Facial Expressions Decoded: A Survey of Facial Emotion Recognition

Ahad Almasoudi
Faculty of Computing and IT
King Abdulaziz University
Jeddah 21589, Saudi Arabia

Souad Baowidan
Faculty of Computing and IT
King Abdulaziz University
Jeddah 21589, Saudi Arabia

Shahenda Sarhan
Faculty of Computing and IT
King Abdulaziz University
Jeddah 21589, Saudi Arabia

ABSTRACT

Facial emotion recognition is a process that involves detecting and interpreting emotions from facial expressions. This field draws from a range of disciplines, including computer science, psychology, and neuroscience. The ability to accurately recognize emotions from facial expressions has broad implications for human-computer interaction, healthcare, security, and marketing. This paper presents a thorough overview of the current state of the art in facial emotion recognition, covering topics such as facial expression theories, types of facial emotion recognition, datasets, techniques, evaluation metrics, and applications. Additionally, the paper addresses ethical considerations related to facial emotion recognition. The aim of this survey is to provide a comprehensive understanding of the present state of facial emotion recognition technology.

General Terms

Facial Emotion Recognition, Facial Expression

Keywords

Facial Emotion Recognition, Deep Learning, Computer Vision, FER, Affective computing

1. INTRODUCTION

"Faces are the mirrors of the soul"

Facial expressions and gestures can reveal our inner emotions, thoughts, and feelings. This is the foundation of facial emotion recognition (FER), a field of study that aims to develop algorithms and systems capable of automatically identifying and categorizing emotions displayed on human faces. FER is an active area of research that has attracted significant attention from the scientific community in recent years. It is a multidisciplinary field that combines knowledge from computer science, psychology, and neuroscience to detect and interpret emotions from facial expressions. The accurate detection of emotions from facial expressions has wide-ranging implications for fields such as human-computer interaction, healthcare, security, and marketing.

The main motivation for studying FER is the potential for its applications in various domains. For instance, in human-computer interaction, FER can enhance the user experience by enabling computers to respond to the user's emotional state. In healthcare, FER can detect and monitor patients' emotional states, aiding in the diagnosis and treatment of mental health conditions. In security and surveillance, FER can identify and track individuals based on their facial expressions. Lastly, in marketing, FER can analyze customers' emotional responses to products and advertisements.

This survey paper aims to provide a comprehensive overview of the state of the art in FER. The paper is divided into six sections, each focusing on a specific aspect of FER. Section 2 discusses various facial expression theories and provides an overview of different facial expression theories. In Section 3, the focus is on the different types of FER. Section 4 covers various datasets used for FER research. Section 5 discusses various techniques used for FER. In Section 6, various evaluation metrics used to measure the performance of FER systems are covered. Finally, in Section 7, the paper discusses various applications of FER and the ethical considerations associated with them. The paper concludes with a discussion of the future directions and open research challenges in the field of FER.

2. FACIAL EXPRESSION THEORY

The field of facial expression theory aims to comprehend the range of facial expressions that individuals use to communicate their emotions and other mental states. This includes both universal expressions that are recognized across cultures and culturally specific expressions.

The implications of facial expression theory extend to various fields, including psychology, sociology, anthropology, and computer science. For instance, a better comprehension of facial expressions can assist in the development of FER systems and enhance their ability to accurately interpret emotions.

In this section, an exploration of various theories and models proposed to explain facial expression and their role in human communication will be undertaken. Additionally, the limitations and criticisms of these theories will also be discussed.

2.1 Overview Of Different Facial Expression Theories

There are several theories that have been proposed to explain the nature and function of facial expressions, including:

—Paul Ekman's theory of six universal emotions [11]: According to this theory, there are six basic emotions that are universally recognized and expressed across cultures. These emotions include happiness, sadness, anger, fear, surprise, and disgust. Ekman argued that these emotions are biologically based and are signaled by specific facial expressions that are universal across cultures.

—Charles Darwin's theory of facial expressions [9]: In his book "The Expression of the Emotions in Man and Animals," Darwin proposed that facial expressions evolved as a way to communicate emotions and intentions. He argued that facial expressions are innate and universal, and that they serve important social functions such as attracting attention and signaling aggression or submission.

—The social function theory of facial expressions [32]: This theory suggests that facial expressions serve a social and communicative function, rather than being tied to specific emotions. According to this theory, people use facial expressions to convey meaning and intentions to others, and the specific expression used depends on the context and the intended message.

—The facial feedback hypothesis [6]: This theory proposes that facial expressions can influence and shape emotional experiences. According to this theory, the muscles in the face can feed back to the brain and influence emotional states, so that smiling can make us feel happier and frowning can make us feel sadder.

These are just a few examples of the various theories that have been proposed to explain the nature and function of facial expressions.

