

## Research Article

# Methods for Facial Expression Recognition with Applications in Challenging Situations

Anil Audumbar Pise <sup>1,2</sup> Mejdal A. Alqahtani <sup>3</sup> Priti Verma,<sup>4</sup> Purushothama K,<sup>5</sup> Dimitrios A. Karras,<sup>6</sup> Prathibha S,<sup>7</sup> and Awal Halifa <sup>8</sup>

<sup>1</sup>Computer Science and Applied Mathematics University of the Witwatersrand Johannesburg, Johannesburg, South Africa

<sup>2</sup>Department of Sustainable Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Saveetha Nagar, Thandalam, Chennai 602105, Tamilnadu, India

<sup>3</sup>Department of Industrial Engineering, King Saud University, Riyadh, Saudi Arabia

<sup>4</sup>School of Business Studies, Sharda University, Greater Noida, India

<sup>5</sup>Department of Computer Science and Engineering, Shri Venkateswara College of Engineering, Vidya Nagara Airport Road, Bangalore, India

<sup>6</sup>National and Kapodistrian, University of Athens (NKUA), School of Science Department General, Athens, Greece

<sup>7</sup>Department of Electronics and Communication, Government Engineering College Ramanagara, Ramanagara, India

<sup>8</sup>Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Correspondence should be addressed to Awal Halifa; [ahalifa@tatu.edu.gh](mailto:ahalifa@tatu.edu.gh)

Received 21 February 2022; Revised 12 April 2022; Accepted 18 April 2022; Published 25 May 2022

Academic Editor: Vijay Kumar

Copyright © 2022 Anil Audumbar Pise et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the last few years, a great deal of interesting research has been achieved on automatic facial emotion recognition (FER). FER has been used in a number of ways to make human-machine interactions better, including human center computing and the new trends of emotional artificial intelligence (EAI). Researchers in the EAI field aim to make computers better at predicting and analyzing the facial expressions and behavior of human under different scenarios and cases. Deep learning has had the greatest influence on such a field since neural networks have evolved significantly in recent years, and accordingly, different architectures are being developed to solve more and more difficult problems. This article will address the latest advances in computational intelligence-related automated emotion recognition using recent deep learning models. We show that both deep learning-based FER and models that use architecture-related methods, such as databases, can collaborate well in delivering highly accurate results.

## 1. Introduction

A human face has significant and distinguishing characteristics that aid in the recognition of facial expressions. FER is defined as a change in facial expression caused by an individual's internal emotional state. It is used in a wide range of human-computer interaction (HCI) applications, such as face image processing, facial video surveillance, and facial animation, as well as in the fields of computer vision, digital image processing, and artificial intelligence. Automatic facial expression recognition is a difficult topic that has piqued the interest of many researchers in recent years. In FER, the stage of feature extraction is critical. Alek et al. [1]

demonstrated in the literature that facial expression accounts for 55% of total transmission while vocal and spoken communication contributes 38% and 7%, respectively.

There are two primary techniques for designing a FER system. As an initial step, some systems employ a sequence of images ranging from a neutral face to the peak level of emotions. In comparison, some systems use a single image of the face to recognize related emotions, and because they have access to less information, they often perform worse than leading approaches [2,3]. Apart from the approach type modeled by a FER system, another classification is based on the type of features employed in the recognition process, with a FER system utilizing one or both of these feature

categories. The first set of traits is obtained from the facial organs' posture and the skin's texture. The second type of feature is geometric features, which hold information about various positions and points on the face and are used to analyze a static image or a sequence of photos by utilizing the movement of the positions and points within the sequence. Using face landmarks as a starting point for extracting geometric features is one way. Landmarks are significant places on the face that provide useful information for facial analysis. Numerous studies have been undertaken on the subject of facial landmark identification; however, they are outside the scope of this work. This work employs the *Python* module *dlib* to detect these points [4].

Two different things, which are involved in automatically identifying human emotions and psychology, are part of artificial intelligence. The big question that researchers are attempting to answer in the area of psychology and artificial intelligence is the identification of emotions. One of them includes both topics such as mood and accent, which is generated in both vocal and nonverbal sensors such as the tonality and aural alterations [5] which are widely accessible, for instance [6], and a quick mood assessment can be obtained, as well as other sources [7]. Results from Mehrrian's study [8] demonstrated that 55% of information was sensory (emotional and verbal), with the remaining 7% being percent of having an unspecified physical component. The first indication a person gives of their emotional state being in a state is facial expressions, so many researchers are very interested in this modality.

First working in the extraction feature space to add new features to an existing representation can be a good thing because it will help the other features as well. Ekman and Friesen [9] noted that the Facial Action Coding System (FACS) and facial movement action units (AUs) assume that each coded movement in FACS involves at least one facial muscle. Ekman and Friesen first recognized that FACS facial movement, in FACS facial AUs, is utilized in such a way that face muscles and facial muscles are coded for each of their head movements (between several different individuals and/or races of subjects).

*1.1. Motivation of This Study.* Humans have a basic set of emotions, which are communicated via universal and basic facial expressions. Automatic emotion identification in images and videos will be possible if an algorithm that identifies, extracts, and assesses these facial expressions in real time is developed. In a social setting, facial expressions are powerful tools for communicating personal feelings and intentions. They play an important role in human social interaction. Face perception in the context of the situations in which they are seen offers important contextual information for facial expression processing. The ability to detect and communicate human emotions is critical in interpersonal relationships. Since the beginning of time, automatic emotion detection has been a controversial scientific subject. As a consequence, significant progress has been made in this area. Emotions are expressed via a number of means, including words, hand and body movements, and facial

expressions. As a result, the ability to extract and comprehend emotion is critical for effective human-machine communication.

### 1.2. Contribution of This Study

- (1) Artificial neural networks (ANNs), backpropagation, multilayer perceptron (MLP), support vector machine (SVM), convolutional neural network (CNN), random forest, RNN, deep belief neural networks (DBNN), genetic algorithm, long short-term memory (LSTM), ResNet, SqueezeNet, SqueezeNet-TRN, and ImageNet are among the methods and techniques covered in detail
- (2) The many articles written for the research and published in the bibliographies have also highlighted the feature selection methods utilized in the different categorization methods
- (3) In addition, data with different degrees of precision have been given in order to evaluate the potential of the best approach
- (4) Furthermore, the details of the data gathering, which includes a variety of facial images, have been revealed
- (5) There is also a summary of the many facial emotions that have been classified throughout time by various researchers
- (6) A collection of literature on facial expression detection using the image and video classification has been researched and collected in order to provide unified knowledge on a single platform for new researchers interested in this area
- (7) Furthermore, the limitations of each study have been recognized and addressed in order to offer a basis for future research

In this paper, we examine the current frontiers of emotion detection through different architectures with regard to the full range of expressive modes, given facial cues. The recent results from 2016 to 2021 are reported along with an analysis of the most prevalent problems and recent contributions to resolving them. It is set up in the following manner. Section 2 begins with the basic types to describe facial expressions such as FACS and prototypic emotional expressions. Section 3 represents face detection and emotion recognition system structure. Section 4 focuses on recent FER findings with real-time applications. Section 5 gives a short summary of FER challenges in the area and speculations on what lies ahead. Section 6 presents links to some public databases used in FER tasks followed by basic types of emotion recognition in Section 7. Deep learning-based facial emotion recognition is briefly explained in Section 8. Section 9 represents a comparative analysis and discussion on FER. Finally, we end with an outlook on what the future will hold with the conclusion.

## 2. Basic Types to Describe Facial Expressions

When it comes to describing facial expressions, there are two main methods to consider.

**2.1. Facial Action Coding System.** The FACS [11] identifies little changes in facial characteristics. This widely used method in psychology is based on a human observer's observation and consists of 44 action units connected to the tightening of groups of facial muscles in order to detect facial emotions (see Figure 1; muscles of facial expression are 1. frontalis; 2. orbicularis oculi; 3. zygomaticus major; 4. risorius; 5. platysma; 6. depressor anguli oris). In addition, a couple of the action units are shown in Figure 2. FACS is often coded and labeled manually by skilled individuals who examine slow-motion video footage of face muscle contractions before coding and labeling them. In recent years, many efforts to automate this procedure have been undertaken [13]. The system's dependence on descriptive data labels rather than inferential data labels, on the other hand, may provide a difficulty since it allows for the capture of nuanced facial expressions, among other things. To estimate emotions using FACS data, the FACS data must first be transformed into a system capable of doing so. This kind of behavior is modeled by the Emotional Facial Action System (EMFACS) [14].

**2.2. Prototypic Emotional Expressions.** The majority of FER systems do not define face characteristics in depth; instead, they begin with prototype expressions. The human universal facial expression of emotion set, which includes six kinds of fundamental emotions [15], is the most frequently used collection of prototype facial emotion expressions. It is the most often used set of facial expression prototypes. These fundamental terms are utilized because they are cross-ethnic and cultural barriers (Figure 3). This indicates that these emotions exist in everyone and may be seen in a range of situations [17]. They include fear, anger, joy, sadness, disgust, surprise, and a neutral remark. This system may be used in two ways: as a traditional classifier to identify the emotion of the person shown in the image, or as a probabilistic estimator to estimate the likelihood of the person displaying emotion in the image. In the second instance, it performs the function of a fuzzy classifier.

### 3. Face Detection and Emotion Recognition System Structure

FER may be utilized as a standalone facial recognition system or as a module inside an existing facial recognition system. As a result, it is prudent to investigate the overall design of the system. In general, the system is made up of four components, as shown in Figure 3. The face detection component is responsible for detecting whether or not a face is present in the input media.

