Journal of Autonomous Intelligence (2023) Volume 6 Issue 2
doi: 10.32629/jai.v6i2.631

REVIEW ARTICLE

An extensive study of facial expression recognition using artificial intelligence techniques with different datasets

Sridhar Reddy Karra[1,*], Arun L. Kakhandki[2]

[1] *Research Scholar, Visvesvaraya Technological University, Belagavi 590001, India*

[2] *Department of ECE, KLS Vishwanathrao Deshpande Institute of Technology, Haliyal 581329, India*

*** Corresponding author:** Sridhar Reddy Karra, Sridharreddy.karra1983@gmail.com

ABSTRACT

Machine and deep learning (DL) algorithms have advanced to a point where a wide range of crucial real-world computer vision problems can be solved. Facial Expression Recognition (FER) is one of these applications; it is the foremost non-verbal intentions and a fascinating study of symmetry. A prevalent application of deep learning has become the area of vision, where facial expression recognition has emerged as one of the most promising new frontiers. Latterly deep learning-based FER models have been plagued by technical problems, including under-fitting and over-fitting. Probably inadequate information is used for training and expressing ideas. With these considerations in mind, this article gives a systematic and complete survey of the most cutting-edge AI strategies and gives a conclusion to address the aforementioned problems. It is also a scheme of classification for existing facial proposals in compact. This survey analyses the structure of the usual FER method and discusses the feasible technologies that may be used in its respective elements. In addition, this study provides a summary of seventeen widely-used FER datasets that reviews functioning novel machine and DL networks suggested by academics and outline their benefits and liability in the context of facial expression acknowledgment based on static replicas. Finally, this study discusses the research obstacles and open consequences of that well-conditioned face expression recognition scheme.

Keywords: artificial intelligence; deep learning; facial expression recognition; symmetry; over-fitting; insufficient training data

ARTICLE INFO

Received: 14 May 2023
Accepted: 19 June 2023
Available online: 15 August 2023

1. Introduction

Day-to-day human interrelationship relates to transparent and potent emotions. People's faces are the clearest indicator of their inner state. Even though (FER) is Complex and time-consuming, it has many valuable applications, particularly in healthcare[1–3], emotionally intelligent robots, and human-computer interaction. Even though FER has become more efficient due to upgrades in technologies, still achieving determinacy is difficult[4]. Anger, joy, sorrow, fear, and surprise are the five most common human emotions. Additionally, "hatred" was included as a preliminary emotion[5].

Illustration, i.e., an obstacle on the face like the hand, aging, and sunglasses, can all have a significant impact on FER's accuracy, making it a challenging task overall. In order to attain the highest degree of precision required for FER modeling, the expertise of this ground is taking these aspects into account[6]. Facial emotions involve disgust, fear, joy, grief, pleasure, and surprise. The hatred was subsequently added to the emotion list. To achieve facial emotion recognition, there is a fundamental inaugural step divided into three

primary phases. In the first preparatory step, facial characteristics are extracted from the full frame of the video. Some examples of facial characteristics include the eyebrows, nose, mouth, and chin[7]. In the second stage, more detailed characteristics are isolated from various parts of the face. The second level also eliminates additional descriptive elements from various parts of the face. In the end, a classifier is prepared with the use of training data until assigning labels to the Emotions, as a result of its applicability in areas such as intelligent robotics, medical treatment, IoT, fatigue, security surveillance, and criminal psychological investigation. Facial Emotion Recognition using Computer vision has become the subject of a large number of academic studies[8,9]. Because people share their lives online through videos and photographs, it is essential that the latest technology is used upon people's emotions to ensure user-friendliness and extreme user happiness.

Emotions are just a mental phase or condition that people link with specific emotional states. These feelings are frequently crooked with behaviour, notion, personality, stances, and drives. Under a variety of psychological conditions, these feelings can be reduced to two distinct categories: positive (pleasure) and negative (displeasure). Such feelings bend people's mind intellectually, which result in a person's behavioural change over time. Humans deal with these feelings in a variety of ways, including behavioral responses, psychological states produced by external events or by the presence of another person, and the individual's own interpretation of the procedures. Since emotions are based on a delicate collection of behaviours, it is difficult for humans to recreate them intentionally. Researchers use their own rendition of emotions and theories. Since there is a great deal of difference across studies and no overarching conclusion can be made, studying human facial expressions is difficult.

1.1. Scope of the survey

Though numerous books and articles have been written previously about FER, many scholars are continuously looking forward to working on exceptions and results. Approaches like (SVM), Artificial Neural Networks (ANN), and decision trees have been the primary focus of the survey[10]. Researchers in the same discipline occasionally investigated the DL approaches[11]. Accordingly, we conducted a relative study of surveys about FER and reported our findings in this research. For instance, Hemalatha and Sumathi[12] reviewed many techniques for detecting faces, extracting facial features, and classifying FER, but they failed to provide a thorough comparison of approaches they calculated into the account or the datasets they employed. Thereafter, the authors in also provided a review on FER, but they ignored datasets applicable to emotion recognition[13]. Chengeta and Viriri[14] also conducted a literature review on conventional feature extraction methods, including principal component analysis (PCA), Linear Discriminant Analysis (LDA), and Locally Linear Embedding (LLE), subsequently alleged an ensemble classifier. The most pioneering method in FER is deep learning, yet they didn't include that in their comparison. Again, Baskar and Kumar[15] don't provide enough detail to illustrate the several DL methods that exist.

Recent research into DL-based FER methods may be found in, which provides in-depth reviews but no context for FER. As a result, we conduct a comprehensive review of existing FER databases, face identification, facial difficulties, and present concerns in FER in the proposed survey[16]. The goal is to deliver a comprehensive review of all cutting-edge systematic methods to FER, which will be useful to researchers interested in investigating deeper into the topic.

1.2. Research aids

- In this study, we conduct the review of a work on FER, concentrating on machine learning and deep learning, datasets, and methodologies that have been employed to categorize human emotions. You may quickly summarise the paper's contributions with the following points.

- The study provides a comprehensive overview of FER techniques and data sources. Then, we focus on the cutting-edge techniques that have been implemented for FER, and we evaluate these techniques side by side.
- The study proposes a categorization scheme for FER approaches that utilize face detection, feature extraction, and emotion labeling.
- At last, this effort outlined the questions that still need to be answered and the obstacles to research in the FER.

2. Incentive

FER schemes have vast applicability in numerous sectors, including computer interfaces, healthcare, and public advertisement. However, real-world photos and videos present enormous difficulties for face expression analysis because of the delicate and fleeting motions of the spotlight humans and the complicated, buzzing atmosphere at the backdrop[17]. There are three key problems created by light fluctuation, and they widely impact the effectiveness of the FER system. The cutting-edge methods used in FER were brought into the laboratory but didn't work in real-world practice.

A primary step in a FER system for emotion detection is extracting features. A FER scheme reliably functions when the extracted features minimize within-class differences and maximize between-class differences. An effective and precise recognition procedure may be ensured by a feature representation. Geometric and appearance-based methods[18] are used for identifying featured base facial expressions.

The second crucial phase of a FER scheme is the categorization of feelings. To classify an input face with one of the expression labels, a classifier is trained using the feature vectors. High-dimensional feature, which affects the performance and speed of FER[19,20]. Some systems have only dealt with static images of a single person's face; however, there are methods that can deal with recordings from dynamic settings. One may classify a video-based FER system as frame-based and sequence-based. Investigators have tried to address the issues in the real-world requests of FER systems by training classifiers with current classification techniques, but there is no complete, trustworthy, and robust classification approach exists yet.

3. Facial Expression Recognition (FER)

Systems that use existing artificial intelligence techniques, especially neural networks, to categorize basic human emotions are the heart of this research topic. There are three main architectures in FER, and they are pre-processing, extraction of features, and classification. The layout is shown in **Figure 1**.

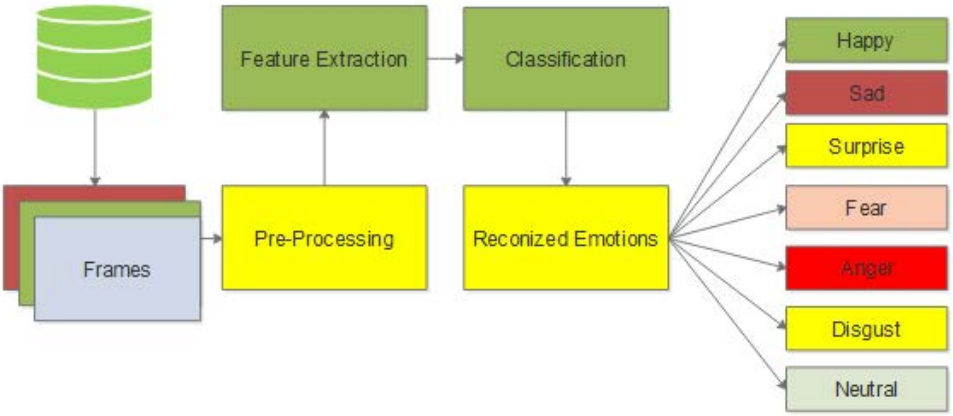

Figure 1. Facial emotion classification process.

Pre-processing: experiments in deep learning for image and signal processing frequently involve data processing to boost results. Facial orientation, gray-to-color picture transfer, 2-dimensional, noise-canceling, adaptive filtering, image sharpening achieved by sharp masking, and information expansion are all covered here. Images acquired in the actual world it can be faced in any sum of directions and come in a wide variety of sizes and degrees of visibility.

Feature extraction: facial feature extraction from the source image or video is the first stage. The major focus of the current body of research is on identifying distinguished traits among the six core expressions. As part of the extraction procedure, various intricate attributes are generated to elucidate the manifestation of morphological or textural changes to the face. The SDM uses texture characteristics such Spatio-Temporal Texture Map (STTM) and Local Background Projection (LBP) to characterize alterations to the texture of face organs. Shape and texture distinctions are made clear in a number of works thanks to the features. The extracted features were utilized to label facial expressions.

Classification: this is the last step of the FER procedure when the labels are applied to the mapping units that represent emotions. Classification is the stage when the retrieved functions are used to train a classifier using various techniques. Without the need for hand-made features, it can be learned accurately feature by training in deep network construction with millions of structures.

4. Terminologies

To complement the theoretical foundation of FER knowledge, we first offer some associated terminology before analyzing the methodologies of FER. The conversion of facial movement into emotion is the subject of the Facial Action Coding System (FACS). Classification of categories based on expressions are as follows Basic Emotions (B.E.s), Compound Emotions (C.E.s), and Micro Expressions (M.E.s). These ideas and phrases are from the foundation of existing research on FER.

4.1. Facial Landmarks (F.L.s)

Figure 2 depicts a person whose facial landmarks have been visually highlighted. These include the alae of the nose, end of the eyebrow, and angle of the mouth. The positions of the F.L.s near the facial structure and contour, which record changes in the face caused by the subject's head position and expressions. Establishing a face's feature vector by point-to-point correspondences of landmarks.