2.2 Limitations and Criticisms of These Theories

There are several limitations and criticisms of the theories of facial expression. Here are a few examples:

—Lack of cultural universality: Some researchers [15, 12, 40] have argued that facial expressions are not universal across cultures, and that what is considered to be a universal emotion in one culture may not be recognized as such in another culture. This suggests that the concept of universal emotions may not be accurate, and that facial expressions may be more culturally specific than previously thought.

—Limited focus on negative emotions: Many of the early theories of facial expression focused primarily on negative emotions [9], such as anger, fear, and disgust. This has led to a bias in research towards studying these emotions, and there has been less research on the expression of positive emotions such as happiness and love.

—Limited focus on facial expressions: Some researchers [14, 38, 33] have argued that facial expressions are just one part of the complex system of emotional communication, and that other nonverbal cues such as body language and vocal intonation also play important roles. This suggests that facial expression theory may be too narrow in its focus, and that a more comprehensive understanding of emotional communication requires considering a range of nonverbal cues.

—Lack of clear definitions: Some critics have argued that the definitions of facial expressions used in some theories are too broad or vague [5], making it difficult to test these theories and compare the results of different studies. This has led to a lack of consistency in the way that facial expressions are defined and measured, which has made it difficult to draw definitive conclusions about the nature and function of facial expressions.

Overall, while the theories of facial expression have contributed to our understanding of how emotions are communicated through

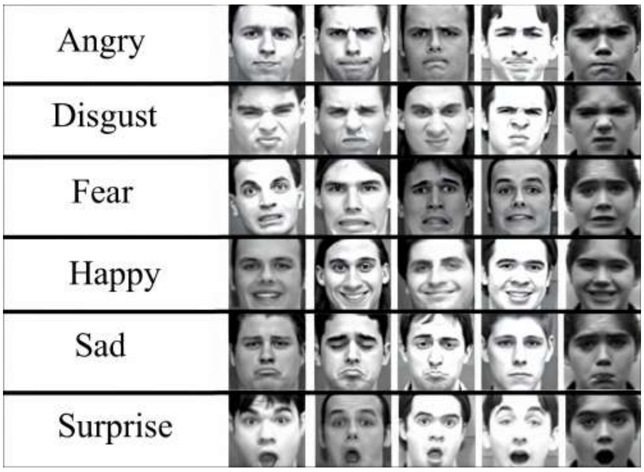

Fig. 1. Sample of Images Express Basic Emotions From Ck+ Dataset [29].

facial movements, there are limitations and criticisms of these theories that need to be considered when interpreting the findings of research on facial expression.

3. FACIAL EMOTION RECOGNITION TYPES

There are two main types of FER based on emotions' complexity: basic emotion recognition and compound emotion recognition. Basic emotion recognition focuses on identifying and interpreting primary or core emotions, such as happiness, sadness, anger, fear, surprise, and disgust [11]. On the other hand, compound emotion recognition involves identifying and interpreting more complex and nuanced emotions that are a combination of basic emotions [10]. These emotions may include contentment and excitement, among others. This section will explain the differences between these two types of FER and provide examples of how they are used in practical applications.

3.1 Basic Facial Emotion Recognition

In FER, basic emotions are considered to be a set of primary or core emotions that are universally recognized and have distinct physiological and behavioral expressions [11]. Examples of basic emotions include happiness, sadness, anger, fear, disgust, and surprise. These emotions are thought to be hardwired into the human brain and are easily recognizable based on universal facial expressions. See Figure 1 for sample images.

Basic FER systems are relatively simple to implement and are generally considered to be less complex and more cost-effective than other types of FER.

The study by Zadeh et al. [49] is significant because it proposes a deep learning-based approach for FER that combines Gabor filters and CNN. This method is an improvement over traditional FER methods because it achieves better accuracy and efficiency. The use of Gabor filters to extract features from the face images allows for a more detailed and accurate representation of the facial expressions. The subsequent use of CNN for classification enables the system to learn and identify complex patterns in the facial expressions. The use of the JAFFE dataset in the study is also noteworthy because it is a widely used benchmark dataset for FER research, and the results demonstrate the effectiveness of the proposed method in comparison to traditional FER methods. Overall, this study contributes to the development of more accurate and efficient FER systems

Table 1. Comparison of Some Techniques Based on Basic FER.