If the input medium is video, face recognition is only done on crucial frames, with the other frames monitored using a tracking technique. This is done to enhance the system's overall resilience. Face alignment, on the other hand, is comparable to face detection in that it gives a more precise position of the identified face. This phase involves identifying face features such as a person's nose, eyes, and brows, as well as other facial features. The image is then

subjected to a technique known as geometric normalization, which alters photometric characteristics such as brightness and contrast. Then, feature extraction is utilized to categorize labels like gender, identity, or expression. The extracted feature may be sent into a classifier or compared to training data, depending on the conditions.

The components of face detection, face alignment, and feature extraction are briefly explained in Sections 3.1, 3.2, and 3.3, respectively.

**3.1. Face Detection.** Face detection is the initial stage in the face recognition process and is essential to the system's overall effectiveness [14]. Faces in films may be recognized using visual cues such as facial expression, skin tone, or movement in the film. Numerous effective methods are restricted to improving the appearance of the face [18]. This may be because these algorithms avoid the challenges associated with representing 3D structures like faces. The face/nonface border, on the other hand, may be extremely complicated, and it is necessary to use 3D variations in order to identify facial emotions. As a consequence, many solutions to this issue have been proposed since the 1990s [19].

Kenli and Ai [20] created a detection technique that uses Eigen decomposition to find abnormalities. They combine a generic face with a range of "eigenfaces". The researchers [21] distinguished this from Sung and Poggio, who focused only on 'eigenfaces'. They did, however, use Bayes' rule on nonfaces to determine the likelihood of occurrence. Rowley et al. [22] utilized neural networks to distinguish between images with and without faces, while Osuna et al. [23] trained a Kernel support vector machine to distinguish between images with and without faces. To retrain the SVM, a bootstrap approach was employed, and the results were encouraging.

Additionally, Schneiderman and Kanade [24] used AdaBoost to build a classifier based on a picture's wavelet shape. As a consequence, the method requires a significant amount of computing time. Viola and Jones [25] overcame this problem by substituting Haar features [26] for the wavelets. Haar features were computationally less expensive than wavelets. This is the first demonstration of a real-time frontal view facial recognition system [27].

Several enhancements to Viola's framework have been proposed. The Haar characteristics were rotated in-plane by Lienhart and colleagues [28]. Li et al. [29,30] suggested using a detector pyramid to cope with out-of-plane rotation, and this system may be utilized for multiview face recognition as well. Eigenface and AdaBoost were introduced as techniques for facial detection. Eigenface is considered the most simple technique for face detection, whereas AdaBoost is regarded as the most successful. AdaBoost may also be used to extract face characteristics.

**3.2. Face Alignment.** Face alignment, which involves the detection of facial feature points, may result in more accurate face localization when performed in conjunction with face localization. A comparison of face recognition algorithms and facial alignment methods is shown in Figure 3.

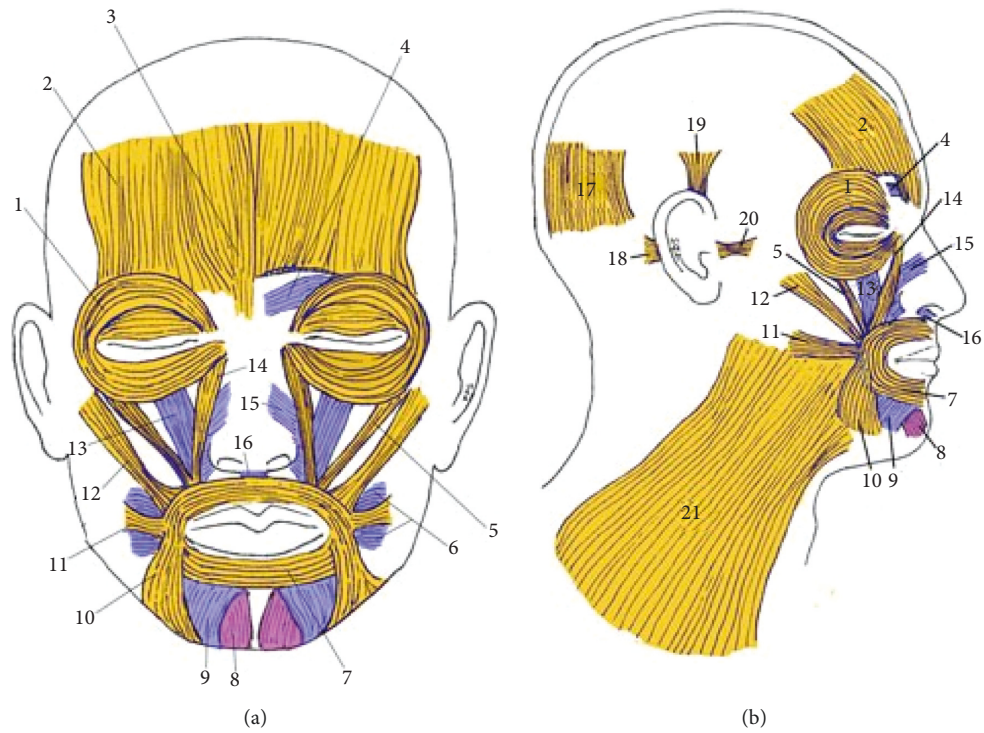


FIGURE 1: Muscles of facial expression [10].

Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink
Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

FIGURE 2: FACS action units [12].



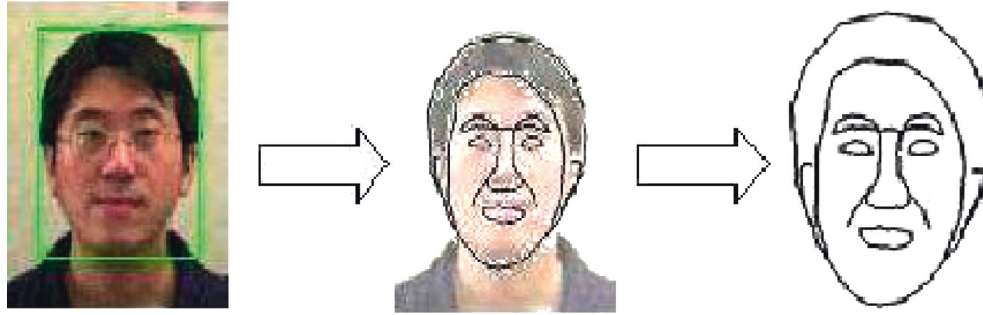


FIGURE 3: Face detection and alignment processes [16].

Face detection, as illustrated in the picture, assesses regions of an image, while facial alignment is accurate to the pixel level.

Various methods to address this problem have been suggested since the 1990s. Gu et al. [31] used histograms to identify the corners of the mouth and eyes in a picture. Marian and colleagues [32] used Gabor filters in images to identify the medial cleft and pupils. Despite the fact that many methods have been tried, the Active Shape Model [33] is the most effective of the curve fitting algorithms currently available.

Cootes et al. [33–35] proposed the Active Shape Model (ASM) for usage with images that include faces. The ASM's durability, speed, and accuracy have all increased considerably since then. Li et al. [36, 37] created the Direct Appearance Model by coupling Gabor filters with ASM, which was subsequently validated by other studies. Other authors [38] have enhanced ASM for local searches by including 2D local textures in the search results.

**3.3. Feature Extraction.** The feature extraction technique transforms pixel data into higher-level representations of the face in the image, such as texture, color, motion, contours, and the spatial arrangement of the face. This collected data is then used to assist in the detection of trends throughout future categorization procedures. During the feature extraction procedure, the dimensions of the input space are typically decreased. It is critical to retain information with high stability and discrimination while also maintaining a high level of stability during this process. To identify a person's face, a number of unique characteristics are utilized [39].

Recently, the coefficients of eigenface have been used as features; [40] used an extension of eigenface, called Tensorface, which has shown promise. The Active Appearance Model [41] deconstructs the image of the face into "shape" and "texture". The shape vector refers to the contours of the face, while the texture vector refers to the "shape-free" textures of the face. Potential Net was used by Matsuno et al. [42] to extract features with a two-dimensional mesh. The above-mentioned methods are seen as holistic features, because they look at the overall structure of an image. Another type of feature that only focuses on small regions is called local features. Local features can be used directly as image subwindows as in the case of Colmenarez et al. [43].

They used nine subwindows positioned around the facial features. Gabor Filter, a popular wavelet filter, has also been used [44, 45] and has had reasonable success regarding the visualization in the primary visual cortex. Primitive topographic features have also been used by Yin and Wei [46] to represent faces. Yu and Bhanu [47] utilized an evolutionary algorithm to automatically generate features instead of defining the features beforehand. In video-based FER, the dynamic changes of expression can also be seen as a feature. The Geometric Deformation Feature that was proposed by [48] has the ability to geometrically displace landmarked nodes. Facial Animation Parameters as used by Aleksic and Katsaggelos [1] are based on the Active Shape Model.

**3.4. Emotion Classification.** The automatic expression recognition problem has received a lot of attention with a variety of classifiers being applied to it. One solution, offered by Matsuno et al. [49], looked at the threshold of normalized Euclidean distance between features to categorize a facial expression. Another solution [43] makes use of Bayesian recognition to find facial expressions, which maximizes the likelihood of the image. Other methods include Locally Linear Embedding [50], Fisher discrimination analysis [51], and Higher-Order Singular Value Decomposition [52] to name only a few [1,53,54]. Currently, support vector machines [55–57] and neural networks [58–61] offer the best accuracy and are seen as most successful when applied to the automatic expression recognition problem.