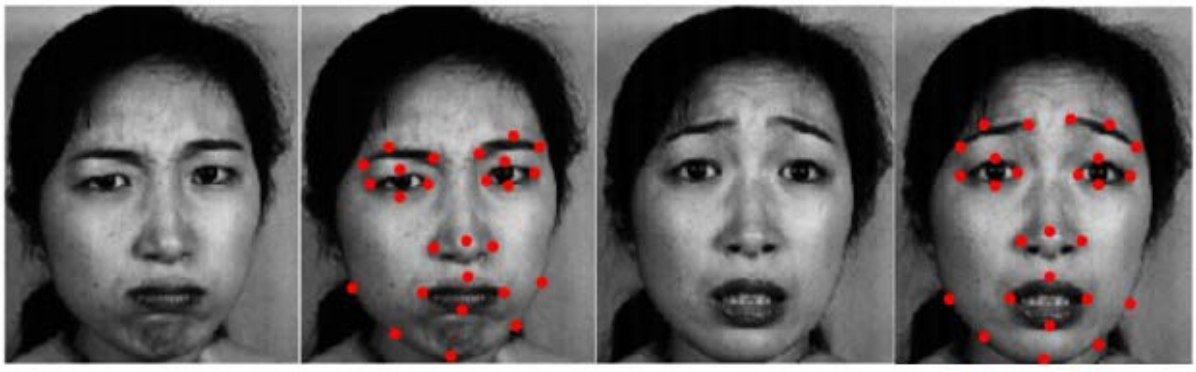

Figure 2. Sample of facial landmarks (from JAFFE dataset).

4.2. Facial Action Units (A.U.s)

Each facial expression encodes a certain emotion through the coordinated actions of 46 facial action units. Some instances are shown in **Figure 3**. Through analysis of observed facial A.U. pairings, the FER system categorizes different types of expressions. A picture can be categorized as an "Awed" emotion if it is labeled with 1, 2, 5, and 25 A.U.s, for instance.

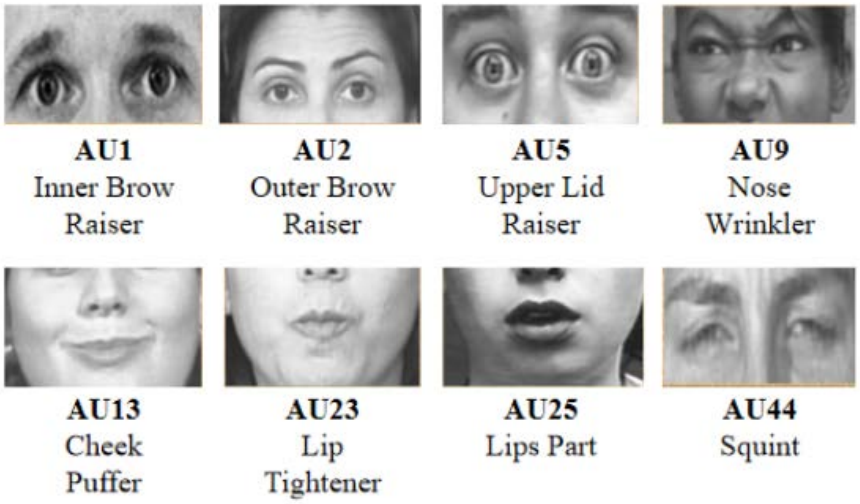

Figure 3. Around instances of AU-Coded facial expression database (from CK + dataset).

4.3. Facial Action Coding Scheme (FACS)

Renowned use biofeedback and observation to illustrate the link between muscle activity and emotional expression. They start by segmenting the entire face into numerous A.U.s based on anatomical factors and then analyze the traits of each A.U. Individually. **Table 1** displays the average A.U.s for both the basic and complex emotion groups. The FACS system provides an instance standard for describing the muscle movements associated with different facial expressions.

Table 1. Ideal A.U.s seen in rudimentary and multiple emotion classes.

Category	Grouping	A.U.s	A.U.s
Happily disgusted	Awed	10, 12, 25	1, 2, 5, 25
Sadly fearful	Appalled	1, 4, 15, 25	4, 9, 10
Sadly angry	Hatred	4, 7, 15	4, 7, 10
Angry	Fearfully disgusted	4, 7, 24	1, 4, 10, 20, 25
Surprised	Angrily disgusted	1, 2, 25, 26	4, 25, 26
Disgusted	Disgusted surprised	9, 10, 17	1, 2, 5, 10
Happily sad	Happily fearfully	4, 6, 12, 25	1, 2, 12, 25, 26
Happy	Sadly disgusted	12, 25	4, 10
Sad	Fearfully angry	4, 15	4, 20, 25
Fearful	Fearfully surprised	1, 4, 20, 25	1, 2, 5, 20, 25
Happily surprised	Angrily disgusted	1, 2, 12, 25	4, 10, 17
Sadly surprised	Category	1, 4, 25, 26	-

4.4. Basic Emotions (B.E.s)

There are six primary human emotions discussed: joy, shock, sorrow, rage, contempt, and fear. Typically, these six B.E.s are used to designate datasets associated with FER.

4.5. Compound Emotions (C.E.s)

Two simple emotions can be combined to form a compound emotion. There are a total of 22 emotions discussed: 7 primary emotions (6 primary and 1 neutral emotion), 12 compound emotions usually uttered by humans, and 3 supplementary emotions.

4.6. Micro Expressions (M.E.s)

Micro expressions are sudden and tricky facial expressions that people express unavoidably. That expression tends to show a person's honesty and perspective expressions for a short retro of time. Microexpressions are fleeting, often lasting between 1/25 and 1/3 of a second. Microexpressions are frequently applicable in forensics and psychology.

5. Datasets

5.1. Basic information

Initial research often employs static 2D photos; afterward, FER research makes use of dynamic 2D video sequences because they can be used to recognize the expression in several dimensions. At the same time, 2D data makes it heavy to analyze face-depth information. Posture and lighting can have a major impact on productive FER works. The micro-behaviors and structural differences of the human face are also handled by using several 3D-based datasets. Some researchers prefer to employ "in the wild" datasets versus "in the lab" in order to address applications practically elegant. Also, certain datasets are tailored specifically to the study of individual facial expressions, making them ideal for training classifiers that deal with this domain. In this part, we deliver a brief introduction to many widely used datasets pertaining to FER. **Figure 4** depicts some example pictures.

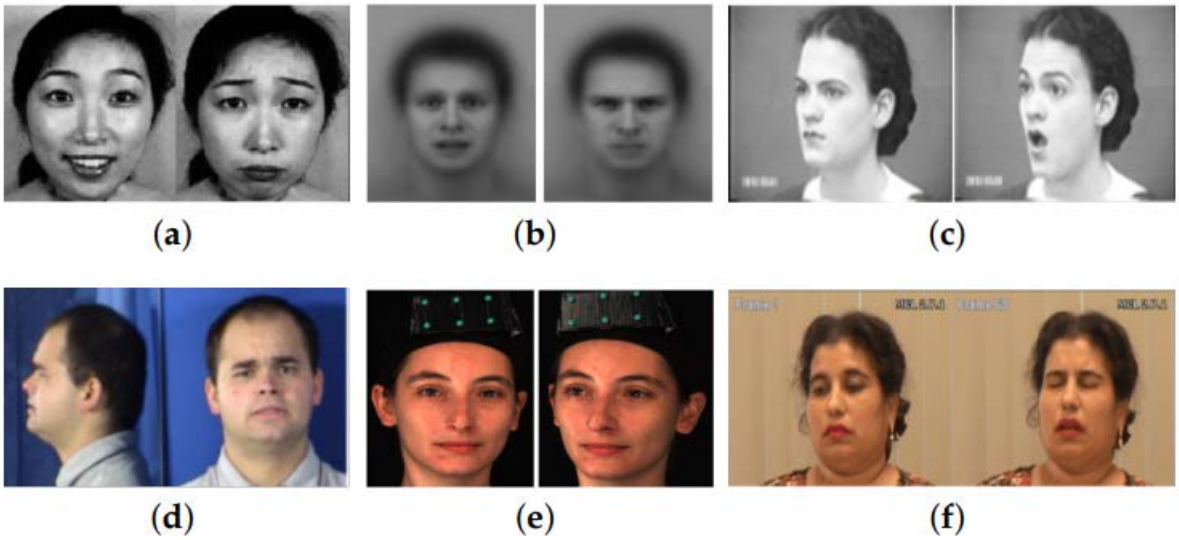

Figure 4. Instances of six illustrative datasets connected to FER. **(a)** JAFFE; **(b)** KDEF; **(c)** CK+; **(d)** MMI; (e) MPI; (f) UNBC.

5.1.1. Japanese Female Facial Terminologies (JAFFE)

Ten Japanese women squatted for 213 photos in the JAFFE dataset[21], which depicts seven different facial expressions (six fundamental emotions and one neutral). Sixty Japanese participants score each picture on a scale of one to six across a range of emotions. The dimensions of the original photos are 256 on 256.

5.1.2. Extended Cohn–Kanade (CK+)

The JCK+ dataset[22] expands upon the C.K. dataset by adding 593 video segments and immovable photos representing seven facial expressions. Both the photographs and the films are staged in a laboratory environment. Among the 123 participants, the median age is 19. Both the width and height of the picture are 640 pixels, the depth is 480 pixels, and the greyscale has 8 bits of accuracy.

5.1.3. Compound Emotion dataset (C.E.)

5060 photos hold the C.E. dataset[23]. These images represent 22 B.E.s and C.E.s from 230 participants, with a moderate age of 23 and a wide range of racial backgrounds. The close-up facial view is minimized without sunglasses. Male participants are expected to have a clean shave, and everyone is stimulated to keep their eyebrows fully grown. Pictures are color images 3000 by 4000 pixels in size.

5.1.4. Impulsive facial action dataset (DISFA)

There are 130,000 stereo movies from 27 people of diverse gender and race in the DISFA dataset[24]. High determination of 1024 × 768 pixel photos are captured, and the strength of A.U.s are assessed for each frame of each movie (0–5 scale). The sum of 66 face landmarks is marked for each image in the collection.

5.1.5. MMI Facial expression dataset

More than 2900 videos and high-resolution immovable photos from 75 participants are included in the MMI Facial Expression collection[25]. The videos include fully annotated by every single A.U.s. The original dimensions of the face photos are 720 px by 576 px.

5.1.6. Binghamton 3D facial appearance (BU-3DFE)

The purpose of BU-3DFE[26] is to advance our understanding of human behavior via studies of 3D faces and facial expressions. There are a total of 100 people in the sample (56 women and 44 men) from a wide range of racial and cultural backgrounds. The ages of those samples range from 18 to 70. There are a total of 6 feelings included in the data set. In this dataset, each person has 25 unique 3D facial emotion models, and all of them are connected to a collection of 83. Each of the original face photos measures 1040 pixels wide by 1329 pixels tall.

5.1.7. Binghamton-pittsburgh 3D dynamic spontaneous

Forty-one people have been organized into BP4D-spontaneous[27]. Emotion elicit protocols are created with the intention of raising strong feelings from study participants. Interviews and a set of steps were designed in eight distinct tasks to elicit eight distinct feelings. Each activity has its own set of 3D and 2D video guides. In the meanwhile, the metadata contains manually annotated action units (A.U.), impulsively tracking results in head position, and a set of 2D and 3D face landmarks. Each of the original face photos measures 1040 pixels wide by 1329 pixels tall.