Ref.	Year	Technique	Dataset	Advantages	Disadvantages
[49]	2019	Gabor filters to extract features and CNN for classification	JAFFE	Improved efficiency and accuracy; Fast and accurate emotion recognition	Gabor filters need more computer power, can lose some information from the original image and they are harder to use and train
[23]	2020	Light Gabor convolutional network	FER2013, FER-Plus and RAF	Increases both the speed of training process and accuracy	Same previous
[1]	2021	Deep CNN with transfer learning	KDEF and JAFFE	High accuracy; Performed well on profile views in the KDEF dataset; Improved accuracy compared to training the model from scratch	Low accuracy when trained from scratch; Images with low resolution or imbalanced distribution may need additional pre-processing and modifications to the method

In 2020, Jiang et al. presented a Gabor Convolutional Network (GCN) for facial expression recognition that also utilizes Gabor filters [23]. The GCN is composed of six layers and extracts facial features from relevant areas using Gabor filters. The authors compared the performance of the GCN with other well-known CNNs and found that it was more accurate in recognizing expressions and required fewer computational resources. The GCN was tested on three facial expression recognition datasets: FER2013, FERPlus, and Real-world Affective Faces (RAF) databases. However, while Gabor filters can improve accuracy, they may also require more computational resources and lead to loss of information during the feature extraction process. This may make it challenging to implement and train for some applications.
In terms of analysis, the use of Gabor filters is a popular approach in facial expression recognition, as they can capture both high and low-frequency information in images. Both the Zadeh et al. [49] and Jiang et al. [23] studies demonstrated the effectiveness of Gabor filters in combination with CNNs for emotion recognition. However, as noted, the use of Gabor filters may also lead to some trade-offs, such as increased computational resources and information loss. Thus, it is important to consider the specific requirements and constraints of the application when deciding whether to use Gabor filters or other methods for feature extraction in FER.
The study in [1] used a combination of deep convolutional neural networks and transfer learning to recognize human facial emotions. The use of transfer learning involved taking a pre-trained deep CNN model and adapting it to facial emotion data by replacing the upper layers that are appropriate for FER. The method achieved high accuracy in recognizing facial emotions, with the best results being 96.51% and 99.52% on the KDEF and JAFFE datasets, respectively. The method also performed well on profile views in the KDEF dataset, highlighting its potential for real-life applications.
One limitation of the study is that training the model from scratch resulted in low accuracy compared to the transfer learning method, with test set accuracy being 23.35% and 37.82% for the KDEF and JAFFE datasets, respectively. This result suggests that transfer learning is a crucial step in achieving high accuracy in FER. Another limitation is that images with low resolution or an imbalanced distribution may require additional pre-processing and modifications to the method.
Overall, this study demonstrated the potential of transfer learning in improving the accuracy of FER and highlighted the importance of careful consideration of data pre-processing techniques in FER. Lastly, brief comparison is presented in Table 1 .

3.2 Compound Facial Emotion Recognition

Compound FER involves identifying and classifying emotions that are composed of multiple basic emotions. For instance, compound FER aims to identify emotions that are expressed as a combination of basic emotions, such as happiness and surprise (happiely surprised) [10].
Compound FER presents a greater challenge than basic emotion recognition, as it requires a deeper understanding of human emotions and how they are expressed. See Figure 2 to show sample images.
In 2018, Guo et al. presented three techniques for classifying FER [20]. The first technique used a Convolutional Neural Network (CNN) to recognize emotions through landmark displacement as a geometric representation. The second technique employed unsupervised learning and multiple Support Vector Machine (SVM) classifiers. The third technique combined a CNN inception-v3 with a center loss function.
The study used the iCV-MEFED dataset, and the proposed method showed several advantages over existing methods. For instance, the use of geometric representation of emotions resulted in improved performance compared to using texture-only information. Additionally, the technique utilized wider shallow networks, which achieved higher accuracy compared to deeper networks. The center loss function improved the model's ability to differentiate between similar samples, leading to an overall better performance.
However, the method also had some limitations. For example, it had difficulties in recognizing CFER and differentiating between dominant and complementary emotions. Furthermore, the second and third techniques had longer computational times compared to the first technique. The second technique utilized a shallow CNN, but the adoption of 50 classifiers added to the computational time. The third technique used the deeper inception-V3 structure, which required even more computational power.
Overall, this study demonstrated that the combination of geometric representation of emotions, wider shallow networks, and center loss

Fig. 2. Sample of Images Express Compound Emotions From RAF Dataset [28].

function can improve FER performance. However, further research is needed to address the limitations of this method, especially in recognizing compound emotions.

In a 2019 study by Khadija Slimani and colleagues [41], the Highway CNN technique was used on the CFEE (Compound Facial Expressions of Emotion) dataset to recognize facial expressions. One of the main advantages of this method is that it avoids the need for hand-designing effective features, which can be a problem with traditional machine learning methods. Moreover, the CFEE dataset included a diverse range of ethnicities and races, with representation from Caucasian, Asian, African-American, and Hispanic individuals. The use of highway layers in the CNN also provided a direct and indirect error correction strategy, which improved the overall performance of the network.