## 4. Real-Time Applications of FER

It is only in the emotional mechanisms that a lack of progress was recently discovered [62]. It has been found that emotional mechanisms take precedence over rational processes in the brain, which can be seen as either being advantageous or detrimental depending on their presence or absence [63]. Worse feelings give rise to negative thoughts, which tend to dampen a person's creativity when looking for solutions to the challenges at hand and lead to them getting you into deeper trouble. It has been found that states such as anger, sadness, fear, and happiness each have their own distinct patterns of blood flow to the brain and have an influence on that of mood [64]. A great deal of research has demonstrated that positive emotions like joy, acceptance, trust, and satisfaction can assist learning, while negative emotions can

bring about learning disabilities and affect the process. Anxiety and depression can hinder memorization in various ways. These states can show up in different ways, such as causing stress, which increases with despair, and leads to increased feelings of anger and fear, or fear, or stress itself may lead to worse than depression. When it is difficult for a student to acquire information, intelligent feedback can help them overcome their lack of motivation. For the latter, the computer should be able to know what learners are feeling, give learners opportunities to expand on their understanding, handle their interests, and provide them with pertinent information and timely feedback [8]. Figure 4 depicts end-to-end face recognition processing flow.

According to [65], enabling students to provide students with affective and intelligent feedback in an e-learning system, for instance, involves the use of virtual persons, namely, embodied conversational agents (or other virtual agents that are capable of communicating both verbally and nonverbally, such as animated graphical characters) that can convey emotions or other sentiments and provide information to them using body language. It is much more effective in having a machine that talks to the user but does not take any of his or her reactions into consideration. Such systems may be easier said than done, but an immense challenge exists in implementing them, given our currently crude ability to recognize human feelings and behavior.

In MobileNets, they developed a set of efficient convolutional neural models, Andrew G. Howard et al. [66]. A class of efficient MobileNets convolution models were developed by him, as part of a class called MobileNets. The neural architecture has been designed with a novel type of convolution called depth separable and is made up of factors, instead of connected layers, of depthwise convolutions. Depthwise convolutions are two layers: they are separable, and the first is a depthwise layer consisting of two separable convolutions. According to normal convolution, it is over 10 times more efficient. However, it only applies lowpass and highpass effects to the signal; it does not add any new features. To do this, they did the addition of other operations such as pointwise convolution (e.g., which does the sum of pointwise convolution outputs) and implemented 1x pointwise convolution, respectively. By adopting two additional hyperparameters, they try to increase their efficiency. Thus, the network can be made thinner by the increase in network width by a multiplicative factor  $\alpha$  and resolution by a nonuniform  $\rho$ , and cost reduction can be done at each layer. Different hyperparameters allow the model builder to choose a model that will have just the right number of parameters for their own application, but without error. Various methods and functions are demonstrated by this model with different examples, with facial features as well as the measurement of object dimensionality.

Recognizing emotions is difficult because they are ambiguous and therefore prone to error, but in many cases, there are various things that can be used to discover them. Ekman claims that there are eleven basic emotions found in human facial expressions which can be classified into seven groups: happiness, anger, sadness, fear, disgust, surprise, and contempt [67]. This after the turn of the millennium got a

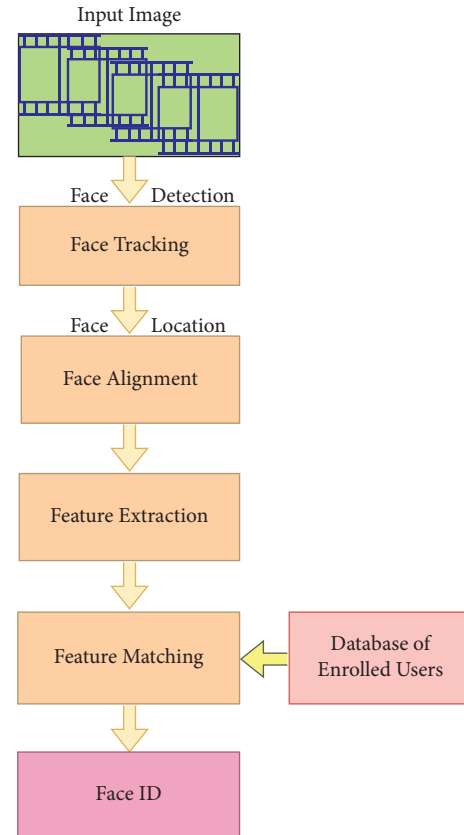


FIGURE 4: Face recognition processing flow.

boost from successful experiments in face recognition and audio-visual media and has paved the way for further research on automatic affect recognition of effect. Face expression popularly suggests that recognition of emotions is simply a method to look for patterns that indicate whether or not a person is empathetic popularly known as “to judge by the look on their face”. The FACS is used to code different types of facial actions, which include facial motions, and numerous AUs are created; each one of these classes contains unique entities; finally, emotions are determined by AU designations.

Bartlett and Colback have done extensive research on emotional recognition results, which is accurate and comprehensive, as Bartlett and Mattivi [68] reported. Features that are based on geometry (e.g., the shape of the eye, or the angle of the eyebrows) measure the face’s facial elements such as the length of the nose or the width of the mouth. Empirical approaches use different machine learning algorithms to analyze the detected face, all of which are capable of determining features, with the goal of classifying it into an emotional state. At this time, it is possible to recognize emotions using facial expressions using the information from the applications that were developed by [69]. Six basic emotions are identified by the facial analysis software tool known as the Face Reader, with an accuracy of 89%. In the work referenced in [70], facial recognition researchers have done significant progress by leveraging local characteristics in a database of people. A face can reveal both stress levels as

well as the degree of emotional interest as well as being energetic endurance, which is something that is not widely understood.

There is also significant information that can be gleaned from gestures and postures in regard to an individual's emotional state and attentive state. This research is lacking; however, these topics have not been extensively studied. When applied to the analysis of the user's previous and current interactions with the web data, information from the previously stated sources can help prove the user's current cognitive state [71]. Generally speaking, the development of an affective guidance system depends on emotional state and entertainment in addition to the study of the various kinds of influencing factors, with the result that feedback is appropriate. Robotic tutors are expected to offer virtual classes in a redesigned, so they can better reflect their pupils' personalities and respond to pupils' emotional states. The researchers in [72] contend that the presence of frustration, boredom, motivation, and confidence is equally essential for a computer tutor, and they conduct an analysis of each method of feedback they have used to gauge the other. Basic emotions like fear, sadness, and happiness are found in [73] articles by the authors. A style of ECAs known as "expansive empathy" is performed before performing "reactive empathy" which is then presented with expression and voice.

Although studies have demonstrated that automatic expansion is the most difficult; here, the authors show different views of the complexity of expanding every case individually (possible), a complexity scale, formed by Philipp et al. [9] that shows different techniques in use. The critical variables to consider when using this technique include head pose, skin condition, and age, which may vary depending on when and if one is taking photos in high light versus low light conditions, as well as their position, and additionally the issue of occlusion created by the scarf or other illumination. Several techniques are used for the extraction of facial features, for example, geometric features, and texture features, for example, LBP [74] and Gaborlet unit activity, while the Generalized Local Binary Pattern Classification (LBC) and Directionalized Gabor (GDA) are used to extract facial landmarks [75]. Since it has started being widely used in recent years, mainly due to the application of convolutional neural networks and recurrent neural networks, it has proven to be a very successful and efficient technique for emotion recognition. Several neural networks have been developed in this area to help with the development of deep architectures, all of which produce commendable results [76].

## 5. Current Problems/Challenges in Face Detection and Emotion Recognition

This section explores cutting-edge methods for analyzing and understanding facial expressions. The following essential issues must be solved while designing a facial emotion recognition system: face detection and alignment, normalization of the facial image, extraction of critical attributes, and, finally, classification. Currently, the majority of systems conduct these processes sequentially and independently. As

a result, this section will first discuss the problems associated with identifying facial emotions, followed by an examination of how the aforementioned processes have been handled in various studies.

However, distinguishing the features of a human face and interpreting its emotional state are both difficult undertakings. The fundamental problem originates from the nonuniformity of the human face, as well as additional restrictions connected with lighting, shadows, facial location, and orientation concerns in various circumstances [77]. Humans are born with the ability to perceive and grasp facial expressions and emotions with little or no effort; nevertheless, computer systems continue to face significant hurdles in recognizing effective and robust facial expressions. Numerous deep learning techniques, including multilayer perceptron (MLP) neural networks and support vector machines, have been studied as a family of techniques for improving the robustness and performance of fundamental machine learning classification techniques, such as MLP neural networks. To be effective, human behavior analysis must be adaptable to a wide range of contexts. Deep learning algorithms may be able to deliver the required robustness and scalability on new types of data.

The parts that follow will go through the most important challenges in automated facial expression recognition in detail. In this situation, acquiring task-representative data, overcoming ground truth collection challenges, dealing with occlusions, and modeling dynamics are all key difficulties.

Procedures utilized in standard FER techniques are depicted in Figure 5: face region and facial landmarks are detected in input images, spatial and temporal features are extracted from the face components and landmarks, and facial expression is determined using pretrained pattern classifiers based on one of the facial categories (face images are taken from the CK + dataset [78]).