5.1.8. Large MPI facial appearance database (MPI)

A validated collection of expressive and linguistic facial expressions are obtained in the MPI Facial Appearance Database[28]. There are a total of 55 unique phrases from 19 German individuals in the sample (ten females and nine males). Method acting protocol is used to elicit performances, and the technique ensures both clear and natural observation. The entire Facial expression is viewed in three distinct ways: three times at lower intensity, thrice at higher intensity, and from three different bursting angles. With the specified frame annotation, we can generate both stationary and powerful versions of the dataset. To verify the authenticity of the video sequences and the background situations, trails are provided under two conditions.

5.1.9. Karolinska Directed Expressive Face (KDEF)

There are 4,900 pictures of people's facial terms of different emotions in the KDEF dataset[29]. There are 70 unique individuals in the dataset, and for each one, there are five vantage points showing seven distinct facial expressions. The source image was 562 × 762 pixel.

5.1.10. NVIE dataset

More than a hundred people's facial expressions were subjected to both instinctive and immovable poses and were captured under three distinct lighting conditions to create NVIE[30], an infrared facial expression dataset. Images of the acme expression, with and without spectacles, are included in the posed dataset. It has labels for the six different expressions, the level of expression, and the Arousal-Valence scale.

5.1.11. CMU multi-pie database

Studying face recognition in different lighting conditions requires the CMU Multi-PIE database[31]. There are 337 subjects represented by almost 750,000 photos collected across 15 camera angles and 19 lighting setups over four separate recording sessions. There are 39 and 68 feature points on the AAM-style labels.

5.1.12. Oulu-CASIA NIR-VIS record (Oulu-CASIA)

It consists of 2,880 picture sequences from 80 participants aged 23 to 58 included in the Oulu-CASIA NIR-VIS database[32]. All facial expressions are taken in the anterior direction: normal, weak, and dim lighting. Participants were instructed to mimic an expression depicted in a series of photographs. With a frame of 25 fps and a resolution of 320 × 240 pixels, the imaging technology is more than capable of capturing high-quality images.

5.1.13. FER2013 face dataset

Kaggle competition is delivered by the FER2013 dataset[33]. There are a total of 35,887 greyscale photos of faces in the dataset, comprising 28,709 training sets, 3589 verification sets, and 589 test sets. Each image consists of a greyscale in the dataset that is 48 pixels by 48 pixels in size. The samples are classified into seven main emotions: anger, contempt, fear, happiness, neutral, sadness, and amazement. Datasets are implemented in real life, and each sample is very different in terms of age, facial orientation, and other characteristics.

5.1.14. Gemep-fera

Ten actors were recorded expressing 18 distinct emotions using a variety of vocal tones, facial expressions, and body language for the GEneva Multimodal Emotion Portrayals (GEMEP)[34]. More than seven thousand audio-video depictions of emotional states are included in this corpus, including 18 states (including delicate emotions examined infrequently) and performed by ten trained actors under the direction of a director.

5.1.15. Acted Facial Terms in the Wild Dataset (AFEW)

This paper presents AFEW[35], a dynamic corpus with spontaneous expressions, different head postures, occlusions, and illuminations all represent the actual world. Seven different types of feelings are used to classify the samples. Train (773 samples), Val (383 samples), and Test (653 samples) are the three data partitions independently that makeup AFEW, and each contains data that is guaranteed to come from different movies or performers. The SFEW[36] was created by cherry-picking frames from AFEW, which split into three collections: train and test (100 samples) (372 samples). The seven different expressions on the face are clearly labeled.

5.1.16. Real-world Affective Database (RAF-DB)

There are 29,672 photos of real human faces in RAF-DB[37], which results in a large-scale facial expression database. About 40 annotators have labeled each image individually using seven basic emotion classes, twelve compound emotion classes, five precise landmark locations, and thirty-seven automated landmark sites. The images in this collection span a wide range of ages, sexes, and ethnicities and feature individuals in various positions, lighting setups, occlusions, and post-processing techniques. The Real-Faces Multi-Label (RAF-ML) dataset[38] is a multi-label collection of face expressions taken from the real world. Each 4908 given real-world photo has a 6-dimensional expression distribution vector and the coordinates of

notable landmarks in the scene. In particular, we utilize 315 highly-trained annotators to guarantee sufficient unique picture annotations.

5.1.17. GENKI-4K dataset

The MPLab GENKI dataset[39] is an extensive collection of photographs that include human faces in various poses and settings, with people of all ages, genders, and races represented. The GENKI-4K subset was created explicitly for smile recognition, and it contains 4,000 photos of faces manually annotated by human coders to indicate whether corresponding faces were smiling. Splits into intersecting subsets respectively own labeled and described.

5.1.18. The UNBC-Mc master shoulder pain expression collection dataset

It is hoped that this dataset[40] will help advance research in pain and increase current data sets in this area. The dataset consists of 200 video sequences with natural facial expressions, 48,398 FACS-coded frames, pain ratings observer assessments, and 66 points from the AAM landmarks.

5.2. Pre-processing

It is the primary phase in FER and also plays a vital role in ML technique. Pre-processing is a crucial step. Raw data analysis without any sort of filtering might result in inaccurate findings. For this reason, ensuring the data will be better before extracting relative characteristics. The most widely used and fruitful pre-processing methods from the literature review are presented here.

5.2.1. Face detection

Facial landmarking is the first phase in most FER systems. This method determines the region of interest (ROI) of an input picture for further processing by the FER system. The Viola-Jones face detector[41] from 2004 was utilized in the majority of the studies.

As a machine learning-based method, the Viola-Jones face detector uses a cascade function that is learned using both “positive” (pictures containing faces) and “negative” (images without faces). This technique employs the Haar features, which are applied to each training picture in order to determine the optimal threshold to distinguish faces as positive or negative detections.

The Dlib package[42] and the Multitask Cascade Convolution Neural Network (MTCNN)[43] are two examples of alternative face detectors that have limited application. An example of a face being discovered from the CK+ database and its 68 landmarks using the Dlib library is displayed in **Figure 5**.

Figure 5. Sensed face from the CK+ record using the Dlib reference library.

The MTCNN system analyses three faces each time using a different CNN to get an accurate detection. **Figure 6** depicts the five discovered facial landmarks of a CK+ database using MTCNN.

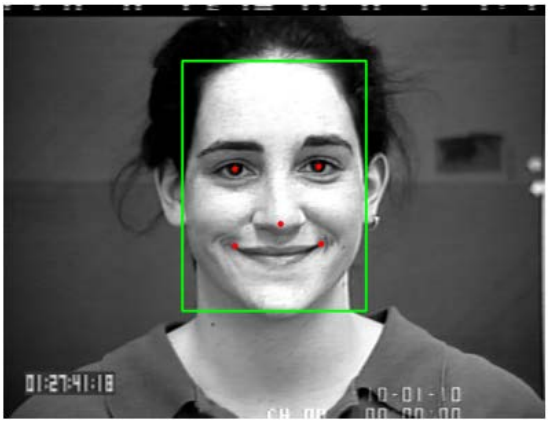

Figure 6. Noticed the face from the CK+ file using MTCNN.

5.2.2. Geometric transformations

Certain re-examined work applied geometric adjustments to faces that are not identified under ideal settings by using facial landmark, which is outputted by a face detector, resulting in rotation correction in images. In general, the examined studies took into account that two facial landmarks make a zero-degree angle in the horizontal axis by aligning the face. After applying a rotation transformation to a rotated face, the two landmarks on the face are brought into horizontal alignment until the angle generates zero degrees, completing the face alignment process. Rotational adjustments made to a face in the CK+ record are displayed in **Figure 7**.

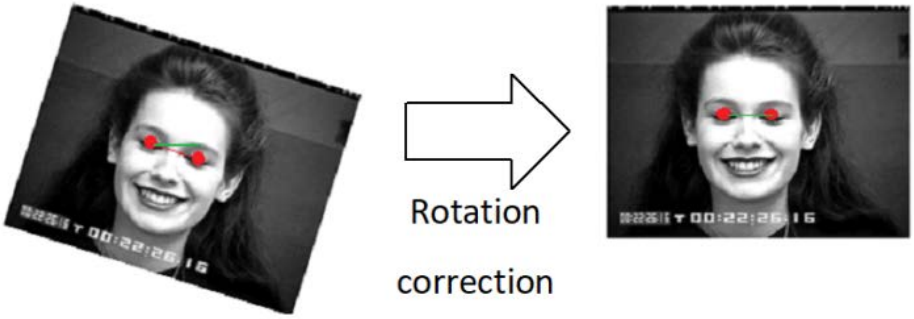

Figure 7. Revolution alteration on a face from the CK+ record.

The scale problem is caused by the fact that the size of the region of interest (ROI) varies based on the distance at which a face was detected. The evaluated works include a reduction in the size of each ROI to a constant level since this is vital to the continued operation of the FER system with consistent data (spatial normalization).

Generally speaking, ambient noise is the most disruptive factor in the pinpointed ROIs. The initial area of interest (ROI) photos can be improved by training a classifier to ignore the backdrop. However, some researchers sought to crop the ROI to filter the backdrop, while others just ignored this sort of noise. Using the face detector's output bounding box with the face detector's facial landmarks is the most typical approach. The ROI dimension collected by the face detector may be reduced by allowing the filter out of background noise. For example, the background of a CK+ database face can be removed, as seen in **Figure 8**.

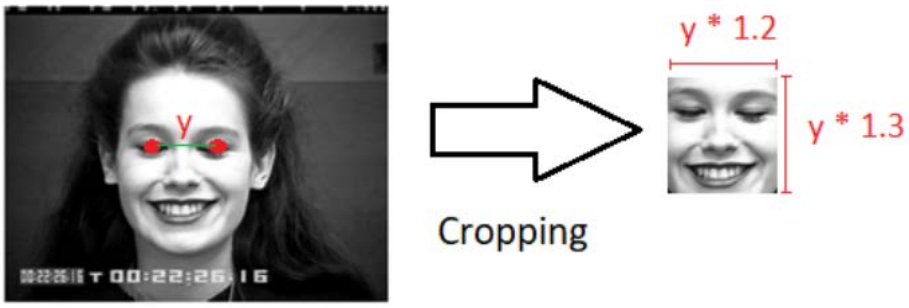

Figure 8. Background elimination using the detachment among eyes of a face from the CK+ record.

5.3. Image processing

In certain cases, preparing picture data, it's not only related to proper transformed ROI. The majority of the examined FER systems included the following image-processing approaches to highlight important features that would ultimately be used in the classifier.

5.3.1. Smoothing

For many image processing tasks, smoothing is frequently needed. By applying a smoothing filter to an image, one can isolate the important details while eliminating the background clutter. Thus, the smoothing process can add stability to the data before analysis. A bilateral filter[44] or a Gaussian filter[45] are two of the most frequently used methods in the evaluated publications for picture smoothing.