However, the CFEE dataset was relatively small for CNN-based methods, which made it more challenging to train the highway CNN architecture to recognize facial expressions accurately. Furthermore, it was difficult for the model to distinguish emotions from the basic categories that compose them. The study also found that the model had poor performance on some emotions, indicating that more work is needed to improve the technique's ability to recognize certain types of compound emotions.

Overall, the use of the Highway CNN technique shows promise in recognizing compound emotions and overcoming the limitations of traditional machine learning methods. However, more extensive datasets and better techniques to differentiate between basic emotions are needed to improve the performance of the technique.

The AHDCNN framework proposed by Thuseethan et al. in 2020 [44] is an interesting approach that leverages active learning for facial expression recognition. The use of active learning can greatly reduce the cost of labeling training data, which is often a time-consuming and expensive process, and can also improve the performance of the model. However, as noted, active learning also adds complexity to the model and may result in slower results. Furthermore, the lack of manual annotations during the training process may make it difficult to understand how the model arrived at its decisions. It is important to note that the study focused on the Compound Emotion Dataset, which is a relatively small dataset, and further research is needed to evaluate the performance of the AHDCNN on larger datasets. Overall, the AHDCNN framework presents an innovative approach to facial expression recognition that has the potential to improve performance and reduce the cost of training data labeling.

In a study by A. Swaminathan et al. in 2020 [43], an algorithm called FERCE was proposed for facial expression recognition using multiple datasets, including JAFFE, Mug face, CK, Yale face database, CEFEE and FERCE. The FERCE algorithm was used to derive combined emotions and a Support Vector Machine (SVM) was used to classify the emotions. The study found that emotions are asymmetric in nature by using the Facial Action Coding System (FACS) and statistical analysis of combined emotions. The proposed method performed well in classification except for two basic emotions and three combined emotions. However, the study also found that the system is less effective due to the limited number of data in most of the databases. Additionally, the system works effectively just with the images in which the faces of the people are front-posed, and the categorization of emotions can be imprecise and failed to clarify the converse emotions such as disgustedly angry and angrily disgusted.

In a study by Yuanlun Xie and others in 2020 [47], the technique of transfer learning was applied to the CFEE (Compound Facial Expressions of Emotion) dataset for facial expression recognition. The advantages of using transfer learning in this context include the fact that it requires less data compared to conventional deep learning methods, and that it can minimize training time while maintaining good accuracy, especially when the amount of data is limited. However, the dataset size is small and has a small number of categories which may cover just a small portion of all possible compound emotions, this can be a disadvantage of this approach.

In 2021, Kaminska et al. published a study that proposed a two-stage algorithm for facial expression recognition using the iCV-MEFED dataset [24]. This dataset has the advantage of covering a diverse range of ethnicity and hairstyles and being assessed by psychologists for expression realism. The two-stage approach addresses the issue of symmetrical label misclassification by leveraging both appearance and facial point information for compound emotion recognition. This method results in improved performance compared to using a single stage. However, the iCV-MEFED dataset is limited in size, and additional information beyond facial appearance features would be necessary to further improve performance.

Recently,In 2022, Pendhari et al. conducted a study that used the InceptionResNet-v2 architecture for recognizing compound facial expressions of emotion [37]. The use of the InceptionResNet-v2 architecture is a popular choice in computer vision tasks, as it has shown state-of-the-art performance in image classification and object detection tasks. Utilizing transfer learning with a pretrained network as a feature extractor can help to improve training performance with small datasets, which is particularly relevant for facial expression recognition since labeled data can be scarce and expensive to obtain. However, the limitations of the CFEE dataset may affect the generalizability of the results to real-world scenarios, where people may express emotions in different ways or with varying intensities. It would be interesting to see if the approach proposed by Pendhari et al. can be extended to larger and more diverse datasets to test its effectiveness in more realistic settings.

Finally, See Table 2 to read brief comparison between presented compound FER techniques.

4. DATASETS

The FER datasets comprise collections of images displaying various emotions on faces, and these images are employed to train machine learning algorithms to recognize emotions. These datasets offer significant value to researchers and developers focused on FER systems since they provide a large and varied set of instances that can instruct the algorithms in recognizing different emotions. This section presents some of the most popular datasets in the field, and Table 3 outlines the critical characteristics of these datasets, such as the number of images, the number of subjects, the number of emotions, and the data source.

4.1 CK+ dataset

The CK+ (Extended Cohn-Kanade) dataset [29] is a FER dataset that contains more than 500 images of faces showing eight basic emotions (angry, disgust, fear, happy, sad, surprise, contempt, and neutral). It is often used to evaluate the performance of FER systems.

The images in the CK+ dataset were collected from 123 subjects, and each image is annotated with one of the eight basic emotions. The dataset includes a wide range of expressions and poses, and it is considered to be one of the more challenging datasets for FER due to the subtlety of the expressions and the large intra-class variations.