## 6. Databases Used for Facial Emotion Recognition

Facial recognition is getting better and more prevalent each year, and consequently, facial databases have expanded tremendously [79]. When modeling of recognition requires visual or audio examples, you have to give, model enhancement or training of the model requires a database of those kinds, and this, as well as class labels for them, which gets progressively larger and larger as the number of examples increases, expand is required [80]. For example, there are various possible applications for emotional recognition, ranging from simple human-robot collaboration [81] to being used to identify people suffering from depression to serving as a depression detector [82].

Although the algorithm most commonly accepts image/portrait datasets, which are uniformly lit and fixed in position, this form, in the top portion, an alternative version is one where the top and bottom halves are aligned but cropped differently. To be able to compare to the pixelated versions, the NIST mugshot database [83] also offers a clear, grayscale option for finding image IDs of 1573 individuals on a neutral background. But it also takes the authors to go

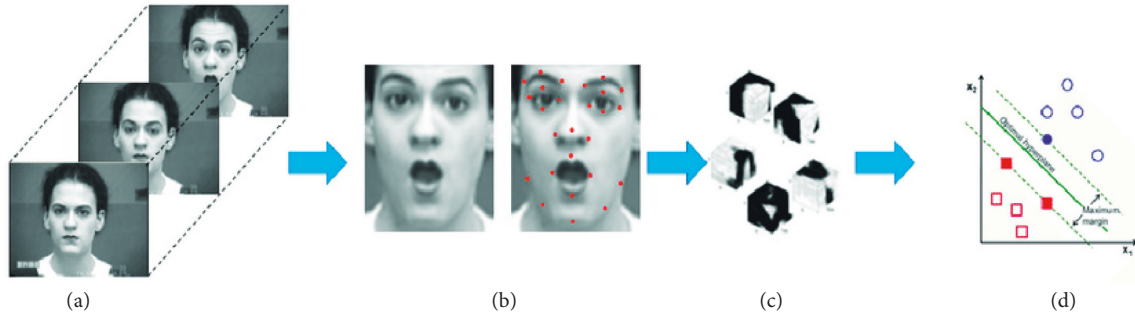


FIGURE 5: Conventional FER method [28]. (a) Input images. (b) Face detection and landmark detection. (c) Feature extraction. (d) FE classification.

out into the real world to realize how light conditions and occlusions work in the context of real life in order to comprehend the situations [84]. Using this method [85], the subjects were easily rotated and the effects of different lighting on their appearance were studied in the M2VTS database which features the faces of 37 subjects in a variety of rotated and lit positions. The emotions included in a database define its function. Many databases, like CK, MMI, eNTERFACE, and NVIE, opt to record the six basic emotion categories proposed by Ekman. Many databases, like the SMO, AAI, and ISL meeting corpus, try to classify or contain general positive and negative emotions. Some try to evaluate deception and honesty, such as the CSC corpus database. The most well-known 3D datasets are BU-3DFE, BU-4DFE, Bosphorus, and BP4D. BU-3DFE and BU-4DFE both have six-expression posed datasets, with the latter having a higher resolution. Bosphorus tries to address the issue of having a wider range of facial expressions, while BP4D is the only one of the four that uses induced rather than posed emotions. The main benefit of deep learning is that it opens the neural networks to various databases, allowing them to grow with the addition of a wide range of new inputs, examples, facial expressions, and constant changes in expressions.

## 7. Facial Expression, Speech Emotion, and Multimodal Emotion Recognition

We have covered the three main types of emotion recognition techniques in this section: facial expression recognition, speech emotion recognition, and multimodal emotion recognition utilizing visual representations. We have also spoken about how these methods may be used in a variety of situations.

**7.1. Facial Expression Recognition.** Facial movements are essential in nonverbal communication for expressing emotions. Facial expression recognition is important in a broad range of applications, including human-computer interaction and health care. Mehrabian discovered that 7% of information is conveyed between people via writing, 38% through conversation, and 55% through facial expression. Ekman and colleagues [86] defined six basic emotions: pleasure, sadness, surprise, fear, and anger. He demonstrated that people, regardless of culture, have these emotions. Feldman et al. [32]

propose that emotions may be represented using two orthogonal dimensions: valence and arousal. He observed that everyone displays their feelings in a unique way.

Furthermore, when people are asked to express their feelings on a regular basis, their responses vary greatly [87]. Arousal levels vary from calm to eager, while valences range from positive to negative [88]. This research would categorize the information based on valence and arousal changes. Researchers first developed methods for manually extracting facial expressions by developing algorithms for extracted functions such as the Gabor wavelet, the Weber Local Descriptor (WLD), the Local Binary Pattern (LBP), and multifeature fusion. These properties are susceptible to topic imbalances and may result in a substantial loss of texture information from the original image. The use of deep neural network models to face expression analysis is now the most popular subject in facial recognition. Furthermore, FER provides a wide range of social life applications, such as intelligent protection, deception detection, and intelligent medical practice. The authors of [89] discussed facial expression recognition models developed using deep learning methods such as DBN, deep CNN, and long short-term memory (LSTM) [90], as well as their combination.

**7.2. Speech Emotion Recognition.** Speech recognition is a key component of human-computer interaction systems. They will communicate their feelings via their words and facial expressions. Emotions are often identified using speech recognition algorithms [91]. The early attempts at emotion detection in a speech focused on categorizing speech by extracting artificial features. Liscombe et al. (2003) looked at the relationship between various emotions and a set of continuous speech parameters based on basic pitch, amplitude, and spectral tilt. Throughout the years, many algorithms for detecting emotions in human speech have been created [92]. Many machine learning techniques have been proposed, including support vector machines, hidden Markov models, and Gaussian mixture models. Deep learning has been widely used in a wide range of speech domains, most notably voice recognition [93]. Convolutional neural networks have also been used to detect emotions in speech; they show that bi-directional multimodal emotion recognitional RNNs (Bi-LSTM) are more successful at extracting important speech characteristics, thereby increasing speech recognition



performance [1]. Figure 6 illustrates the end-to-end “Speech Emotion Recognition” system.

**7.3. Multimodal Emotion Recognition.** In research, multimodal emotion processing is still extensively utilized. Through the utilization of new study modalities, this expansion would assist in a better understanding of emotions (video, audio, sensor data, etc.). To achieve the study’s goal, a variety of techniques and tactics are used. Many of them use big data techniques, semantics, and deep learning. Emotions are complicated psychophysiological processes that occur nonverbally, making identification difficult. Multimodal learning is much more effective than unimodal learning [94].

There is a basis for a neural network for multimodal emotion recognition, with a focus on visual input in recognized faces. Their approach was inspired by the winners of the 2013 and 2014 EmotiW challenges. Chen et al. [95] proposed their approach in answer to a multimodal emotion detection issue (MEC 2016). This technique retrieves multimodal features in order to determine the emotion of the character in the video. Among them, the facial CNN feature has the highest discriminative power for emotion recognition. Previously [96], we retrieved several features using both traditional and deep convolutional neural network (DCNN) methods. On testing sets, this approach delivers an exceptionally promising result. We describe the techniques used to generate the team submissions for Beijing Normal University’s 2017 Multimodal Emotion Recognition Challenge (MEC 2017). Many features were retrieved, including an autoencoder (AE), a CNN, a dense SIFT, and an audio feature, and a Dempster-Shafer theory fusion method was provided for merging different prediction results based on these features. The framework for multimodal emotion recognition NN is shown in Figure 7.

Furthermore, research has tried to combine data from different modalities, such as facial expressions and audio, audio and written text, physiological signals, and various combinations of these modalities [97]. This technique is currently being enhanced to increase the accuracy of emotion detection. A multimodal fusion model may generate emotion detection results by integrating physiological data in a number of ways. Recent advancements in deep learning (DL) architectures have made it possible to use deep learning for multimodal emotion recognition. The deep belief network, the deep convolutional neural network, the LSTM [55], the support vector machine (SVM) [98], and their combinations are all deep learning techniques.

## 8. Deep Learning-Based Facial Emotion Recognition

Deep learning algorithms have recently emerged as a viable replacement to traditional feature design techniques since they provide automated feature learning instantly. Deep learning research may lead to better representations and the creation of new models for learning these representations from unlabeled data. Because of the introduction of powerful GPU processors that allow high-performance numerical

computing in graphics cards, these techniques have become computationally feasible. Deep learning techniques such as convolutional neural networks, deep Boltzmann machines, deep belief networks, and stacked autoencoders are used in practical applications such as pattern analysis, audio recognition, computer vision, and image recognition, producing impressive results on a variety of tasks. Li et al. [100] provided a thorough evaluation of the aforementioned DL techniques adapted to the FER issue lately. Ginne et al. [5] have given an overview of CNN-based FER techniques. Deep convolution has been used extensively in FER research. The focus of network research has been on improving expression recognition accuracy. It is important to observe how, at the end of the day, a smaller CNN architecture with the same level of accuracy is feasible. Deep convolution neural networks based on FER are depicted in Figure 8 providing more effective scattered training, as well as a more controllable parameter model, as well as improved deployment suitability on memory-constrained devices, reducing costs and allowing for greater distribution.

In the last decade, CNNs have done well for FER, as shown by their use in a number of cutting-edge algorithms. Many FER competitions [101], including the previous year’s EmotiW challenge, were won by a kind of CNN architecture with few layers. Facial emotion recognition has served the public well for decades prior to the field of deep learning breaking, and a group of brilliant researchers has tried to stay abreast of the current research efforts in that field, while others have undertaken to learn from its methods and discoveries. In recent times, many researchers offered novel and recurring practices for applying deep learning in order to security problems in an effort to enhance detection. Validation users currently do additional validation on a number of static or sequential databases before allowing their information to be used in a live database.