5.3.2. Histogram equalization

Histograms are graphs of an image's intensity data. The accuracy of facial expressions is expected to drop because of the varying lighting conditions. Underexposed areas of a face can have their intensity normalized by histogram-based techniques in Computer Vision. This approach enhances contrast, draws attention to key face characteristics, and minimizes the impact of varying lighting. However, this results in accelerating the foreground sound and background noise[46]. Histogram equalization was a common topic in the review literature that employed this pre-processing technique (HE). The results of a few searches on the CK+ database using HE are displayed in **Figure 9**.

Figure 9. Consequences in some faces of the CK+ record using HE.

5.3.3. Data augmentation

Most datasets used for emotion identification are few, which is problematic for machine learning classifiers. training on a few data leads to overfitting[47], which is a prevalent issue in machine learning models. A model is well at categorizing data from the training set but suffers a significant reduction in performance while classifying data from independent datasets. Overfitting can be detected during model training because the model performs decently on the training data but poorly on the validation data. This issue is generally a result of employing insufficiently sized datasets throughout the training process; however, DA is one technique to combat this.

5.3.4. Principal component analysis

(PCA)[48] is a technique for reducing the dimension of a large number of characteristics while retaining the majority of their information. To Simplify data, huge number of variables should be reduced. This approach may be used to eliminate unnecessary face characteristics in a FER system, which improves the system's computing efficiency.

5.4. Feature extraction

After initial processing is complete, the desired highlighted characteristics may be extracted. Features of the face are typically used as input in traditional FER systems. Since the accuracy of the system was achieved by the dignity of characteristics, numerous feature extraction algorithms have been industrialized in Computer Vision. The following are the most frequently used feature extraction methods across the surveyed literature.

5.4.1. Local binary patterns

One of the most effective techniques for texture processing is LBP[49]. The purpose of this method is to evaluate a central pixel in relation to its surrounding 3-by-3-pixel areas. The value "1" is assigned if the neighbor pixel value is larger than or equal to the value of the central pixel, while the value "0" is assigned otherwise. The resultant binary code may then be used to assign a pixel. **Figure 10** depicts a typical instance of this procedure.

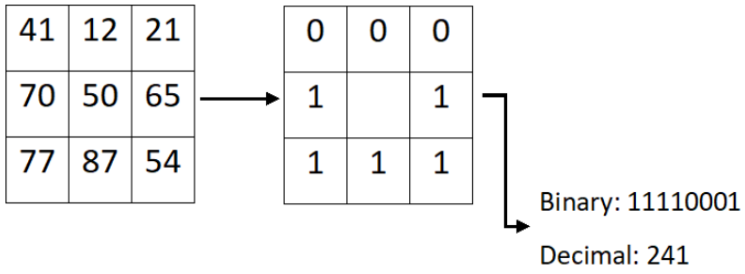

Figure 10. LBP operation.

LBP can highlight key facial features for emotion recognition in FER systems, including the eyebrows, eyes, nose, and mouth. Although, it is affected by image noise when processing a region with almost uniform intensity, as it is a method reliant on intensity alterations. See **Figure 11** for an illustration of LBP's use in feature extraction.

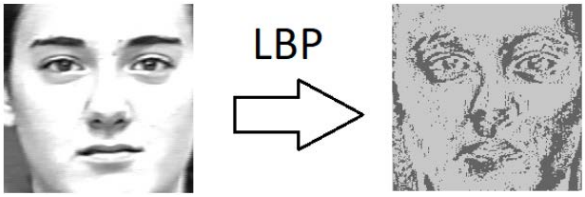

Figure 11. LBP extraction using a face from the CK+ record.

5.4.2 Optical flow

The purpose of is to evaluate the amount and direction of motion, hence the approach can only be used in the video sequence. This method basically determines the change in position between two frames pixel-by-pixel. That's why it gives you a vector that details how the pixels in the image changed between the first and second snapshots. This technique, however, is very sensitive to noise and occlusions, and its success depends on the quality of the initial tracked features and their evolution over time.

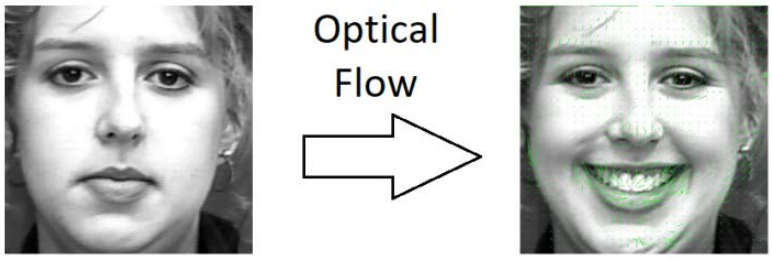

Figure 12. OF of an arrangement from the CK+ database.

Since there is a clear motion in the face can be calculated anytime someone transitions from a neutral expression to a peak facial expression, this approach can be efficiently applied in a FER system. Using a CK+ record, **Figure 12** shows how this strategy may be applied in a FER system.

5.4.3. Active appearance model

The computer vision algorithm active appearance model (AAM)[50] is responsible for matching the object's form and appearance to a fresh picture. This means it isolates only the face of the picture, and other backgrounds are removed. The AAM's shape data is utilized to calculate a set of optimal parameters that brings out the characteristics of the face. Images that vary in an instance, emotion, and lighting were left out of the training set, although the approach is still quite sensitive to these factors. The procedure is illustrated in **Figure 13**.

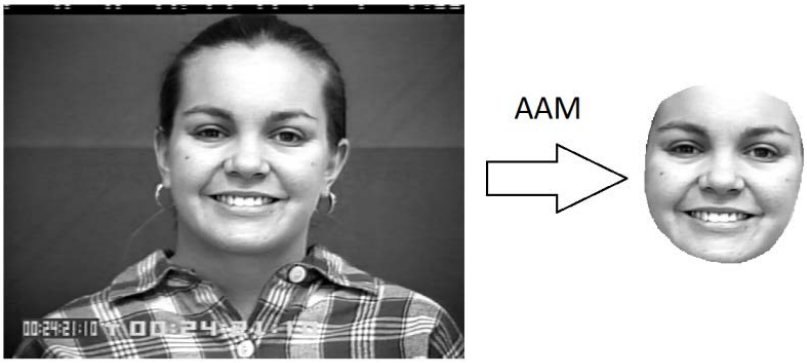

Figure 13. AAM shape approximation in a face from the CK+ record.

5.4.5. Facial animation parameters

The Facial Parameters[51] show 66 different ways in which the feature points can be moved and rotated, which is relative to the default location of the face. Movements of the face and muscle movement are the basis for FAPs. To depict emotions, they reflect the whole range of fundamental facial movements. As a second definition, FAPs are bounded as the average distances between different facial features. However, the FAP extraction is reactive to noise, such as display setting, which leads to small errors in the facial region. Extracting FAPs from the CK+ database is shown as an example in **Figure 14**.

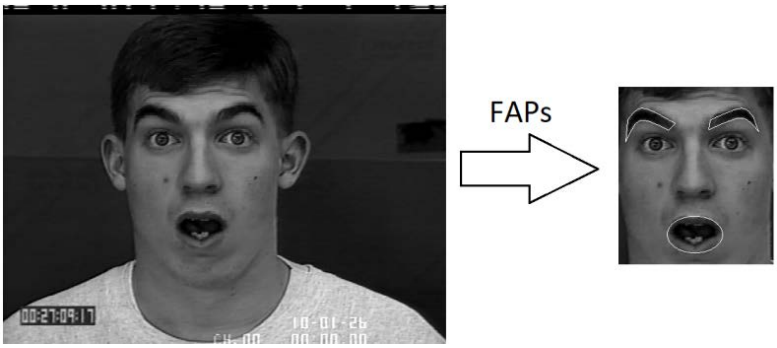

Figure 14. FAPs extraction using the CK+ record.

5.4.6. Gabor filter

Information about textures is represented with the help of the Gabor filter. Despite being unfazed by intense lighting, it is able to make a defining choice about orientation and scale. Specifically, it can extract fine-grained local modifications from a picture together with frequency, location, and orientation data. However, this approach has limitations because of the enormous computational cost incurred when dealing with Gabor feature spaces, making it impracticable for use in real-time settings. Real-time performance requires using simplified Gabor features, which are light-sensitive. As seen in **Figure 15**, Gabor features were used to create a face map from the CK+ record.

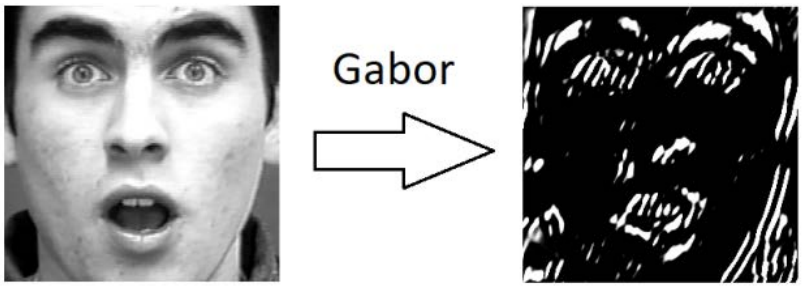

Figure 15. Feature map of Gabor from the CK+ record.

5.4.7. Scale-invariant feature transform

The computer vision approach for identifying and characterizing image-level features is the (SIFT). SIFT features are blurry and affine-change sensitive, yet they are invariant to uniform illumination vicissitudes. This durability stems from the image's translation into a vast set of feature vectors, each of which is invariant under the aforementioned requirements. Facial characteristics such as the eyes, nose, and mouth can be identified in FER systems using the Simplified Iris Feature Transform (SIFT) technique. This feature extraction approach has been included in several studies to determine facial feature motion. Extracting local characteristics from a CK+ database face is demonstrated in **Figure 16**.

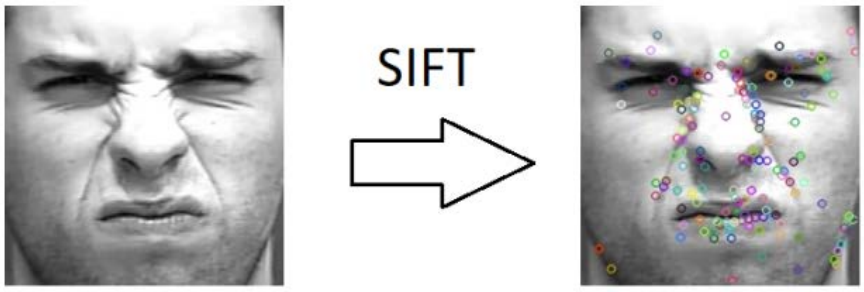

Figure 16. SIFT structures of a face from the CK+ folder.

5.5. Classification/regression

Predicting a label from an input picture or set of data is the job of a classification model. The function of regression is to establish the link between the reliant variable and the independent ones. Both are in existing work in that classification is a frequently used method. These are the most popular classification and regression algorithms.

5.5.1. Convolutional neural network

Due to their versatility in solving a wide variety of image classification problems, CNNs find widespread use in Computer Vision. Since CNNs are able to discover and recognize fundamental patterns that are too complicated for the human eye, they can even outperform humans in certain of these situations. CNN uses multiple hidden layers to break down an input picture into features. These traits are used in a classification setting, often using a Softmax function that selects the class with the highest probability from a given distribution.