4.2 iCV-MEFED dataset

The iCV-MEFED (iCV Multi-Emotion Facial Expression) dataset [19] is a large dataset that was developed for compound recognition purposes. It comprises 31,250 facial expressions that display various emotions, including the eight basic emotions (anger, contempt, disgust, fear, happiness, sadness, surprise, and neutral) as well as compound emotions, from 125 subjects with roughly equal gender distribution. Each subject displays 50 different emotions, and for each emotion, five samples were recorded. The images were supervised and labeled by psychologists, and the subjects were trained in acting out the emotions they were asked to display. The labels for the images use a complimentary-dominant format, with the primary emotion listed first and the secondary emotion listed second. For instance, 5_7 would indicate an expression of happily surprised.

4.3 FER-2013 dataset

The FER-2013 (Facial Expression Recognition 2013) dataset [18] is a collection of grayscale images of faces showing various emotional expressions with a size of 48x48 pixels. The dataset contains 35,887 images of faces, each labeled with one of seven emotional expressions: angry, disgusted, fearful, happy, neutral, sad, and surprised. This dataset was created to train and evaluate facial expression recognition systems. The images are divided into three sets: training (28,709 images), validation (3,589 images), and test (3,589 images).

4.4 RAF-DB dataset

The RAF-DB (Real-world Affective Faces) dataset [28] is a large collection of facial expression images that includes approximately 30,000 images sourced from the internet. These images are annotated by approximately 40 annotators and are diverse in terms of subject characteristics (age, gender, ethnicity, head pose, lighting, occlusions, etc.). The RAF-DB dataset provides a 7-dimensional expression vector for each image and includes two subsets: one with 7 classes of basic emotions and another with 12 classes of compound emotions. It also includes annotations such as landmark locations, bounding boxes, subject attributes, and classifier outputs for both basic and compound emotions. For objective performance evaluation, the dataset is split into a training set and a test set, with the training set being five times larger and having a similar expression distribution as the test set.

4.5 JAFFE dataset

The JAFFE (Japanese Female Facial Expression) dataset [30] is a collection of images of Japanese female faces displaying various facial expressions. The dataset was created by the Misaki Intelligent Systems Research Center in Japan and contains 213 images of 10 female Japanese models displaying each of the seven basic emotions: anger, disgust, fear, happiness, sadness, and surprise, as well as a neutral expression. Each image is labeled with the emotion that the model is expressing. The JAFFE dataset is widely used in research on facial expression recognition, especially for cross-cultural studies.

4.6 CFEE dataset

The CFEE (Compound Facial Expressions of Emotions) dataset [10] is a collection of standardized, non-profit photos of facial expressions from 230 individuals around the age of 23 from various ethnicities and races, including Caucasians, Asians, African-Americans, and Hispanics. The participants, including 130 women,

Table 2. Comparison of Some Techniques Based on Basic FER.

Ref.	Year	Technique	Dataset	Advantages	Disadvantages
[20]	2018	(1) CNN and it exploited landmark displacement as geometric (2) Adopted unsupervised learning with multiple SVM classifier (3) Combined CNN with a center loss function	iCV-MEFED	1st improved performance compared to using texture-only information; 2nd achieve higher accuracy compared to deeper networks with wider shallow networks; in 3rd center loss function improved differentiation between similar samples	Difficulties in recognizing emotions; 2nd and 3rd have longer computational time compared to the 1st
[41]	2019	Highway CNN	CFEE	Overcame the problem of hand-designing effective features; CFEE dataset included representation from multiple ethnicities and races; highway layers in the CNN provided a direct and indirect error correction strategy	Small dataset for CNN; difficult to train the highway CNN architecture to recognize expressions; poor performance on some emotions
[44]	2020	Active Hybrid Deep CNN with fusion mechanism	CFEE	Active learning improve performances; It capable to actively learn and enhance the classification capability	Added complexity from the use of active learning, slower to get results; Lack of understanding of how the model arrived at its decisions
[43]	2020	FERCE algorithm and using SVM	JAFFE, Mug face, CK, Yale face, CEFEE and FERCE	Prove that emotions are asymmetric in nature by employing the FACS and statistical analysis of combined emotions; Work performs well using the SVM classification	Limited number of data in most of the databases; less effective system; Works effectively just with front-posed images
[47]	2020	Transfer learning	CFEE	Requires less data compared to conventional deep learning methods; Can minimize training time while maintaining good accuracy	Small dataset size and a small number of categories that may cover just a small portion of all possible compound emotions
[24]	2021	2 stages: coarse recognition then fine recognition	iCV-MEFED	The dataset covers diverse ethnicity and hairstyles and assessed by psychologists; Addresses symmetrical label misclassification; Improves performance compared to single stage	Requires additional information beyond facial appearance features to improve performance
[37]	2022	InceptionResNet-v2 architecture	CFEE	Improved training performance with a small dataset; Reduced training time	Limited in size of CFEE dataset; Limited representation of all possible compound emotions