The VGG-16 model (developed by the University of Oxford’s Visual Geometry Group (VGG)) may be considered a watershed moment in the history of deep CNN models [102]. It was pretrained using the ImageNet database [103] to extract features from images that might be used to distinguish between image classes. Numerous recent studies show that VGG-16 performs well on image recognition and classification datasets from a variety of fields.

Marco et al. [104] proposed Deep Convolution Neural Networks (DCNNs) which are used in the cross-database search. After that, facial images had to be reduced to  $48 \times 48$  pixels; the rest of the same pictures had to be searched for locations and landmarks to be extracted. Finally, they had augmented the database with additional data, and only then did they were able to create it. Subsequently, the data moves on to two classification stages where the softmax (SF) is expanded and fed into the fully connected softmax (XF) network after the first classification stage. To avoid overfitting, they suggest using local CNNs in combination with convolutional layers that are fine-tuned for specific use cases.

In [105], the authors have shown that the results prior to training were used to discover how to influence the final outcome. When it expanded, the first CNN expansion, when it lowered the size to  $32 \times 32$  and also used data

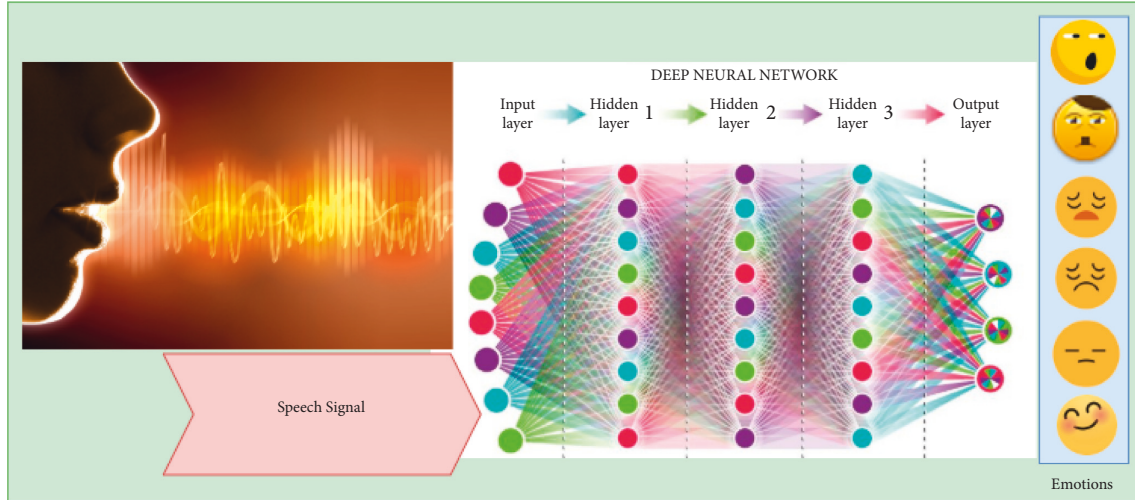


FIGURE 6: Speech emotion recognition [8].

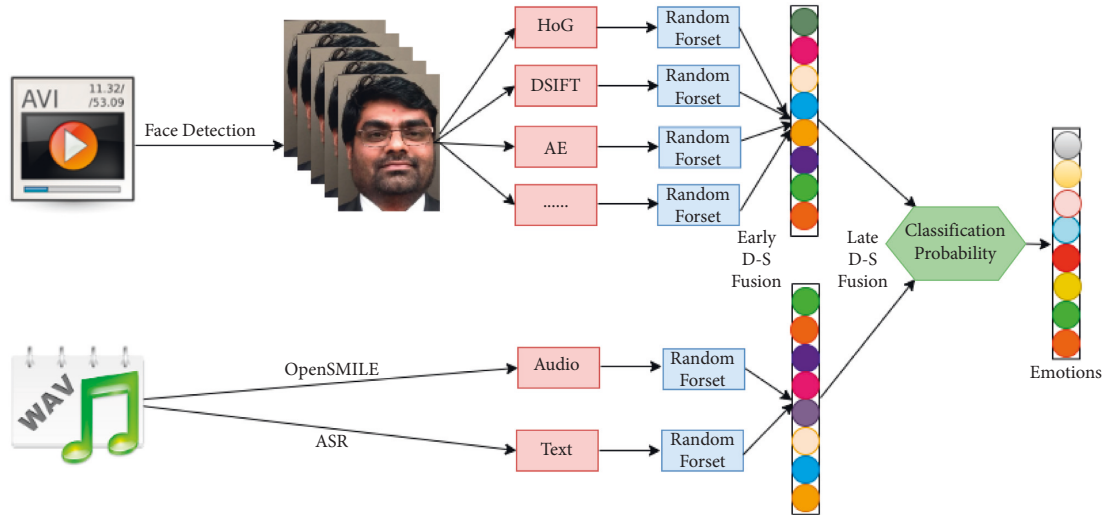


FIGURE 7: The framework for multimodal emotion recognition NN [94].

normalization with 8-connected pools followed by down-sampling (normalization of  $32 \times 32$  to a 256 final dimension), and when that was done, cropping was employed. Gaining the most mass is something that happens only at the competition, so the athletes who have gained the most muscle will play in the games. The information used this third party search tool to assemble a total of three transparently accessible databases: the CK+ and JAFFE, as well as the BU3DF. One also discovers a wide range of beneficial practices when considering these studies, such as utilizing all of these techniques and products together; studies show the difference between these things yielding different results.

The preprocessing techniques employed by Anil and his associates [8] have also been applied in the study by the authors. They are devising a new CNN face recognition algorithm for people who have not yet been recognized. They have two convolutions allowing layers and a dropout layer

that gives the net activation of one in order to predict more accurate results. They use maxing with one extra activation as well and a final convolution (expanding) layer in the last step to increase accuracy and flexibility.

An important concern raised by Cai et al. [106] deals with the fact that the closing or the disappearance of small-town public libraries is managed by solving CNN, which employs Sparsity Batch Normalization. Dropout may be added to network building to help against overfitting and SBP (Support, Gradient, and Regularization) as a second stage to improve model generalization capacity, with the property of being used in networks twice (as a support for and then starting with 2-convol reg and ending with SGD), to strengthen the network. Li et al. [107] proposed using a CNN to tackle the facial distortion problem, in particular; the authors are doing so by first extracting the data from the VGG network and then running the ACN. *Affect*. Also, this



To figure out the significance of the CNN attributes in FER2013, researchers investigated and added to the already discovered findings of Agrawal et al. [109] (this also included research on Agrawalwarsh et al.) in 2019. Beyond that is an image memory pool at  $64 \times 64$  pixel resolution, the network will have a certain type of an allowable number of convolution layers, and ad hoc pooling will take the second position, followed by other admissible filters before classifiers. The results of the study demonstrate that determining models achieve a 61.23% and 63.77% of their accuracy using isolated units, compared to adjudging, or dropout models, but do not have well-connected layers.

In the facial expression sequence level, DCNN was postulated by Liang [111] consisting of two deep layers, one of which handles spatial features and the other temporal features, which are treated as features that are then merged and expanded into vectors of 256 dimensions to form the large facial emotion category vector; that is, the expression differentiated into six basic emotions is utilized. They went through the Multitask Cascade Computational Net for face detection, after which they broadened the database with the technique of data augmentation. It is based on all of the scientists' opinions about classifying the basic emotions that

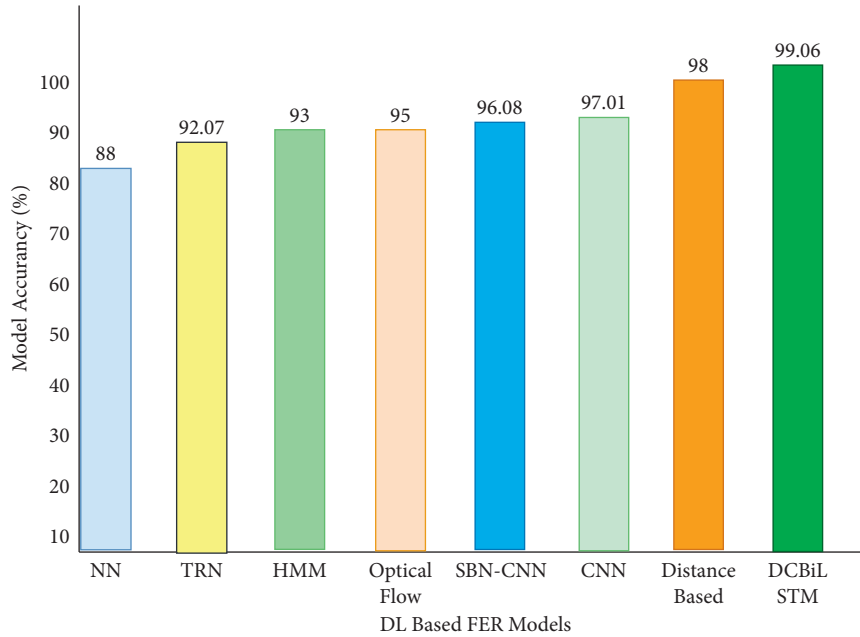


FIGURE 9: Deep learning-based FER models.

previously state that the emotion categories are happiness, fear, surprise, sadness, and neutral. Instead of presenting the theories and thoughts of other researchers, we exhibit different proposals from those who recommended.

