. Parallel Distrib. Comput. 131 (2019) 97–102.

[133] N. Sun, Q. Li, R. Huan, J. Liu, G. Han, Deep spatial-temporal feature fusion for facial expression recognition in static images, Pattern Recogn. Lett. 119 (2019) 49–61.

[134] M.S. Hossain, G. Muhammad, Emotion recognition using secure edge and cloud computing, Inf. Sci. 504 (2019) 589–601.

[135] M.S. Hossain, G. Muhammad, Emotion recognition using deep learning approach from audio–visual emotional big data, Information Fusion 49 (2019) 69–78.

[136] W. Hua, F. Dai, L. Huang, J. Xiong, G. Gui, HERO: Human emotions recognition for realizing intelligent Internet of Things, IEEE Access 7 (2019) 24321–24332.

[137] X. Zhenghao, Y. Niu, J. Chen, X. Kan, H. Liu, Facial Expression Recognition of Industrial Internet of Things by Parallel Neural Networks Combining Texture Features, IEEE Trans. Ind. Inf. (2020).

[138] A. Agrawal, N. Mittal, Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy, Vis. Comput. 36 (2) (2020) 405–412.

[139] X. Sun, S. Zheng, H. Fu, ROI-attention vectorized CNN model for static facial expression recognition, IEEE Access 8 (2020) 7183–7194.

[140] S. Rajan, P. Chenniappan, S. Devaraj, N. Madian, Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM, IET Image Proc. 14 (7) (2020) 1373–1381.

[141] C. Li, A. Pourtaherian, L. van Onzenoort, W. T. a Ten, and P. de With, "Infant facial expression analysis: towards a real-time video monitoring system using r-cnn and hmm," *IEEE Journal of Biomedical and Health Informatics,* vol. 25, no. 5, pp. 1429-1440, 2020.

[142] Q. Xu and N. Zhao, "A facial expression recognition algorithm based on CNN and LBP feature," in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 2020, vol. 1: IEEE, pp. 2304-2308.

[143] G. Bargshady, X. Zhou, R.C. Deo, J. Soar, F. Whittaker, H. Wang, Enhanced deep learning algorithm development to detect pain intensity from facial expression images, Expert Syst. Appl. 149 (2020) 113305.

[144] G. Muhammad and M. S. Hossain, "Emotion Recognition for Cognitive Edge Computing Using Deep Learning," *IEEE Internet of Things Journal,* 2021.

[145] A. Shirian, S. Tripathi, T. Guha, Dynamic Emotion Modeling with Learnable Graphs and Graph Inception Network, IEEE Trans. Multimedia (2021).

[146] D. Duncan, G. Shine, C. English, "Facial emotion recognition in real time,", Comput. Sci. (2016) 1–7.

[147] T. Zhang, W. Jia, X. He, J. Yang, Discriminative dictionary learning with motion weber local descriptor for violence detection, IEEE Trans. Circuits Syst. Video Technol. 27 (3) (2016) 696–709.

[148] J. Jeon et al, A real-time facial expression recognizer using deep neural network, in: *Proceedings of the 10th international conference on ubiquitous information management and communication*, 2016, pp. 1–4.

[149] S. Zhang, S. Zhang, T. Huang, W. Gao, Multimodal deep convolutional neural network for audio-visual emotion recognition, in: *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, 2016, pp. 281–284.

[150] Y. Fan, X. Lu, D. Li, Y. Liu, Video-based emotion recognition using CNN-RNN and C3D hybrid networks, in: *Proceedings of the 18th ACM international conference on multimodal interaction*, 2016, pp. 445–450.

[151] Y. Gan, "Facial expression recognition using convolutional neural network," in *Proceedings of the 2nd international conference on vision, image and signal processing*, 2018, pp. 1-5.

[152] X. Peng, Z. Xia, L. Li, and X. Feng, "Towards facial expression recognition in the wild: A new database and deep recognition system," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 93-99.

[153] K. Talele, A. Shirsat, T. Uplenchwar, and K. Tuckley, "Facial expression recognition using general regression neural network," in *2016 IEEE Bombay Section Symposium (IBSS)*, 2016: IEEE, pp. 1-6.

[154] L. Chao, J. Tao, M. Yang, Y. Li, and Z. Wen, "Long short term memory recurrent neural network based encoding method for emotion recognition in video," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016: IEEE, pp. 2752-2756.