Table 3. Summary of Popular FER Datasets

Dataset	Type	No. of Images	No. of Subjects	No. of Emotions
CK+ [29]	Basic	over 500	123	8 basic emotions
iCV-MEFED [19]	Basic and Compound	31250	125	8 basic and 42 compound emotions
FER-2013 [18]	Basic	35,887	-	7 basic emotions
RAF-DB [28]	Basic and Compound	$\approx$ 30,000	-	7 basic and 12 compound emotions
JAFFE [30]	Basic	213	10	7 basic emotions
CFEE [10]	Compound	5060	230	15 compound emotions

were photographed without glasses or hair obstruction on their faces and were compensated for their contribution. The dataset contains a total of 5060 images, and special arrangements were made for those who needed corrective lenses.

5. FACIAL EMOTION RECOGNITION TECHNIQUES

FER techniques refer to algorithms or methods used to interpret and understand the emotional state of a person based on their facial expressions. This section reviews the various FER techniques that have been developed, discusses the strengths and limitations of each technique, and provides examples of their usage. Table 4 summarizes these strengths and limitations.

5.1 Rule-based techniques

Rule-based approaches to FER involve defining a set of rules or heuristics to identify emotions based on specific facial features or expressions. These approaches are typically simple to implement and can be effective when used with well-defined emotions, such as happiness or sadness. However, they are limited to identifying a small number of pre-defined emotions and can be sensitive to variations in facial expression.

An example of a rule-based approach is the ”facial action coding system” (FACS) [13], which involves defining a set of action units (AUs) corresponding to specific facial muscle movements. These AUs can be combined to create a wide range of facial expressions, which can be used to identify a variety of emotions. However, accurately detecting and coding AUs can require specialized hardware and may be sensitive to changes in lighting or other factors that affect the appearance of facial features.

One advantage of rule-based approaches is their relative simplicity of implementation and effectiveness when used with well-defined emotions such as happiness or sadness. However, they are limited in their ability to recognize a wide range of emotions and can be sensitive to variations in facial expression, such as a slight smile that may not be recognized as a smile.

Rule-based approaches are also limited in their ability to handle complex or nuanced emotions and may not be able to accurately recognize more subtle emotional states. In addition, they are typically limited to identifying a small number of pre-defined emotions, which can limit their applicability in real-world situations.

Many researchers have proposed FER systems that are rule-based and rely on the FACS system. For example, Gupta et al. [21] investigated the utilization of rule-based fuzzy systems for identifying human emotions from facial expressions. Additionally, a study by Das et al. [35] uses machine learning techniques to predict facial emotions based on FACS.

These studies suggest that while rule-based approaches to FER have limitations, such as their inability to handle complex or nuanced emotions, they can still be effective for identifying certain emotions with well-defined facial expressions. The use of machine learning techniques and fuzzy logic can also help to improve the accuracy of rule-based approaches by taking into account variations in facial expressions and other factors that can affect emotion recognition.

However, it is important to note that FACS itself has limitations, such as its reliance on static images and the need for specialized hardware for accurate detection and coding of AUs. Additionally, relying solely on rule-based approaches can limit the applicability of FER systems in real-world situations where emotions may be more complex and nuanced.

Therefore, a combination of rule-based and machine learning approaches, along with the use of diverse datasets and annotations, may be necessary to improve the accuracy and applicability of FER systems.

Overall, rule-based approaches can be useful for identifying simple emotions with well-defined facial expressions but may not be suitable for more complex or subtle emotions.

5.2 Feature-based techniques

In feature-based approaches to FER, specific features are extracted from an image or video of a face and used to classify the emotion being expressed. These approaches can handle a larger number of emotions than rule-based approaches and can be more robust to variations in facial expression.

Manual feature extraction, where specific facial features are identified and measured by hand, and automated feature extraction, where features are identified and extracted using machine learning algorithms, are two common ways to extract features in a feature-based approach.

The geometric feature-based (GFB) method is one example of a feature-based approach that involves extracting geometric features such as the distance between the eyes, the width of the mouth, and the angle of the eyebrows. Other feature-based approaches may use texture or color features, or a combination of multiple types of features. Arora et al. [4] and Alreshidi et al. [2] have used feature-based approaches in FER.

A major advantage of feature-based approaches is their ability to handle a larger number of emotions than rule-based approaches. They can also be more robust to variations in facial expression, as they do not rely on specific facial configurations to identify emotions. However, feature-based approaches require a large amount of labeled training data to be effective and can be sensitive to changes in lighting or other factors that affect the appearance of facial features.

Overall, feature-based approaches can be effective for classifying a wide range of emotions but may require a large amount of labeled training data and may be sensitive to changes in lighting or appearance.