## 9. Comparative Analysis and Discussion on FER

We made it clear in this paper that there is a strong demand for expanded research on the topic of FER far beyond shallow learning techniques. The full form of the automatic FER goes through four main data processing, a few proposed architectures, and finally getting to the main model's emotion recognition. Many preprocessing techniques were mentioned in this review, such as the cropping and resizing of images to speed up training and normalization, as well as the overfitting of data. Lopes et al. [120] have presented all these techniques well, according to Lopes and his colleagues. Figure 9 illustrates deep learning-based FER models. The various methods and contributions presented in this review achieved high levels of accuracy. Moselhi et al. [121] demonstrated the key significance of the use of neural networks and connectivity expansion layers in neural network architectures. Like many before them, the authors Mohammadpour et al. [122] prefer to extract AU from the face, rather than do face-to-recognition first, first. This study is being conducted to determine whether occlusion images exist or not, as well as to try to gain greater insight into the network. Pise et al. [8] have examined the incorporation of the leftover blocks. While text images allow only for larger eyes and smaller faces, the addition of the iconized face to the

network improves accuracy when using small images, as is demonstrated by Yolu and Ayiv [108]. Two-concept CNN architecture expansion was added after a long and thorough analysis of the recognition rate by offering two more feature articles to know the impact of CNN parameters. Favorable results have been observed in most of the methods attempted, which means more than 90% of these projects achieved some level of success. Researchers who study spatial and temporal features first provided several combinations: the combination of CNN-L and 3D-CNN is typically applied to give a boost to spatial features but boosts temporal features too. As can be demonstrated according to the work of Yu and his colleagues, the methods proposed by Kim et al. [118] and Liang et al. [111] provide a better level of precision than the one that was performed by the Kim group [118]. That comes out to an effective volume expansion factor of about 99%.

In CNN applications, both temporal and spatial networks have demonstrated their accuracy. That is why the researchers chose LSTM, which is effective for sequential data in general, but especially for time-dependent data in order to achieve high accuracy in FER. To date, CNN parametric modeling and the most difficult algorithms used by CNN researchers are softmax and Adam optimization. To validate the neural network architecture, we also tested the model across multiple databases, and our findings indicate that there are no significant differences in results.

Table 1 summarizes the arguments made previously, with particular emphasis on the architecture, database, and recognition rate discussed in the linked articles.



TABLE 1: Comparison between FER models.

Approach	Technique	Groups	Sub	Authors	Acc (%)
DCBiLSTM	Fusion	6	123	Liang et al. [111]	99.6
Dist-based	Optical flow	5	8	Essa & pentland [112]	98
CNN	Facial AUs	7	123	Hashemi et al.,	97.01
SBN-CNN	Batch norm	7	10	Wei et al., [113]	96.8
Rule-based	Optical flow	6	32	Yacoob & davis [114]	95
HMM	2-D FT optical flow	6	4	Otsuka & Ohya	93
TRN	Relational reasoning	8	27	Pise et al. [8,115]	92.7
Rule-based	Parametric model	6	40	Black & Yacoob [116]	92
NN	Optical flow	2	32	Rosenblum et al. [117]	88

## 10. Conclusion

In this study, recent developments in FER were presented, and presented studies allowed us to track the latest developments in this area. Over the past year or so, a number of different researchers have devised different CNN architectures and some outside the lab produced reference databases. To facilitate an accurate emotional detection, we need to have provided previously obtained as well as experimental tables (spontaneous as well as lab) (emotion as reference). Also, we introduce a discussion which emphasizes the fact that machines are already able to recognize more complex emotions, implying that the emergence of human-machine collaboration will become more and more commonplace.

## 11. Future Work

While FER is an important source of information about an individual's emotional state, it is always limited by learning only the six basic emotions plus neutral. It is in conflict with what is present in everyday life, which contains more complex emotions. This will encourage researchers to expand their databases and develop powerful deep learning architectures capable of recognizing all basic and secondary emotions in the future. Additionally, emotion recognition has evolved from a unimodal analysis to a complex system multimodal analysis in the modern era. Leon et al. in [123] demonstrate that multimodality is a necessary condition for optimal emotion detection. Researchers are now focusing their efforts on developing and commercializing powerful multimodal deep learning architectures and databases, such as the fusion of audio and visual modalities investigated by Zhang et al. [124] and Ringeval et al. [125] for audio-visual and physiological modalities.

## Data Availability

The data that support the findings of this study are available on request from the corresponding author.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors extend their appreciation to King Saud University for funding this work through Researchers Supporting Project number (RSP-2022R426), King Saud University, Riyadh, Saudi Arabia.

## References

- [1] P. S. Aleksic and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream hmms," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 1, pp. 3–11, 2006.
- [2] H. Chouhayebi, J. Riffi, M. A. Mahraz, Y. Ali, and T. Hamid, "Facial expression recognition using machine learning," in *Proceedings of the 2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS)*, pp. 1–6, IEEE, Seoul, Korea, November 2021.
- [3] Anil. Audumbar Pise, H. Vadapalli, and I. Sanders, "Relational reasoning using neural networks: a survey," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 29, pp. 237–258, 2021.
- [4] J. Kim, M. Ricci, and T. Serre, "Not-So-CLEVR: learning same-different relations strains feedforward neural networks," *Interface focus*, vol. 8, no. 4, 2018.
- [5] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: a survey of registration, representation, and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1113–1133, 2014.
- [6] R. Brooks, "A robust layered control system for a mobile robot," *IEEE Journal of Robotics and Automation*, vol. 2, no. 1, pp. 14–23, 1986.
- [7] C. Chen, K. Li, S. G. Teo, X. Zou, K. Li, and Z. Zeng, "Citywide traffic flow prediction based on multiple gated spatio-temporal convolutional neural networks," *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 4, pp. 1–23, 2020.
- [8] A. Pise, H. Vadapalli, and I. Sanders, "Facial Emotion Recognition Using Temporal Relational Network: An Application to E-Learning," *Multimedia Tools and Applications*, pp. 1–21, 2020.
- [9] P. V. Rouast, M. Adam, and C. Raymond, "Deep Learning for Human Affect Recognition: Insights and New Developments," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, 2019.
- [10] M. Anne and J. F. Cohn, *Anatomy of Face*, 2009.

- [11] E. Paul, D. Matsumoto, and V. F. Wallace, "Facial expression in affective disorders," *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*, vol. 2, pp. 331–342, 1997.
- [12] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of the the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pp. 46–53, IEEE, Grenoble, France, March 2000.
- [13] J. Mei, K. Li, and K. Li, "Customer-satisfaction-aware optimal multiserver configuration for profit maximization in cloud computing," *IEEE Transactions on Sustainable Computing*, vol. 2, no. 1, pp. 17–29, 2017.
- [14] V. F. Wallace, E. Paul et al., "Emfacs-7: emotional facial action coding system," *University of California at San Francisco*, vol. 2, no. 36, p. 1, 1983.
- [15] G. Schubert, "Human ethology and evolutionary epistemology: the strange case of dmEibesfeldthuman eieeyaPp. xvi, 848, \$69.95," *Journal of Social and Biological Systems*, vol. 13, no. 4, pp. 355–387, 1990.
- [16] E. S. Jaha and L. Ghouti, "Color face recognition using quaternion pca," in *Proceedings of the the 4th International Conference on Imaging for Crime Detection and Prevention 2011*, pp. 1–6, ICDP, 2011.
- [17] P. Ekman, Di Perrett, and H. D. Ellis, "Facial expressions of emotion: an old controversy and new findings: Discussion," *Philosophical Transactions of the Royal Society of London, Series A B*, vol. 335, p. 69, 1992.
- [18] X. Zhou, K. Li, G. Xiao, Y. Zhou, and K. Li, "Top f probabilistic products queries," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 10, pp. 2808–2821, 2016.
- [19] M. Bichsel and A. P. Pentland, "Human face recognition and the face image Set's topology," *CVGIP: Image Understanding*, vol. 59, no. 2, pp. 254–261, 1994.
- [20] K. Li, W. Ai, Z. Tang et al., "Hadoop recognition of biomedical named entity using conditional random fields," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 11, pp. 3040–3051, 2014.
- [21] X. Tang, K. Li, Z. Zeng, and B. Veeravalli, "A novel security-driven scheduling algorithm for precedence-constrained tasks in heterogeneous distributed systems," *IEEE Transactions on Computers*, vol. 60, no. 7, pp. 1017–1029, 2010.
- [22] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, 1998.
- [23] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," in *Proceedings of the the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, CVPR '97, p. 130, IEEE Computer Society, Washington, DC, USA, 1997.
- [24] J. Chen, K. Li, Z. Tang et al., "A parallel random forest algorithm for big data in a spark cloud computing environment," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 4, pp. 919–933, 2016.
- [25] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR, p. I, Kauai, HI, USA, Dec 2001*.
- [26] C. Franklin, "Crow. Summed-area tables for texture mapping," in *Proceedings of the the 11th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '84*, pp. 207–212, ACM, New York, NY, USA, January 1984.
- [27] H. Li, K. Li, J. An, and K. Li, "Msgd: a novel matrix factorization approach for large-scale collaborative filtering recommender systems on gpus," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 7, pp. 1530–1544, 2017.
- [28] F. Liu, X. Lin, S. Z. Li, and Y. Shi, "Multi-modal face tracking using bayesian network," in *Proceedings of the the 2003 IEEE International SOI Conference. Proceedings (Cat. No.03CH37443)*, pp. 135–142, Nice, France, October 2003.
- [29] S. Z. Li and Z. Q. Zhenqiu Zhang, "Floatboost learning and statistical face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1112–1123, 2004.
- [30] Y. Xu, K. Li, L. He, L. Zhang, and K. Li, "A hybrid chemical reaction optimization scheme for task scheduling on heterogeneous computing systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 12, pp. 3208–3222, 2014.
- [31] H. Gu, G. Su, and Du Cheng, "Feature points extraction from faces," *Image and vision computing NZ*, vol. 26, pp. 154–158, 2003.
- [32] I. R. Fasel, M. S. Bartlett, and J. R. Movellan, "A comparison of gabor filter methods for automatic detection of facial landmarks," in *Proceedings of the Fifth IEEE international conference on automatic face gesture recognition*, pp. 242–246, IEEE, Washington, DC, USA, May 2002.
- [33] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
- [34] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743–756, 1997.
- [35] T. F. Cootes, C. J. Taylor, and A. Lanitis, "Multi-resolution search with active shape models," in *Proceedings of the 12th International Conference on Pattern Recognition*, pp. 610–612, IEEE, Jerusalem, Israel, October 1994.
- [36] F. Jiao, S. Li, Heung-Yeung. Shum, and S. Dale, "Face alignment using statistical models and wavelet features," in *Proceedings of the 2003 IEEE computer society conference on computer vision and pattern recognition*, vol. 1, p. I, IEEE, Madison, WI, USA, June 2003.
- [37] Z. Stan, H. J. Zhang, and Q. S. Cheng, "Multi-view face alignment using direct appearance models," in *Proceedings of the the Fifth IEEE International Conference on Automatic Face Gesture Recognition*, pp. 324–329, IEEE, Washington, DC, USA, May 2002.
- [38] Y. Chen, K. Li, W. Yang, G. Xiao, X. Xie, and T. Li, "Performance-aware model for sparse matrix-matrix multiplication on the sunway taihulight supercomputer," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 4, pp. 923–938, 2018.
- [39] R. Xue, S. Yu, and X. Zhang, "Identification of parameters in 2d-fem of valve piping system within npp utilizing seismic response," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 789–805, 2020.
- [40] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *European Conference on Computer Vision*, pp. 447–460, Springer, Berlin, Germany, 2002.
- [41] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