It's notable that various issues require different CNN models to ensure high-accuracy categories due to overfitting and underfitting. When applied to data that is not included in the training set, its classification accuracy suffers significantly. This is known as underfitting, and it occurs when the model fails to accurately predict outcomes on both the training set and novel data. This issue can be answered in a few different ways:

(1) Adding extra layers to the model.
(2) Dropout layers are layers added to a model to prevent it from memorizing patterns during training.
(3) Adjusting training-time variables such as epochs, batch size, learning rate, and class weight.
(4) Adding more samples will train on additional data.
(5) Transfer Learning (T.L.) can be used when an insufficient amount of data is accessible locally; this is a typical issue with publically available databases for emotion identification. In TL, utilize the predefined model, which is trained on a large collection of data, and then it is fine-tuned with a smaller database for a specific classification task.

Promising outcomes for Deep Learning based classifiers were found in a large percentage of the examined papers that employed this method.

5.5.2. Support vector machine

The SVM is commonly applied for classification and regression. SVM models are spatial representations of data in which the characteristics associated with each class are separated by a sharp, as large a gap as feasible. Next, the input features are projected onto the same space, and class membership is predicted according to the features where they drop. This predictive map is built during the training phase. This classifier shines when confronted with complicated nonlinear data, and it also has the advantage of being resistant to overfitting. However, it will not give a complete solution with big datasets, and also it consumes a huge amount of time to compute, and it's very tough to choose the proper kernel function for tuning.

5.5.3. K-Nearest neighbour

K-nearest neighbor (KNN) is a classification and regression approach that operates on an instance-by-instance basis. Vectors with characteristic multidimensional space in training data are labeled with classes. During the training phase of the algorithm, these feature vectors and their associated classes are simply stored. The input feature is anticipated in the classification phase by associating them with the class that closely matches the input in terms of the characteristics that are being predicted. Average distance measures like Euclidean distance (E.D.) and Hamming distance to determine which characteristics are closer to the input. This classifier's strong points are its ease of use and quick training time. However, it has poor performance on high-dimensional data, takes up a lot of storage space, and has slow testing speeds and sensitivity to noise. A

further issue with the classification and regression method will produce misleading results if the classes are not evenly distributed. Fixing this issue can be done in part by adjusting class weights.

5.5.4. Naive bayes

Based on the Bayes theorem, the Naive Bayes classifier is a part of the machine learning classifier, which suspects strong self-reliance between features. The following equation describes Bayes's theorem:

$$P(A|B) = \frac{P(B|A)(A)}{P(B)} \quad (1)$$

Assuming *B* is true, we can use this equation to determine the probability that *A* will occur. The Naive Bayes classifier, for example, in a FER system will not correlate features while generating a prediction since it assumes that features are independent. This is problematic since there are clear associations between the parts of the face used to indicate emotion, such as the lips and the eyes, in the shocked look. The benefits of this classifier, however, lay in its ease of installation and ability to scale to massive data sets. In text classification issues, these classifiers are often used.

5.5.5. Hidden markov model

A Hidden Markov Model (HMM) helps to conclude a series of unknown variables from a series of known variables. Example: using facial expression recognition (FER) to infer someone's happy feeling (hidden variable) from the presence of a grin (observed variable). This classifier's merits are its ability to model free-form features extracted from observations, its flexibility in combining several HMMs to categorize large datasets, and its capacity to incorporate prior information. On the other hand, this classifier is both computationally costly and conflicts with overfitting.

5.5.6. Decision tree

The Decision Tree (D.T.) is analogous to a tree diagram of a process flowchart. Classification "leaves" are the subsets of the database that remain after all possible divisions have been generated using a D.T. This classifier's advantage includes its simplicity, ability to work with high-dimensional data, and ability to learn nonlinear correlations between data points. However, overfitting is a major drawback of this classifier since it might keep splitting apart until it learns to recall the training data.

5.5.7. Random forest

For classification purposes, Random Forest (R.F.) is equivalent to a collection of D.T.s working together as an ensemble classifier. Each D.T. generates predictions, and a vote superiorly decides the final prediction; Overfitting is reduced compared to a single D.T., and the bias is decreased by using a predicting ensemble of D.T.s. However, as its complexity grows, When D.T.s are added in bulk, it will inevitably slow down.

6. Related works

Using facial region approaches with CNN and LBP and a classifier SVM, Ravi et al.[52] presented facial expression recognition. With CNN architecture, images are scaled to be analyzed quickly while retaining accurate forecasts. One of CNN's key benefits is that it may increase the recognition rate on the ck+ dataset by 97.32 percent. However, the low identification rate of 31.82% on the YALE data set was a significant restriction.

A strategy based on deep learning is presented for recognizing facial expressions, as described by Divya et al.[53]. CNN, SVM, VGG-16, ResNet-50, transfer learning, and ensemble learning are some methods employed. This study's facial expression identification dataset comes from Kaggle and Karelinska's Guided Emotional Faces. Once only possible with black and white photos, the suggested work now extends to color photos and even a live video feed.

The convolutional neural network approach is used for face emotion identification, as explained by the research[54,55] for picture categorization. The benefits of static and dynamic picture initialization are discussed, including using a co-evolutionary layer, a pooling layer, etc. The max-pooling technique nourished this function. Sixty-six percent of each feeling and accuracy target is shown in this project.

Anjum et al.[56] explains the adaptive features mapping in deep learning-based facial expression recognition model. Technologies such as cross-domain adaptation, pattern recognition, and image processing are utilized. In most cases, LBP will be the primary criterion. The CNN technique, as described by Xue et al.[57] clustering-based pre-processing approach, delivers a higher accuracy rate in face image processing applications. PCA and LDA are utilized for facial recognition. Eigenvectors, clusters, and a fuzzy c-means (FCM) clustering method as features for clustering-based discriminant analysis with multiple clusters. Facial expression recognition has additional benefits due to the high dimensionality of face images and the fact that the CDA method can effectively reduce their dimensions. Additionally, extraction schemes are helpful for facial expression recognition, but it creates accuracy in results.

As proposed by Maheswari et al.[58], multiview facial expression recognition uses several methodologies, including classifier fusion, principal component analysis, locality preserving projection, LDA methods, classifications, classifications, BU-3DFE data, and pre-processing. Some characteristics are location-preserving projection; closest neighbour; posture emotion cascade; LBD; and SIFT features. The main benefits include easy view recognition from any angle Lakshmi et.al.[59] Unfortunately, this survey feels the necessity for reducing angle intervals.

Combining empirical mode decomposition (EMD) with genetic algorithms is employed by Ismail et al.[60]. To classify an input signal, it is necessary to disentangle the intrinsic mode functions (IMFs) that correspond to its varied energy level. Using the chosen IMFs as building blocks, the filtered signal is next analyzed using higher-order crossings (HOC) to extract relevant features. After forcing signals through the filter mentioned above, the HOC analysis was employed with the EMD methods and a basic genetic algorithm to construct an energy-based filter. To get the feature vector, this method was used.

Mahmood et al.[61] evaluated the controlled classifiers' ability to consistently recognize facial expressions using a metric based on minimal chi-square features. These are the most significant and defining traits that can be used for accurate prediction. Six classifiers—including a multi-layer perceptron (MLP), support vector machine (SVM), decision-making board (DMB), random forest (R.F.), radial bias (R.B.), and K-nearest neighbor (KNN)—use these six characteristics to get a close possible to the true answer. This is accomplished by scrutiny of the classifiers' output. As a data set, CK+ is analyzed. The total accuracy ratio of 94.23%, achieved in the random rainforest, is the most accurate categorization available.

As cited in Abdulrazaq et al.[62], to evaluate the precision of six classifiers based on the Reliever-F method, we will be focusing on employing a small number of characteristics. The study is being evaluated using a number of different types of machine learning and artificial neural network architectures. Using CK+ data, the experiment proves that the most accurate classifier is the K-Nearest neighbor, with an average accuracy of 94.93 percent.

Using the MLP, Naive Bayes, decision tree, and KNN algorithms, Dino and Abdulrazzaq[63] proposed a comparative FER approach to identify the most distinguishing features of face pictures. Expression recognition is the goal of these classifiers, and their relative performance is evaluated and compared. The primary objective is to compare and contrast the performance of different classifiers in order to pick the one that works best with the small number of features. The primary goal of this method is to determine which classifier is the most effective by judging its performance on a set of criteria that is both manageable and representative. A CK+ version of the proposed solution was included in the presentation. Experiments show that KNN can classify data with a 91 percent accuracy with only 30 characteristics.

Together with a CNN idea of rank pooling, known as lively pictures, utilized to recognize micro motion by Le et al.[64], the compact approach is used to collect insightful features on particular regions of interest. In order to extract meaningful information from the dynamic image, only a limited number of face regions are used, each of which is based on the prominent muscle changes detected. The CNN models can handle an emotional grouping test for the final film depiction.

Liu et al. presented the Siamese Action Units Network (SAANet) to build a metric FER learning system mapped with space and time that incorporates the Action Unit Focus Module and the Vigilant Pooling Module, two types of care modules that take care of individual units and the system as a whole[65]. Using the baseline models of a (CNN) and a Recurrent Neural Network (RNN), we can create a metric learning system for complex and nuanced facial expressions. They provide novel A.U. methods for the space industry to more precisely and effectively aggregate spatial qualitative knowledge on the crucial regions from long-range dependencies. A specific pair-wise sampling strategy for this metric learning system is provided to ensure challenges in FER. For this purpose, a careful grouping module is employed to identify temporal commonalities between video frames in the time domain. Experiments have been conducted on the four benchmark datasets (extended Cohn-Kanade (CK+), Oulu-CASIA, Maja Pantic, and Michel Valstar. The results of the experiments support the idea that the new design is superior to the current best practices.

To foretell a client's sentimental at a service location, Chen et al.[66] constructed a Visual Geometry Group (VGG) model coupled with a video over time. The principal results are: 1) combining lengthy emotional changes with brief facial output utilizing the Bidirectional Long Short-Term Memory (BiLSTM) architecture for providing instruction regarding input sequences. The dataset is based on much research and was created to address the shortage of facial expression data sets in device and service contexts. Long-term fluctuations in expression and keyframe acquisition strategies, which were previously disregarded by researchers, have been proven to be successful in extensive tests. This provides a basis for focusing on the generation of picture sequences and aids in the investigation of the temporal structure of the transition phase.

A C3D-based network architecture, 3D-Incept-ResNet, was built by Chen et al.[67] to extract location and time from sequences of facial expressions. To make use of the preserved spatial and channel-integral links between characteristics extracted, a Care Module (STCAM) is suggested. The suggested STCAM is optimized for measuring a channel and a spatial-temporal map to improve functionality by adding characteristics along those dimensions. It is familiar with three widely used dynamic facial expressions to test this process. The results of our experiments imply that we perform well or balanced state-of-the-art approaches.