Table 4. Strengths and Limitations of FER Techniques

Technique	Strengths	Limitations
Rule-based	Simple to implement; Can be effective for well-defined emotions	Limited to a small number of pre-defined emotions; Sensitive to variations in facial expression; Cannot handle complex or nuanced emotions
Feature-based	Can handle a larger number of emotions; More robust to variations in facial expression	Requires a large amount of labeled training data; Can be sensitive to changes in lighting or other appearance factors
Hybrid-based	Combines strengths of rule-based and feature-based techniques	May have limitations of both rule-based and feature-based techniques
Deep learning	Can handle a wide range of emotions; More robust to variations in facial expression; Can learn from large amounts of data and improve over time	Requires a large amount of labeled training data; Can be sensitive to changes in lighting or other appearance factors; Difficult to implement; Require a significant amount of computational resources

5.3 Hyprid-based techniques

Hybrid techniques combine the strengths of rule-based and feature-based techniques to improve the accuracy of emotion recognition. They often use a combination of hand-crafted features and machine learning algorithms to analyze facial expressions. For example, a hybrid system might first use a set of predefined rules to identify certain facial features and then use a machine learning algorithm to classify those features into different emotions. These techniques can improve the accuracy of emotion recognition by leveraging the strengths of both rule-based and feature-based techniques.

5.4 Deep learning techniques

Deep learning approaches to FER involve training a deep neural network to learn how to classify emotions based on labeled examples. These approaches are capable of handling a wide range of emotions and can be more robust to variations in facial expressions than rule-based or feature-based approaches. They are also able to learn from large amounts of data, allowing them to improve their performance over time.
However, deep learning approaches require a large amount of labeled training data to achieve high accuracy, and they can be sensitive to factors that affect the appearance of facial features, such as lighting or occlusions. Moreover, implementing and tuning these approaches can be challenging, as they require significant computational resources and expertise in deep learning techniques. For instance, recent studies like [8] and [25] demonstrate the use of deep learning approaches in FER.
In summary, deep learning approaches are a powerful tool for FER, but they require significant data and technical expertise to implement and may be sensitive to changes in facial appearance.

6. EVALUATION METRICS OF FACIAL EMOTION RECOGNITION SYSTEMS

Evaluation metrics of FER systems is an important step in the development process, as it allows researchers to measure the performance of their algorithms and compare them to other methods. The evaluation of FER systems is a challenging task due to the variability of human emotions and the diversity of datasets. Therefore, multiple evaluation metrics are typically used to provide a comprehensive view of the performance of the system.This section will discuss the different evaluation metrics used in FER research.
There are a different metrics and methods that are commonly used to evaluate the performance of FER systems. Some of the most common ones include:

—**Accuracy**: This is the most basic and straightforward metric, and it simply measures the percentage of emotions that are correctly classified by the system. It is typically calculated as:

$$Accuracy = \frac{Correct\ Predictions}{Total\ of\ Predictions}$$

—**Confusion Matrix**: A confusion matrix is a table that evaluates the performance of a classification algorithm [46]. In the case of facial expression recognition, the matrix displays the number of correctly and incorrectly predicted expressions by the system. The matrix comprises of four measures: True Positives (correctly predicted expressions), True Negatives (correctly not-predicted expressions), False Positives (incorrectly predicted expressions), and False Negatives (incorrectly not-predicted expressions). It is typically organized as follows:

Predicted	**Positive**	**Negative**
Positive	True Positive	False Positive
Negative	False Negative	True Negative

—**F1 Score**: An F1 score is a metric that combines precision and recall, and it is calculated as the harmonic mean of the two [46]. It is often used as a more balanced measure of performance, as it takes into account both the number of false positives and false negatives. It is typically calculated as:

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

—**Receiver Operating Characteristic (ROC) Curve**: A ROC curve is a commonly used graph for evaluating the performance of a binary classifier [17]. It displays the relationship between the true positive rate (TPR) and false positive rate (FPR) as the decision threshold is varied. The TPR represents the proportion of positive instances that are correctly classified as positive, while the FPR represents the proportion of negative instances that are incorrectly classified as positive. The AUC of the ROC curve is the area under the curve and provides a single number summarizing the classifier's performance. A higher AUC value indicates better performance.

—**Cross-Validation**: A Cross-validation is a technique used to evaluate the performance of a model by dividing the data into portions and training the model on one portion while evaluating it on the other portion [46]. This helps to reduce the impact of any specific data sample on the evaluation. A popular method of cross-validation is k-fold, where the data is split into k portions, and the model is trained and evaluated k times, each time using a different portion as the evaluation set while the rest of the portions serve as the training set.