- [42] S. Kimura and M. Yachida, "Facial expression recognition and its degree estimation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 295–300, IEEE, San Juan, PR, USA, June 1997.
- [43] A. Colmenarez, B. Frey, and T. S. Huang, "A probabilistic framework for embedded face and facial expression recognition," in *Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. PR00149)*, pp. 592–597, IEEE, Fort Collins, CO, USA, June 1999.
- [44] R. L. De Valois and K. K. De Valois, "Spatial vision," *Annual Review of Psychology*, vol. 31, no. 1, pp. 309–341, 1980.
- [45] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex," *Journal of Neurophysiology*, vol. 58, no. 6, pp. 1233–1258, 1987.
- [46] Y. Xie, L. Hu, X. Chen, J. Feng, and D. Zhang, "Auxiliary diagnosis based on the knowledge graph of tcm syndrome," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 481–494, 2020.
- [47] J. Yu and B. Bhanu, "Evolutionary feature synthesis for facial expression recognition," *Pattern Recognition Letters*, vol. 27, no. 11, pp. 1289–1298, 2006.
- [48] I. Arpacı, S. Alshehbi, M. Al-Emran et al., "Analysis of twitter data using evolutionary clustering during the covid-19 pandemic," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 193–204, 2020.
- [49] K. Matsuno, C.-W. Lee, S. Kimura, and S. Tsuji, "Automatic recognition of human facial expressions," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 352–359, IEEE, Cambridge, MA, USA, June 1995.
- [50] Y. Shinohara and N. I. Otsu, "Facial expression recognition using Fisher weight maps," in *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings*, pp. 499–504, IEEE, Seoul, Korea, May 2004.
- [51] Yu-K. Wu and S.-H. Lai, "Facial expression recognition based on supervised lle analysis of optical flow and ratio image," in *Proceedings of the International Computer Symposium*, Taipei, Taiwan, 2006.
- [52] H. Wang, "Facial expression decomposition," in *Proceedings of the ninth IEEE international conference on computer vision*, pp. 958–965, IEEE, Nice, France, October 2003.
- [53] L. Yin and X. Wei, "Multi-scale primal feature based facial expression modeling and identification," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pp. 603–608, IEEE, Southampton, UK, April 2006.
- [54] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 172–187, 2006.
- [55] Q. Xu, P. Zhang, W. Pei, L. Yang, and Z. He, "A facial expression recognition approach based on confusion-crossed support vector machine tree," in *Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia*, pp. 309–312, IEEE, Pasadena, CA, USA, December 2006.
- [56] P. Michel and R. El Kaliouby, "Real time facial expression recognition in video using support vector machines," in *Proceedings of the 5th international conference on Multimodal interfaces*, pp. 258–264, ACM, Vancouver British Columbia Canada, November 2003.
- [57] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real time face detection and facial expression recognition: development and applications to human computer interaction," in *Proceedings of the 2003 Conference on computer vision and pattern recognition workshop*, p. 53, April 2003.
- [58] L. Ma and K. Khorasani, "Facial expression recognition using constructive feedforward neural networks," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 3, pp. 1588–1595, 2004.
- [59] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Proceedings of the Third IEEE International Conference on Automatic face and gesture recognition*, pp. 454–459, IEEE, Nara, Japan, April 1998.
- [60] H. Kobayashi, A. Tange, and F. Hara, "Space robot. Real-time recognition of 6 basic facial expressions," *Journal of the robotics society of Japan*, vol. 14, no. 7, pp. 994–1002, 1996.
- [61] T. Ichimura, S. Oeda, and T. Yamashita, "Construction of emotional space from facial expression by parallel sand glass type neural networks," in *Proceedings of the the 2002 International Joint Conference on Neural Networks*, pp. 2422–2427, IEEE, Honolulu, HI, USA, May 2002.
- [62] R. W. Picard, S. Papert, W. Bender et al., "Affective learning - a manifesto," *BT Technology Journal*, vol. 22, no. 4, pp. 253–269, 2004.
- [63] M. Rafiq, A. Ahmadian, A. Raza, D. Baleanu, M. Sarwar Ahsan, and M. Hasan Abdul Sathar, "Numerical control measures of stochastic malaria epidemic model," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 33–51, 2020.
- [64] C. N. Moridis and A. A. Economides, "Toward computer-aided affective learning systems: a literature review," *Journal of Educational Computing Research*, vol. 39, no. 4, pp. 313–337, 2008.
- [65] Y. Tan, J. Qin, X. Xiang, C. Zhang, and Z. Wang, "Coverless Steganography Based on Motion Analysis of Video," *Security and Communication Networks*, vol. 2021, 2021.
- [66] N. Yuan, C. Jia, J. Lu et al., "A drl-based container placement scheme with auxiliary tasks," *Computers, Materials & Continua*, vol. 64, no. 3, pp. 1657–1671, 2020.
- [67] B. Yan, J. Wang, Z. Zhang et al., "An improved method for the fitting and prediction of the number of covid-19 confirmed cases based on lstm," *Computers, Materials & Continua*, vol. 64, no. 3, 2020.
- [68] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 568–573, IEEE, San Diego, CA, USA, June 2005.
- [69] Mj Den Uyl and H. Van Kuilenburg, "The facereader: online facial expression recognition," in *Proceedings of the measuring behavior*, vol. 30, , pp. 589–590, Citeseer, 2005.
- [70] S. L. Happy, A. George, and A. Routray, "A real time facial expression classification system using local binary patterns," in *Proceedings of the 2012 4th International conference on intelligent human computer interaction (IHCI)*, pp. 1–5, IEEE, Kharagpur, India, December 2012.
- [71] P. shou Xie, G. qiang Ma, T. Feng, Y. Yan, and X. ming Han, "Behavioral feature and correlative detection of multiple types of node in the internet of vehicles," *Computers, Materials & Continua*, vol. 64, no. 2, pp. 1127–1137, 2020.
- [72] F. Ao, Z. Gao, X. Song, Ke Ke, T. Xu, and X. Zhang, "Modeling multi-targets sentiment classification via graph