The dynamic kernel-based facial expressions proposed by Perveen et al.[68] are absorbed by local space-time represented in the generic Gaussian mix-model (uGMM). By employing uGMM statistics, these intricate kernels are employed to enhance the global significance of the same name while preserving local correlations. Explicit mapping using three widely used facial expression datasets (Motion Capture Initiative, Affective Expressions of the World, and Binghamton-Pittsburgh 4D Spontaneous Expressions of Emotion) demonstrates the value of dynamic kernel representation using three distinct dynamic kernels (BP4D). The consequences of these evaluations indicate that probability-based kernels are the most biased of all dynamic kernels. The intermediate matching kernels are more capable for computationally efficient.

When it comes to facial expressions in 3D, suggested a method powered by discrete-force field dynamics (DFFD) algorithm created by Ni and Liu[69] to capture realistic facial expressions pushes a universal neutral 3D model of the virtual face, a believable 3D animated expression may be synthesized to match the facial gestures of the actual performers. Limitations in visual, aural, and material transmitted in the associated process are illustrated by the 3D simulated face generated using this method[70]. The technology also depends on the performances of real people.

In the study of Verma et al.[71], acquired expression-sensitive characteristics that can not only offer a comparable result to state-of-the-art documentation but can also be comprehended by humans. Second, the dynamics of facial expression are reflected in patch sequences of local depth without the spatial time features. The authors suggest a two-stage selection approach to resolve the facial features than effectively segregate gestures apart. By feeding expressive information from the respective region into a hierarchical classifier for FER, the effectiveness of the generated face components may be verified. The suggested method is evaluated using the Binghamton University 4D, with findings indicating that the researched sensitive, expressive characteristics would reach comparable reconnaissance efficiency with existing approaches. In addition, the discovered functions in HOG3D can lead to a semantic comprehension of dynamic recognition.

In the study of Dong et al.[72] suggested a two-layer paradigm for identifying human emotional states as a means of investigating the temporal dynamics of human emotions in spatially disentangled settings. To begin, the input video is marked for each frame based on its time stage and the mood it conveys. To determine the ESTP of each video input frame, they employ the disconnected space backdrop to train many (SVM) classifiers. The first layer generates a sequence of frame labels; the second layer utilizes (DTW) to categorize those labels into one of seven emotional states: annoyance, indignation, disgust, anxiety, satisfaction, sadness, and surprise. When dealing with time-related complexity, such as the intensity with which one transmits emotion, the DTW's capacity to resolve vast disparities in periods is crucial. This recognition of the human emotion scheme is made simpler by the use of spatial and temporal approaches independently.

According to Maheswari et al.[73], to capture the nuances of facial movement in videos in a single snapshot, a complicated micro-expression representation was proposed. This data also reveals that LEARNet is able to accurately imprison the subtleties of facial expressions. LEARNet improves qualities by incorporating A.L. into the network. The AL response includes the hybrid feature maps produced by the interconnected convolution layers. The link between convolution layers is another key component of the LEARNet architecture that aids in preserving subtle but significant details about the facial muscles. High and micro-face functionality is preserved in the visual responses of the suggested LEARNet, proving the method's efficacy. To evaluate LEARNet's efficacy, we use the CASME-I, CASME-II, CAS(ME)2, and SMIC benchmarks. The study's experimental findings demonstrate a 4.03 percentage point, 1.90 percentage point, 1.79 percentage point, and 2.82 percentage point development in the CASME-I, CASME-II, CAS(ME)2, and SMIC datasets, respectively.

Lai et al[74] described the CNN topology in which consists of shallow, densely linked short FER roadways. Instead of using thick lines, the shallow CNN structure makes use of dense networking to promote sharing of features. Instead, our method, which injects modest distances for each convolution layer from all preceding layers, has been shown to reach competitive efficiency on the CK+ and Oulu-CASIA benchmark datasets based on extensive experiments. The purpose of this effort is to introduce dense networking for feature sharing, hence reducing the burden on current small-scale datasets. Some research suggests that when connections between levels are relatively brief, information can travel through the network and be learned by the system more quickly. However, we don't use dense intermediate blocks or transition layers as other dense designs do.

7. Challenges and chances

Over the past few decades, numerous people have tried their hands at theoretical analysis and practical implementations of FER algorithms. Several difficulties and prospects are presented and examined here when the FER literature switches its primary attention to the difficult wild environmental circumstances.

7.1. Wild environmental conditions

Especially in the wild, FER faces significant challenges in adapting to complex situations like occlusion and pose variation, which might interfere with the detection of unique facial expressions. Li et al. draw

attention to this problem and present two variants of ACNN to swap out the occluded patches for similar but non-obstruct ones[75]. Multiview face pictures with variable emotions and poses have also been generated using the GAN for use in FER[76].

Some current algorithms can handle FER under certain situations. However, these approaches have low accuracy and can only adjust to very specific fluctuations. To effectively use FER systems in the dynamic and complicated real world, enhancing flexibility is crucial.

7.2. The lack of high-quality publicly available information

The process of FER relies heavily on the collection and analysis of relevant data. When training a network, it is typically necessary to capture minor deformations associated with expressions. The lack of sufficient and high-quality training data is the primary obstacle for deep FER systems. In Section 4, we introduce and examine a few frequently used FER-related datasets. However, one of the most consistent issues is the limited scope of these data. Concurrently, employing several datasets might lead to issues with data bias and inconsistent annotations due to differences in the class nature of the annotation process. Disparities in socioeconomic status also frequently arise.

7.3. The heaviness of high-volume data dispensation

Current FER systems achieve admirably on conventional datasets that make extensive use of tiny capacity and poor pixel resolution. The FER system has difficulties in data storage, transmission, and processing because of the terabyte-scale data used in many research and commercial applications. When running FER on time-series data, data compression is also an absolute necessity. The FACS hypothesis postulates that the face area as a whole is far larger than just the expression-related units. The elimination of superfluous face data might be a useful data-compression strategy.

7.4. Multimodal affect acknowledgement

Although FER is based on visible face pictures, which achieve promising performance on its own, it can be further enhanced by merging it with other models into an integrated scheme. One such method proposed by Ramakrishnan and El Emary[77] uses audio to identify the emotional tone of a speaker's voice. Furthermore, infrared pictures with temporal skin records and 3D face images with depth information are insensitive to light fluctuations, suggesting they may be a useful option for studying natural facial expressions. The valence-arousal (V.A.) model may also be used to establish a connection between facial expression and emotion states, allowing enabling multimodal dimensional emotion identification[78].

7.5. Visual privacy

Camera-smartphones have a significant challenge in the form of rising privacy concerns. Many other FER approaches have been presented, most of which rely on high-resolution photos but pay little to no regard to preserving users' visual privacy. Therefore, FER systems require more reliable and precise privacy protection measures to provide a happy medium between privacy and data value.

8. Future directions

8.1. Super-resolution

The concept of super-resolving images has recently grabbed the interest of academics. Its goal is to improve the visual clarity and detail of a low-resolution image by making it into a high-resolution one. Lower spatial resolution/smaller size or deterioration like blurring will be contributed to an image's "reduced resolution." Here's the equation you'll need to use to link the High Resolution (H.R.) and Low Resolution (L.R.) images:

Low-resolution models may be created from high-resolution source photos using the following formula. The letters D, I_y, and I_x, respectively, stand for "degradation function," "high resolution," "low resolution," and "noise," respectively.

$$I_x = D(I_y; a) \quad (2)$$

The degradation function (D) is unknown, so low resolution is achieved from high-resolution equivalents of a picture.

When dealing with the challenge of a small or blurry image, Super Resolution (S.R.) techniques typically perform better than conventional algorithms like bilinear and bicubic. Reversing the process from a high-resolution to a low-resolution picture is more challenging. Lost low-resolution image elements must be recovered. A deep CNN model operates on both low-resolution and high-resolution pictures, as described in a recent paper[79]. Simply said, it is an easier and more accurate alternative to bicubic interpolation. SRCNN plays a similar vital role as Very Deep Super Resolution (VDSR) was also discussed. If you compare it to SRCNN, though, you will find that it goes into greater detail. Although SRCNN is one method for accomplishing super-resolution, other methods like ESPCN and EDSR are also viable options.

8.2. Transfer learning

Using a model's weights from one dataset to make predictions on a third dataset is known as transfer learning[80]. Since it works even with a small dataset, it has gained a lot of popularity in the field of FER. Transferring learned features onto another dataset is the most efficient option since training the model on all of the millions of datasets would result in numerous inefficiencies.

Learned Transfer for Emotion Recognition: In order to use transfer learning for emotion identification, it is necessary to first extract enormous datasets. Classes identified in the original datasets used to train the model may not correspond to the target classes in the target dataset. To further improve generalization and robustness, an occluded dataset is also employed, which is frequent and realistic in everyday life.

8.3. Domain adaptation

Machine learning's domain adaptation deals with training on one circulation and implementing it on different circulation(similar)[78]. Domain adaptation is a method for addressing novel problems in a target domain by using an existing labeling infrastructure in one or more source domains. Most often, the source and destination domains are similar to achieve an accurate result. Domain adaptation can be challenging when the task space remains unchanged, and the only difference is a divergent input domain.

Emotion Detection Through Domain Adaptation: An innovative way of domain adaption methodology in a new study[80] to identify facial expressions by combining human and non-human features. Intersection scores are used for prediction in the proposed approach. It also recommended utilizing Attentional CNN with pre-trained models for recognizing facial expressions. Experiments conducted on the Flickr picture dataset labeled with fundamental emotions (such as angry, happy, sad, and neutral) demonstrated a 63.87% for emotion detection, which findings.

8.4. Adversarial machine learning

For the purposes of (AML) are nefarious inputs that are intended to cause the model to incorrectly anticipate labels. In order to recreate the missing sceneries in the real environment or discover a method to avoid such challenges, adversaries interrupt the way the model normally predicts. Adversarial machine learning has emerged in recent years as an integral aspect of any job, including FER, activity identification, and object detection.

Competing Machine Learning Approaches to Identifying Emotions: by smearing subsequent fully connected layers, the adversarial approach[81] claims to deliver anonymity to applications. Two classifiers are

used to identify emotions and identify people based on the data it produces. CNN can retain information about emotions and calculated identities.

8.5. Zero-shot learning

At test time[82], zero-shot ML is utilized to distinguish target classes that have never been seen before, even when the training times[81,82] did not include the test label.

In zero-shot learning, the information is as follows:

Classes attended: tag pictures from relevant lectures during your training.

Second, unseen courses: no photographs were ever labeled as belonging to these courses throughout the training phase.

The third piece of context: data for both shown and hidden classes, including descriptions, semantic features, and word embeddings during training time. The data connects the dots between categories that are normally out of sight.

Learning from Scratch in the Field of Emotion Acknowledgement: Generalized zero-shot learning (GZSL) has been proposed for emotion recognition in a recent work[83]. It has 3 different parts: A Prototype-Based Detector (PBD) for predicting unseen gesture groups based on learned data, a stacked autoencoder for classification, and a fusion layer for combining the two. The third fork facilitates emotional identification in a broader context.

8.6. Federated machine learning (F.L.)

It's a novel approach to machine learning in which the procedure is deployed over a network of scattered edge devices or servers that each keep their own copies of the sample data and never share them[84]. This approach is different from the standard centralized machine learning techniques, which need all datasets to be uploaded to a central server. Federated learning requires many players to collaborate to deliver conventional tough learning models without sharing data, and it addresses basic concerns, including diverse data. When using F.L., the model may freely access and learn from data sets in a wide variety of locales without safety precautions. This learning model allows a wide range of industries, including healthcare, pharmaceuticals, military, space, and makers of heavy machinery, to design a speedier, dispersed, and more trustworthy model.