[155] I. Lee, H. Jung, C. H. Ahn, J. Seo, J. Kim, and O. Kwon, "Real-time personalized facial expression recognition system based on deep learning," in *2016 IEEE International Conference on Consumer Electronics (ICCE)*, 2016: IEEE, pp. 267-268.

[156] Y. Guo, D. Tao, J. Yu, H. Xiong, Y. Li, and D. Tao, "Deep neural networks with relativity learning for facial expression recognition," in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2016: IEEE, pp. 1-6.

[157] A. Jaiswal, A. K. Raju, and S. Deb, "Facial emotion detection using deep learning," in *2020 International Conference for Emerging Technology (INCET)*, 2020: IEEE, pp. 1-5.

[158] A. Durmuşoğlu and Y. Kahraman, "Facial expression recognition using geometric features," in *2016 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2016: IEEE, pp. 1-5.

[159] B.-K. Kim, J. Roh, S.-Y. Dong, S.-Y. Lee, Hierarchical committee of deep convolutional neural networks for robust facial expression recognition, J. Multimodal User Interfaces 10 (2) (2016) 173–189.

[160] D. Sokolov and M. Patkin, "Real-time emotion recognition on mobile devices," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018: IEEE, pp. 787-787.

[161] T. Kosch, M. Hassib, R. Reutter, and F. Alt, "Emotions on the Go: Mobile Emotion Assessment in Real-Time using Facial Expressions," in *Proceedings of the International Conference on Advanced Visual Interfaces*, 2020, pp. 1-9.

[162] H. Alshamsi, V. Kepuska, H. Meng, Real time automated facial expression recognition app development on smart phones, IEEE, 2017, pp. 384–392.

[163] M. Suk, B. Prabhakaran, Real-time facial expression recognition on smartphones, IEEE, 2015, pp. 1054–1059.

[164] E. Goeleven, R. De Raedt, L. Leyman, B. Verschuere, The Karolinska directed emotional faces: a validation study, Cogn. Emot. 22 (6) (2008) 1094–1118.

[165] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *2010 ieee computer society conference on computer vision and pattern recognition-workshops*, 2010: IEEE, pp. 94-101.

[166] S. Cheng, I. Kotsia, M. Pantic, and S. Zafeiriou, "4dfab: A large scale 4d database for facial expression analysis and biometric applications," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5117-5126.

[167] M. Valstar and M. Pantic, "Induced disgust, happiness and surprise: an addition to the mmi facial expression database," in *Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect*, 2010: Paris, France., p. 65.

[168] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Proceedings Third IEEE international conference on automatic face and gesture recognition*, 1998: IEEE, pp. 200-205.

[169] Z. Zhang, P. Luo, C.C. Loy, X. Tang, From facial expression recognition to interpersonal relation prediction, Int. J. Comput. Vis. 126 (5) (2018) 550–569.

[170] I.J. Goodfellow et al, Challenges in representation learning: A report on three machine learning contests, in: International conference on neural information processing, Springer, 2013, pp. 117–124.

[171] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon, "From individual to group-level emotion recognition: Emotiw 5.0," in *Proceedings of the 19th ACM international conference on multimodal interaction*, 2017, pp. 524-528.

[172] A. Dhall, O. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon, "Video and image based emotion recognition challenges in the wild: Emotiw 2015," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015, pp. 423-426.

[173] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-pie, Image Vis. Comput. 28 (5) (2010) 807–813.

[174] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *7th international conference on automatic face and gesture recognition (FGR06)*, 2006: IEEE, pp. 211-216.

[175] X. Zhang *et al.*, "A high-resolution spontaneous 3d dynamic facial expression database," in *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, 2013: IEEE, pp. 1-6.

[176] G. Zhao, X. Huang, M. Taini, S.Z. Li, M. PietikäInen, Facial expression recognition from near-infrared videos, Image Vis. Comput. 29 (9) (2011) 607–619.

[177] S. Li, W. Deng, Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition, IEEE Trans. Image Process. 28 (1) (2018) 356–370.

[178] C. Fabian Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5562-5570.

[179] W.-J. Yan et al, CASME II: An improved spontaneous micro-expression database and the baseline evaluation, PLoS One 9 (1) (2014) e86041.

[180] A. Mollahosseini, B. Hasani, M.H. Mahoor, Affectnet: A database for facial expression, valence, and arousal computing in the wild, IEEE Trans. Affect. Comput. 10 (1) (2017) 18–31.