- convolutional networks and auxiliary relation,” *Computers, Materials & Continua*, vol. 2020, 2020.
- [73] Y. C. Shiah, S.-C. Huang, and M. R. Hematiyan, “Efficient 2d analysis of interfacial thermoelastic stresses in multiply bonded anisotropic composites with thin adhesives,” *Computers, Materials & Continua*, vol. 64, no. 2, pp. 701–727, 2020.
- [74] S. Zhang, X. Zhao, and B. Lei, “Facial expression recognition based on local binary patterns and local Fisher discriminant analysis,” *WSEAS transactions on signal processing*, vol. 8, no. 1, pp. 21–31, 2012.
- [75] S. Zhou, M. Ke, and P. Luo, “Multi-camera transfer gan for person re-identification,” *Journal of Visual Communication and Image Representation*, vol. 59, pp. 393–400, 2019.
- [76] Yu Fei, Q. Tang, W. Wang, and H. Wu, “A 2.7 ghz low-phase-noise lc-qvco using the gate-modulated coupling technique,” *Wireless Personal Communications*, vol. 86, no. 2, pp. 671–681, 2016.
- [77] M. Long and Y. Zeng, “Detecting iris liveness with batch normalized convolutional neural network,” *Computers, Materials & Continua*, vol. 58, no. 2, pp. 493–504, 2019.
- [78] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression,” in *Proceedings of the 2010 IEEE computer society conference on computer vision and pattern recognition-workshops*, pp. 94–101, IEEE.
- [79] L. Liu and M. T. Özsu, *Encyclopedia of Database Systems*, Springer, New York, NY, USA, 2009.
- [80] M. Daneshmand, A. Abels, and G. Anbarjafari, “Real-time, automatic digi-tailor mannequin robot adjustment based on human body classification through supervised learning,” *International Journal of Advanced Robotic Systems*, vol. 14, no. 3, Article ID 1729881417707169, 2017.
- [81] A. Bolotnikova, H. Demirel, and G. Anbarjafari, “Real-time ensemble based face recognition system for nao humanoids using local binary pattern,” *Analog Integrated Circuits and Signal Processing*, vol. 92, no. 3, pp. 467–475, 2017.
- [82] M. Valstar, B. Schuller, K. Smith et al., “Avec 2013: the continuous audio/visual emotion and depression recognition challenge,” in *Proceedings of the 3rd ACM International Workshop on Audio/visual Emotion challenge*, pp. 3–10, Barcelona Spain, October 2013.
- [83] C. Watson and P. Flanagan, “Nist special database 18 mugshot identification database,” 2016.
- [84] V. Bruce and A. Young, “Understanding face recognition,” *British Journal of Psychology*, vol. 77, no. 3, pp. 305–327, 1986.
- [85] G. Richard, Y. Mengay, I. Guis et al., “Multi modal verification for teleservices and security applications (m2vts),” in *Proceedings IEEE International Conference on Multimedia Computing and Systems*, pp. 1061–1064, IEEE, Florence, Italy, June 1999.
- [86] E. Paul, “Basic emotions,” *Handbook of cognition and emotion*, vol. 98, no. 45-60, p. 16, 1999.
- [87] R. Oppermann and R. Rasher, “Adaptability and adaptivity in learning systems,” *Knowledge transfer*, vol. 2, pp. 173–179, 1997.
- [88] E. Popescu, “Diagnosing students’ learning style in an educational hypermedia system,” in *Cognitive and Emotional Processes in Web-Based Education: Integrating Human Factors and Personalization*, pp. 187–208, IGI Global, Pennsylvania, United States, 2009.
- [89] D. Prakash, J. Van Haneghan, W. Blackwell, S. Jackson, G. Murugesan, and K. S. Tamilselvan, “Classroom engagement evaluation using computer vision techniques,” in *Pattern Recognition and Tracking XXX*, M. S. Alam, Ed., pp. 192–199, International Society for Optics and Photonics, SPIE, 1995.
- [90] J. Li, P. Wang, and Y. Xu, “Prognostic value of programmed cell death ligand 1 expression in patients with head and neck cancer: a systematic review and meta-analysis,” *PLoS One*, vol. 12, no. 6, pp. 1–16, 2017.
- [91] M. S. H. Aung, F. Alquaddoomi, C.-K. Hsieh et al., “Leveraging multi-modal sensing for mobile health: a case review in chronic pain,” *IEEE journal of selected topics in signal processing*, vol. 10, no. 5, pp. 962–974, 2016.
- [92] K. O’reagan, “Emotion and e-learning,” *Journal of Asynchronous Learning Networks*, vol. 7, no. 3, pp. 78–92, 2003.
- [93] N. Kemp and R. Grieve, “Face-to-face or face-to-screen? Undergraduates’ opinions and test performance in classroom vs. online learning,” *Frontiers in Psychology*, vol. 5, p. 1278, 2014.
- [94] Q. Xu, Bo Sun, J. He, B. Rong, L. Yu, and P. Rao, “Multi-modal facial expression recognition based on dempster-shafer theory fusion strategy,” in *Proceedings of the 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*, pp. 1–5, IEEE, Beijing, China, May 2018.
- [95] J. Yao, “Multilayer model for on-line learning resources based on cognitive load theory,” *World Trans. on Engng. and Technol. Educ.*, vol. 13, no. 3, pp. 245–250, 2015.
- [96] R. Joseph, S. Divvala, R. Girshick, and F. Ali, “You only look once: unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, May 2016.
- [97] X. Sun, J. Lichtenauer, M. Valstar, A. Nijholt, and M. Pantic, “A multimodal database for mimicry analysis,” in *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, pp. 367–376, Springer, Memphis, TN, USA, October 2011.
- [98] K. Sivaraman and A. Murthy, “Object recognition under lighting variations using pre-trained networks,” in *Proceedings of the 2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pp. 1–7, IEEE, Washington, DC, USA, October 2018.
- [99] J. Shao and Y. Qian, “Three convolutional neural network models for facial expression recognition in the wild,” *Neurocomputing*, vol. 355, pp. 82–92, 2019.
- [100] S. Stoyanov and P. Kirchner, “Expert concept mapping method for defining the characteristics of adaptive e-learning: alfanet project case,” *Educational Technology Research & Development*, vol. 52, no. 2, pp. 41–54, 2004.
- [101] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [102] D. Varga, “Multi-pooled inception features for no-reference image quality assessment,” *Applied Sciences*, vol. 10, no. 6, p. 2186, 2020.
- [103] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 60, no. 6, pp. 1097–1105, 2012.
- [104] D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, and S. Tubaro, “Deep convolutional neural networks for pedestrian detection,” *CoRR, abs/*, vol. 1510, p. 03608, 2015.



- [105] A. T. Lopes, E. de Aguiar, A. F. De Souza, T. Oliveira-Santos, and T. Oliveira-Santos, "Facial expression recognition with Convolutional Neural Networks: c," *Pattern Recognition*, vol. 61, no. C, pp. 610–628, 2017.
- [106] J. Cai, O. Chang, X. L. Tang, C. Xue, and C. Wei, "Facial expression recognition method based on sparse batch normalization cnn," in *Proceedings of the 2018 37th Chinese Control Conference (CCC)*, pp. 9608–9613, Wuhan, China, July 2018.
- [107] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion aware facial expression recognition using CNN with attention mechanism," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439–2450, 2019.
- [108] G. Yolcu, I. Oztel, S. Kazan et al., "Facial expression recognition for monitoring neurological disorders based on convolutional neural network," *Multimedia Tools and Applications*, vol. 78, no. 22, pp. 31581–31603, 2019.
- [109] A. Agrawal and N. Mittal, "Using cnn for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy," *The Visual Computer*, vol. 36, no. 2, pp. 405–412, 2020.
- [110] Deepak. Kumar Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognition Letters*, vol. 120, pp. 69–74, 2019.
- [111] D. Liang, H. Liang, Z. Yu, and Y. Zhang, "Deep convolutional bilstm fusion network for facial expression recognition," *The Visual Computer*, vol. 36, no. 3, pp. 499–508, 2020.
- [112] I.A. Essa, A. Pentland, and Darrell Trevor, "Tracking facial motion," in *Proceedings of the 1994 IEEE Workshop on Motion of Non-rigid and Articulated Objects*, pp. 36–42, IEEE, Austin, TX, USA, November 1994.
- [113] C. Wei, J. Cai, O. Chang, X. -L. Tang, and C. Xue, "Facial expression recognition method based on sparse batch normalization cnn," in *Proceedings of the 2018 37th Chinese Control Conference (CCC)*, pp. 9608–9613, Wuhan, China, July 2018.
- [114] Y. Yacoob and L. S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 636–642, 1996.
- [115] Anil. Audumbar Pise, Hima. Vadapalli, and I. Sanders, "Estimation of learning affects experienced by learners: an approach using relational reasoning and adaptive mapping," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–14, 2022.
- [116] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *International Journal of Computer Vision*, vol. 25, pp. 23–48, 1997.
- [117] H. J. Rosenblum, K. Pace-Savitsky, R. J. Perry, J. H. Kramer, B. L. Miller, and R. W. Levenson, "Recognition of emotion in the frontal and temporal variants of frontotemporal dementia," *Dementia and Geriatric Cognitive Disorders*, vol. 17, no. 4, pp. 277–281, 2004.
- [118] D. H. Kim, J. Wissam, J. Jang, and Y. Man Ro, "Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition," *IEEE Transactions on Affective Computing*, vol. 10, no. 2, pp. 223–236, 2017.
- [119] Z. Yu, G. Liu, Q. Liu, and J. Deng, "Spatio-temporal convolutional features with nested lstm for facial expression recognition," *Neurocomputing*, vol. 317, pp. 50–57, 2018.
- [120] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. O. Santos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610–628, 2017.
- [121] M. Ali, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Proceedings of the 2016 IEEE Winter conference on applications of computer vision (WACV)*, pp. 1–10, IEEE, Lake Placid, NY, USA, 7–10 March 2016.
- [122] M. Mohammadpour, H. Khaliliardali, S. M. R Hashemi, and M. M. AlyanNezhadi, "Facial emotion recognition using deep convolutional networks," in *Proceedings of the 2017 IEEE 4th international conference on knowledge-based engineering and innovation (KBEI)*, pp. 0017–0021, IEEE, Tehran, Iran, 22–22 Dec. 2017.
- [123] M. Pantic, L. J Rothkrantz, and M. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," *Proceedings of the the IEEE*, vol. 91, no. 9, pp. 1370–1390, 2003.
- [124] S. Zhang, S. Zhang, T. Huang, and W. Gao, "Multimodal deep convolutional neural network for audio-visual emotion recognition," in *Proceedings of the the 2016 ACM on International Conference on Multimedia Retrieval*, pp. 281–284, IEEE, Bari, Italy, October 2019.
- [125] F. Ringeval, F. Eyben, E. Kroupi et al., "Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data," *Pattern Recognition Letters*, vol. 66, pp. 22–30, 2015.