Recent research[83] has shown feature extraction methods for pictures and audio paving the way for federated learning in emotion detection. The suggested method is able to identify human emotions by analyzing gathered facial and audio data. Both classifiers of a person's categorized emotions provide the result. The proposed classifiers for detecting emotions in both faces and voices have an accuracy of 71.64 and 85.04 percent. Determine whether professional help from a psychologist is necessary based on the outcome.

8.7. Explainable A.I.

An advanced A.I. idea, "explainable A.I.," provides an explanation of the logic behind a decision in a way that a human can understand[84]. Pre-emptive design or retroactive investigation. There is a current trend toward employing such methods in an effort to shed light on the murky realm of artificial intelligence and provide models with greater credibility and plausibility. Only humans can properly assess, label, and describe the whole range of human emotions. However, A.I. only displays the results of what it has learned without providing an explanation for those results; for instance, A.I. can determine if a person has pneumonia or not just by glancing at an X-ray. Still, it won't be believed since it skips a vital step—proposing a doctor's opinion before announcing the results so failed to earn people's trust. In these situations, Explainable A.I. will provide a conclusion and an explanation that is both more accurate and convincing than any prior A.I. model.

9. Facial emotion recognition potential applications

The action of Muscles below the skin causes expressions on the face. As such, they play a vital role in the dissemination of interpersonal news. Insight into the subjective and immediate nature of the emotional content sent by a person's face can be gained via the study of facial expression analysis. The fields of medicine, e-learning, monitoring, entertainment, legislation, etc., are only a few of the benefits of FER. Below, we'll examine how FER has been applied in each of these areas.

9.1. E-Learning

By monitoring their students' reactions, online educators may gauge their level of understanding and tailor their lessons accordingly. Whether via traditional classroom instruction or online courses, students everywhere will benefit from this work toward a stronger educational infrastructure.

9.2. Monitoring

Psychological research has shown that emotions play a crucial part in responsible driving. The driver's ease and security behind the wheel are affected by the driver's internal emotional state. Research on human emotions found that negative ones like anger, hopelessness, and fear all contributed to dangerous behaviours like speeding and irresponsible driving. The risk of having an accident increases with factors including anger, aggression, fatigue, and stress. Both anxiety and depression can have an effect on driving, especially when they occur together. As a result, the FER scheme that constantly analyses driving and detects them will warn the driver and avoid accidents if the expressions fall into one of the categories stated. Critical to police operations, FER analyses a person's facial expressions while withdrawing an amount from an ATM. It is easily understandable whether a person experiences fear or not.

The organization then comes up with a strategy to stop handing out money. Installing a FER tool in businesses allows for the monitoring of customer preferences and satisfaction, providing valuable information that can be analyzed to enhance the shopping experience for the customer[85].

9.3. Medicine

The patient's inability to travel, whether due to illness or old age, the patient's location, or both, where everything makes it difficult to schedule regular checkups. With the help of FER systems, we can prevent a lapse in the administration of necessary medications. Children with autism have a harder time interacting with others because they have a diminished capacity to understand faces. Help autistic children learn to read facial expressions by creating a mobile FER app and distributing it to them. Children who have trouble reading facial expressions might benefit from being assigned an emoji to represent that emotion[86].

9.4. Entertainment

Video game developers can better engage their players emotionally by relying on players' assertions about their real-time user experience[87]. The authors argue that watching and analyzing a player's facial expressions in real time is necessary for determining whether or not a game succeeds in creating a positive user experience.

10. Conclusion

In recent years there has been a rise in interest in FER. Many cutting-edge FER algorithms have emerged in the past decade. This study offers a thorough analysis of the latest developments in FER technology. The study begins with a brief impression of the history of FER research and the definition of key terms. We have separated the current FER approaches into two groups: one is a more traditional one, and the other one relies on deep learning. Specifically, the survey separates the traditional approaches into three distinct phases: picture pre-processing, feature extraction, and expression categorization. Methods of varying degrees of complexity are introduced and addressed at each stage. In addition, 17 new FER datasets are presented. The poll concludes

with a presentation of some of the potential and difficulties facing the FER that further warrant study. The purpose of this survey is to encourage more research into the topic of FER by providing a structured and complete existing in the area. Future research difficulties and open questions are discussed, and it is determined that more work remains to be done in this area, particularly with respect to FER in 3D face-shape replicas and the recognition of emotion in images when they are obscured, among other things.

Author contributions

Conceptualization, SRK and ALK; methodology, SRK and ALK; formal analysis & data curation, SRK and ALK; writing—original draft preparation, SRK; writing—review & editing, ALK; supervision, ALK.

Conflict of interest

The authors declare no conflict of interest.

References

1. Kumari A, Tanwar S, Tyagi S, Kumar N. Fog computing for healthcare 4.0 environment: Opportunities and challenges. *Computers and Electrical Engineering* 2018; 72: 1–13. doi: 10.1016/j.compeleceng.2018.08.015
2. Hathaliya J, Sharma P, Tanwar S, Gupta R. Blockchain-based remote patient monitoring in healthcare 4.0. In: Proceedings of 2019 IEEE 9th International Conference on Advanced Computing (IACC); 13–14 December 2019; Tiruchirappalli, India.
3. Vora J, DevMurari P, Tanwar S, et al. Blind signatures based secured e-healthcare system. In: Proceedings of 2018 International Conference on Computer, Information and Telecommunication Systems (CITS); 11–13 July 2018; Alsace, France.
4. Zhang L, Verma B, Tjondronegoro D, Chandran V. Facial expression analysis under partial occlusion: A survey. *ACM Computing Surveys* 2018; 51(2): 1–49. doi: 10.1145/3158369
5. Maheswari UV, Aluvalu R, Chennam KK. Application of machine learning algorithms for facial expression analysis. *Machine Learning for Sustainable Development* 2021; 9: 77–96. doi: 10.1515/9783110702514-005
6. Swapna M, Viswanadhula UM, Aluvalu R, et al. Bio-Signals in medical applications and challenges using artificial intelligence. *Journal of Sensor and Actuator Networks* 2022; 11(1): 17. doi: 10.3390/jsan11010017
7. Spezialetti M, Placidi G, Rossi S. Emotion recognition for humanrobot interaction: Recent advances and future perspectives. *Journal of Sensor and Actuator Networks* 2020; 7: 145. doi: 10.3389/FROBT.2020.532279
8. Ramis S, Buades JM, Perales FJ. Using a social robot to evaluate facial expressions in the wild. *Sensors* 2020; 20(23): 1–24. doi: 10.3390/s20236716
9. Bhatti YK, Jamil A, Nida N, et al. Facial expression recognition of instructor using deep features and extreme learning machine. *Computational Intelligence and Neuroscience* 2021; 2021: 5570870. doi: 10.1155/2021/5570870
10. Li S, Deng W. Deep facial expression recognition: A survey. *arXiv* 2018; arXiv:1804.08348. doi: 10.1109/TAFFC.2020.2981446
11. Pulmamidi N, Aluvalu R, Maheswari VU. Intelligent travel route suggestion system based on pattern of travel and difficulties. *IOP Conference Series: Materials Science and Engineering* 2021; 1042: 012010. doi: 10.1088/1757-899X/1042/1/012010
12. Hemalatha G, Sumathi C. A study of techniques for facial detection and expression classification. *International Journal of Computer Science and Engineering Survey* 2014; 5(2): 27. doi: 10.5121/ijcses.2014.5203
13. Deodhare D. *Facial Expressions to Emotions: A Study of Computational Paradigms for Facial Emotion Recognition*. Springe; 2015. pp. 173–198.
14. Chengeta K, Viriri S. Facial expression recognition: A survey on local binary and local directional patterns. *Lecture Notes in Computer Science* 2018; 11055: 513–522. doi: 10.1007/978-3-319-98443-8_47
15. Baskar A, Kumar TG. Facial expression classification using machine learning approach: A review. *Data Engineering and Intelligent Computing: Proceedings of IC3T* 2016; 2018: 337–345.
16. Sariyanidi E, Gunes H, Cavallaro A. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2014; 37(6): 1113–1133. doi: 10.1109/TPAMI.2014.2366127
17. Tian Y-I, Kanade T, Cohn JF. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2001; 23: 97–115. doi: 10.1109/34.908962
18. Yan H. Transfer subspace learning for cross-dataset facial expression recognition. *Neurocomputing* 2016; 208: 165–173. doi: 10.1016/j.neucom.2015.11.113