[181] A. Dhall, J. Joshi, I. Radwan, R. Goecke, Finding happiest moments in a social context, in: Asian Conference on Computer Vision, Springer, 2012, pp. 613–626.

[182] W. Chen, X. Xie, X. Jia, L. Shen, Texture Deformation Based Generative Adversarial Networks for Multi-domain Face Editing, in: Pacific Rim International Conference on Artificial Intelligence, Springer, 2019, pp. 257–269.

[183] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2011: IEEE, pp. 921-926.

[184] J. M. Susskind, A. K. Anderson, and G. E. Hinton, "The toronto face database. Department of Computer Science, University of Toronto, Toronto, ON," Canada, Tech. Rep, 3, 2010.

[185] Z. Zhang *et al.*, "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3438-3446.

[186] I. O. Ertugrul, J. F. Cohn, L. A. Jeni, Z. Zhang, L. Yin, and Q. Ji, "Cross-domain au detection: Domains, learning approaches, and measures," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, 2019: IEEE, pp. 1-8.

[187] D. Lundqvist, A. Flykt, and A. Öhman, "The Karolinska directed emotional faces (KDEF)," *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet,* vol. 91, no. 630, pp. 2-2, 1998.

[188] S.M. Pizer et al, Adaptive histogram equalization and its variations, Comput. Vision, Graphics, Image Processing 39 (3) (1987) 355–368.

[189] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," in *2003 IEEE International SOI Conference. Proceedings (Cat. No. 03CH37443)*, 2003: IEEE, pp. 157-164.

[190] M. Savvides, B.V. Kumar, Illumination normalization using logarithm transforms for face authentication, in: International Conference on Audio-and Video-Based Biometric Person Authentication, Springer, 2003, pp. 549–556.

[191] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Generative models for recognition under variable pose and illumination," in *Proceedings fourth ieee international conference on automatic face and gesture recognition (cat. no. pr00580)*, 2000: IEEE, pp. 277-284.

[192] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: Illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.

[193] J. Liu, Y. Feng, H. Wang, Facial expression recognition using pose-guided face alignment and discriminative features based on deep learning, IEEE Access 9 (2021) 69267–69277.

[194] W. Wu, Y. Yin, Y. Wang, X. Wang, and D. Xu, "Facial expression recognition for different pose faces based on special landmark detection," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018: IEEE, pp. 1524-1529.

[195] X. Zhu, Z. He, L. Zhao, Z. Dai, Q. Yang, A Cascade Attention Based Facial Expression Recognition Network by Fusing Multi-Scale Spatio-Temporal Features, Sensors 22 (4) (2022) 1350.

[196] N.C. Ebner, M.K. Johnson, H. Fischer, Neural mechanisms of reading facial emotions in young and older adults, Front. Psychol. 3 (2012) 223.

[197] N.C. Ebner, M.R. Johnson, A. Rieckmann, K.A. Durbin, M. K. Johnson, H. Fischer, Processing own-age vs. other-age faces: neuro-behavioral correlates and effects of emotion, Neuroimage 78 (2013) 363–371.

[198] M. Sajjad, A. Shah, Z. Jan, S.I. Shah, S.W. Baik, I. Mehmood, Facial appearance and texture feature-based robust facial expression recognition framework for sentiment knowledge discovery, Clust. Comput. 21 (1) (2018) 549–567.

[199] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 11065-11074.

[200] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136-144.

[201] M. Sajjad, I. Mehmood, S.W. Baik, Image super-resolution using sparse coding over redundant dictionary based on effective image representations, J. Vis. Commun. Image Represent. 26 (2015) 50–65.

[202] Z. Liu, L. Li, Y. Wu, C. Zhang, Facial expression restoration based on improved graph convolutional networks, in: International Conference on Multimedia Modeling, Springer, 2020, pp. 527–539.

[203] J. Yang, T. Qian, F. Zhang, S.U. Khan, Real-time facial expression recognition based on edge computing, IEEE Access 9 (2021) 76178–76190.

[204] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, 2017: PMLR, pp. 1273-1282.

[205] I. Feki, S. Ammar, Y. Kessentini, K. Muhammad, Federated learning for COVID-19 screening from Chest X-ray images, Appl. Soft Comput. 106 (2021) 107330.

[206] V. Narula and T. Chaspari, "An adversarial learning framework for preserving users' anonymity in face-based emotion recognition," *arXiv preprint arXiv:2001.06103*, 2020.