19. Benini S, Khan K, Leonardi R, et al. Face analysis through semantic face segmentation. *Signal Processing: Image Communication* 2019; 74: 21–31. doi: 10.1016/j.image.2019.01.005
20. Verma VK, Srivastava S, Jain T, Jain A. Local invariant feature-based gender recognition from facial images. In: *Soft Computing for Problem Solving*. Springer; 2019. pp. 869–878.
21. Lyons M, Akamatsu S, Kamachi M, Gyoba J. Coding facial expressions with Gabor wavelets. In: Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition; 14–16 April 1998; Nara, Japan.
22. Lucey P, Cohn JF, Kanade T, et al. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops; 13–18 June 2010; San Francisco, USA.
23. Du S, Tao Y, Martinez AM. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences* 2014; 111(15): E1454–E1462. doi: 10.1073/pnas.1322355111
24. Mavadati SM, Mahoor MH, Bartlett K, et al. Disfa: A spontaneous facial action intensity database. *IEEE Transactions Affective Computing* 2013; 4(2): 151–160. doi: 10.1109/T-AFFC.2013.4
25. Pantic M, Valstar M, Rademaker R, Maat L. Web-based database for facial expression analysis. In: Proceedings of 2005 IEEE International Conference on Multimedia and Expo; 6–6 July 2005; Amsterdam, Netherlands.
26. Yin L, Wei X, Sun Y, et al. A 3D facial expression database for facial behavior research. In: Proceedings of 7th International Conference on Automatic Face and Gesture Recognition (FGR06); 10–12 April 2006, Southampton, UK.
27. Zhang X, Yin L, Cohn JF, et al. BP4D-spontaneous: A high-resolution spontaneous 3D dynamic facial expression database. *Image Vision Computing* 2014; 32: 692–706. doi: 10.1016/j.imavis.2014.06.002
28. Kaulard K, Cunningham DW, Bülthoff HH, Wallraven C. The MPI facial expression database—A validated database of emotional and conversational facial expressions. *PLoS One* 2012; 7(3): e32321. doi: 10.1371/journal.pone.0032321
29. Lundqvist D, Flykt A, Öhman A. The Karolinska directed emotional faces (KDEF). *CD ROM Dep* 1998; 91: 630. doi: 10.1037/t27732-000
30. Wang S, Liu Z, Lv S, et al. A natural visible and infrared facial expression database for expression recognition and emotion inference. *IEEE Transactions on Multimedia* 2010; 12(7): 682–691. doi: 10.1109/TMM.2010.2060716
31. Gross R, Matthews I, Cohn J, et al. Multi-PIE. *Image Vision Computing* 2010; 28(5): 807–813. doi: 10.1016/j.imavis.2009.08.002
32. Zhao G, Huang X, Taini M, et al. Facial expression recognition from near-infrared videos. *Image Vision Computing* 2011; 29(9): 607–619. doi: 10.1016/j.imavis.2011.07.002
33. Carrier PL, Courville A, Goodfellow IJ, et al. *FER-2013 Face Database*. Universit de Montral; 2013.
34. Valstar MF, Mehu M, Jiang B, et al. Meta-analysis of the first facial expression recognition challenge. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 2012; 42(4): 966–979. doi: 10.1109/TSMCB.2012.2200675
35. Dhall A, Goecke R, Lucey S, Gedeon T. Collecting large, richly annotated facial-expression databases from movies. *IEEE Multimedia* 2012; 19(3): 34–41. doi: 10.1109/MMUL.2012.26
36. Dhall A, Goecke R, Lucey S, Gedeon T. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In: Proceedings of 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops); 6–13 November 2011; Barcelona, Spain.
37. Li S, Deng W. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. *IEEE Transions Image Processing* 2019; 28: 356–370. doi: 10.1109/TIP.2018.2868382
38. Li S, Deng W. Blended emotion in-the-wild: Multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning. *International Journal of Computer Vision* 2018; 127: 884–906. doi: 0.1007/s11263-018-1131-1
39. Whitehill J, Littlewort G, Fasel I, et al. Toward practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2009; 31(11): 2106–2111. doi: 10.1109/TPAMI.2009.42
40. Lucey P, Cohn JF, Prkachin KM, et al. Painful data: The UNBC-McMaster shoulder pain expression archive database. In: Proceedings of 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG); 21–25 March 2011; Santa Barbara, USA.
41. Viola P, Jones MJ. Robust real-time face detection. *International Journal of Computer Vision* 2004; 57: 137–154. doi: 10.1023/B:VISI.0000013087.49260.fb
42. Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees. In: Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition; 23–28 June 2014; Columbus, USA.
43. Zhang K, Zhang Z, Li Z, Qiao Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters* 2016; 23(10): 1499–1503. doi: 10.1109/LSP.2016.2603342
44. Tomasi C, Manduchi R. Bilateral filtering for gray and color images. In: Proceedings of Sixth International Conference on Computer Vision; 7–7 January 1998; Bombay, India.
45. Lindenbaum M, Fischer M, Bruckstein A. On Gabor's contribution to image enhancement. *Pattern Recognition* 1994; 27(1): 1–8. doi: 10.1016/0031-3203(94)90013-2

46. Garg P, Jain T. A comparative study on histogram equalization and cumulative histogram equalization. *International Journal of New Technology and Research* 2017; 3: 41–43.
47. Hawkins DM. The problem of overfitting. *Journal of Chemical Information and Computer Sciences* 2004; 44: 1–12. doi: 10.1021/ci0342472
48. Jolliffe I. *Principal Component Analysis*. Springer; 2011.
49. Ojala T, Pietikäinen M, Mäenpää T. Multiresolution Gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2002; 24(7): 971–987. doi: 10.1109/TPAMI.2002.1017623
50. Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2001; 23(6): 681–685. doi: 10.1109/34.927467
51. Pakstas A, Forchheimer R, Pandzic IS. *MPEG-4 Facial Animation: The Standard, Implementation and Applications*. John Wiley & Sons; 2002.
52. Ravi R, Yadhukrishna SV, Prithviraj R. A face expression recognition using CNN and LBP. In: Proceedings of 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC); 11–13 March 2020; Erode, India.
53. Divya M, Reddy ROK, Raghavendra C. Effective facial emotion recognition using convolutional neural network algorithm. *International Journal of Recent Technology and Engineering (IJRTE)* 2019; 8(4): 2277–3878.
54. Shah ANB, Patel N, Dave JA, et al. *Role of Artificial Intelligence and Neural Network in the Health-Care Sector: An Important Guide for Health Prominence*. CRC Press; 2023. pp. 239–263.
55. Pisupati S, Ismail BM. Image registeration method for satellite image sensing using feature based techniques. I*nternational Journal of Advanced Trends in Computer Science and Engineering* 2020; 9(1):490–593. doi: 10.30534/ijatcse/2020/82912020
56. Anjum G, Reddy TB, Ismail BM, et al. Variable block size hybrid fractal technique for image compression. In: Proceedings of 2020 6th International Conference on Advanced Computing and Communication Systems; 6–7 March 2020; Coimbatore, India.
57. Xue M, Mian A, Duan X, Liu W. Learning interpretable expression-sensitive features for 3D dynamic facial expression recognition. In: Proceedings of 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019); 14–18 May 2019; Lille, France.
58. Maheswari VU, Prasad GV, Raju SV. Facial expression analysis using local directional stigma mean patterns and convolutional neural networks. *International Journal of Knowledge-Based and Intelligent Engineering Systems* 2021; 25(1): 119–128. doi: 10.3233/KES-210057
59. Lakshmi KN, Reddy YK, Kireeti M, et al. Design and implementation of student chat bot using AIML and LSA. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* 2019; 8(6): 1742–1746.
60. Ismail M, Vardhan VH, Mounika VA, Padmini KS. An effective heart disease prediction method using artificial neural network. *International Journal of Innovative Technology and Exploring Engineering* 2019; 8(8): 1529–1532.
61. Mahmood MR, Abdulrazzaq MB, Zeebaree S, et al. Classification techniques performance evaluation for facial expression recognition. *Indonesian Journal of Electrical Engineering and Computer Science* 2021; 21:1176–1184.
62. Abdulrazaq MB, Mahmood MR, Zeebaree SR, et al. An analytical appraisal for supervised classifiers' performance on facial expression recognition based on relief-f feature selection. *Journal of Physics: Conference Series* 2021; 1804: 012055. doi: 10.1088/1742-6596/1804/1/012055
63. Dino HI, Abdulrazzaq MB. A comparison of four classification algorithms for facial expression recognition. *Polytechnic Journal* 2020; 10: 74–80. doi: 10.25156/ptj.v10n1y2020.pp74-80
64. Le TTQ, Tran TK, Rege M. Dynamic image for micro-expression recognition on region-based framework. In: Proceedings of 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI); 11–13 August 2020, Las Vegas, USA.
65. Liu D, Ouyang X, Xu S, et al. SAANet: Siamese action-units attention network for improving dynamic facial expression recognition. *Neurocomputing* 2020; 413: 145–157. doi: 10.1016/j.neucom.2020.06.062
66. Chen L, Ouyang Y, Zeng Y, Li Y. Dynamic facial expression recognition model based on BiLSTM-Attention. In: Proceedings of 2020 15th International Conference on Computer Science & Education (ICCSE); 18–22 August 2020; Delft, Netherlands.
67. Chen W, Zhang D, Li M, Lee DJ. STCAM: Spatial-temporal and channel attention module for dynamic facial expression recognition. *IEEE Transactions on Affective Computing* 2020; 14(1): 800–810. doi: 10.1109/TAFFC.2020.3027340
68. Perveen N, Roy D, Chalavadi KM. Facial expression recognition in videos using dynamic kernels. *IEEE Transactions on Image Processing* 2020; 29: 8316–8325. doi: 10.1109/TIP.2020.3011846
69. Ni H, Liu J. 3D face dynamic expression synthesis system based on DFFD. In: Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC); 15–17 March 2019; Chengdu, China.

70. Alazrai R, Yousef KWA, Daoud MI. Emotion recognition based on decoupling the spatial context from the temporal dynamics of facial expressions. In: Proceedings of 2019 International Symposium on Networks, Computers and Communications (ISNCC); 18–20 June 2019; Istanbul, Turkey.
71. Verma M, Vipparthi SK, Singh G, Murala S. LEARNet: Dynamic imaging network for micro expression recognition. *IEEE Transactions on Image Processing* 2019; 29: 1618–1627. doi: 10.1109/TIP.2019.2912358
72. Dong J, Zheng H, Lian L. Dynamic facial expression recognition based on convolutional neural networks with dense connections. In: Proceedings of 2018 24th International Conference on Pattern Recognition (ICPR); 20–24 August 2018; Beijing, China.
73. Maheswari VU, Varaprasad G, Viswanadharaju S. Local double directional stride maximum patterns for facial expression retrieval. *International Journal of Biometrics* 2022; 14(3–4): 439–452. doi: 10.1504/ijbm.2022.124682
74. Lai YH, Lai SH. Emotion-preserving representation learning via generative adversarial network for multiview facial expression recognition. In: Proceedings of 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018); 15–19 May 2018; Xi'an, China.
75. Ramakrishnan S, El Emary IMM. Speech emotion recognition approaches in human computer interaction. *Telecommuniaction Systems* 2013; 52: 1467–1478. doi: 10.1007/s11235-011-9624-z
76. Chang J, Scherer S. Learning representations of emotional speech with deep convolutional generative adversarial networks. In: Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 5–9 March 2017; New Orleans, USA.
77. Vo TH, Lee GS, Yang HJ, Kim SH. Pyramid with super resolution for in-the-wild facial expression recognition. *IEEE Access* 2020; 8: 131988–132001. doi: 10.1109/ACCESS.2020.3010018
78. Yang Q. *An Introduction to Transfer Learning.* Springer; 2008.
79. Maheswari VU, Aluvalu R, Kantipudi MP, et al. Driver drowsiness prediction based on multiple aspects using image processing techniques. *IEEE Access* 2022; 10: 54980–54990. doi: 10.1109/ACCESS.2022.3176451
80. Narula V, Wang ZY, Chaspari T. An adversarial learning framework for preserving users' anonymity in face-based emotion recognition. *arXiv* 2020; arXiv:2001.06103. doi: 10.48550/arXiv.2001.06103
81. Soysal OA, Guzel MS. An introduction to zero-shot learning: An essential review. In: Proceedings of 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA); 26–28 June 2020; Ankara, Turkey.
82. Wu J, Zhang Y, Zhao X, Gao W. A generalized zero-shot framework for emotion recognition from body gestures. *arXiv* 2020; arXiv:2010.06362. doi: 10.48550/arXiv.2010.06362
83. McMahan HB, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* 2017; 54: 1273–1282. doi: 10.48550/arXiv.1602.05629
84. Longo L, Goebel R, Lecue F, et al. Explainable artificial intelligence: Concepts, applications, research challenges and visions. International Cross-Domain Conference for Machine Learning and Knowledge Extraction. In: *Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics* 2020; 12279: 1–16. doi: 10.1007/978-3-030-57321-8_1
85. Generosi A, Ceccacci S, Mengoni M. A deep learning-based system to track and analyze customer behavior in retail store. In: Proceedings of 2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin); 2–5 September 2018; Berlin, Germany.
86. Haque MIU, Valles D. A facial expression recognition approach using DCNN for autistic children to identify emotions. In: Proceedings of 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON); 1–3 November 2018; Vancouver, Canada.
87. Liu X, Lee K. Optimized facial emotion recognition technique for assessing user experience. In: Proceedings of 2018 IEEE Games, Entertainment, Media Conference (GEM); 15–17 August 2018; Galway, Ireland.