7. FACIAL EMOTION RECOGNITION APPLICATIONS

FER is a rapidly growing field with a wide range of potential applications. Some of the key areas where FER technology is being used or has the potential to be used include:

—**Human-computer interaction**: FER technology can enhance the way humans interact with computers by making it more natural and intuitive. One application of FER in human-computer interaction (HCI) is through the use of intelligent user interfaces that can adapt to the user's emotional state. For instance, an intelligent system that incorporates FER technology can adjust the level of assistance it provides to the user based on their facial expressions. If the user appears frustrated, the system can offer more assistance, whereas if the user appears confident, the system may provide less assistance. Other areas where FER technology is used include assistive technology, education, customer service, and entertainment to create more natural and effective communication, to understand students' engagement, to understand customer's needs and emotions and to create immersive and engaging experiences respectively. Research papers such as [8, 31, 7] are a few examples of studies that have applied FER technology in HCI and there are many more studies being conducted in this field.

—**Education**: In the field of education, FER technology can be used to assess student engagement and attention, helping teachers to understand how their students are responding to their teaching.
Also, can be used in the classroom to monitor the students' emotional states in real-time, providing teachers with valuable insights into how students are responding to the material being taught. For example, if a student appears to be disengaged or frustrated, the teacher can adjust their teaching approach to better meet the needs of that student. [26, 45, 3] are examples of a papers that applied FER technology in education.

—**Health care**: In the field of health care, FER technology can be utilized to aid in the diagnosis and treatment of mental health conditions such as depression and anxiety. By analyzing patterns of emotion in patients, healthcare professionals can gain a deeper understanding of the patient's emotional state and provide more effective treatment.
FER technology can be applied during therapy sessions to evaluate a patient's facial expressions, providing valuable insights into their emotional state. This information can aid healthcare professionals in adjusting treatment plans to better meet the patient's needs.
Additionally, FER technology can be employed in research studies to gain a deeper understanding of the neural mechanisms underlying emotional processing in individuals with mental health disorders.
Examples of papers that have applied FER technology in mental health are [16, 27, 22, 39].

—**Marketing**: In the context of marketing, FER technology can be used to analyze the emotional reactions of consumers to advertisements, products, or branding. This information can then be used to improve marketing strategies and create more effective campaigns. For example, a company might use FER to determine which types of ads are most likely to elicit positive emotions in viewers, and then use that information to create more effective ads. Additionally, FER can be used to personalize marketing messages and content based on the emotions of individual consumers.
[42, 36, 34] are examples of a paper that applies FER in marketing.

There are many other potential applications of FER, and the technology is likely to be used in a wide range of fields in the future.

7.1 Ethical Considerations

There are several ethical considerations to take into account when using FER technology in different applications. Some of these include:

—**Privacy**: FER technology relies on the collection and analysis of personal data, which raises concerns about privacy and data protection.

—**Bias**: FER algorithms may be trained on biased data sets, which can lead to biased results [48]. This can have particularly negative impacts when used in security and surveillance or in systems that make decisions that affects people's lives.

—**Accuracy**: There is a risk that FER technology may be inaccurate or unreliable, which could lead to incorrect or misleading conclusions.

—**Fairness**: FER technology should be used in a way that is fair and non-discriminatory. It should not be used to unfairly target certain groups or individuals based on race, gender, sexual orientation, or other characteristics.

—**Misuse**: FER technology can be used for malign purpose such as profiling, surveillance, and discrimination.

It is crucial for researchers, developers, and users of FER technology to be aware of these ethical considerations and to take steps to address them. This may include conducting regular audits, testing for bias and fairness, being transparent about the use of the technology, and engaging in public dialogue about its implications.

8. CONCLUSION

In conclusion, FER is an active area of research that has gained significant attention from the scientific community in recent years. The accurate detection and interpretation of emotions from facial expressions have extensive implications for fields such as psychology, sociology, and computer science. The various applications of FER provide a strong motivation for studying this field.
This survey paper provides a comprehensive overview of the state of the art in FER by covering different aspects of the field including facial expression theories, types of FER, datasets, techniques, evaluation metrics, and applications. The advantages and disadvantages of different types and techniques have been discussed, and ethical considerations associated with FER have been highlighted.
Despite the progress made in this field, there are still many open research challenges that need to be addressed. For example, the development of robust and accurate FER systems that can operate in real-world environments is still an ongoing challenge. Additionally, the ethical implications of FER need to be carefully considered as the technology is increasingly being deployed in various domains.
Overall, FER is a rapidly evolving field with many exciting possibilities and ongoing challenges. Further research in this field has

the potential to lead to significant advancements in various domains and improve our understanding of human emotions.

9. REFERENCES

[1] MAH Akhand, Shuvendu Roy, Nazmul Siddique, Md Abdus Samad Kamal, and Tetsuya Shimamura. Facial emotion recognition using transfer learning in the deep cnn. *Electronics*, 10(9):1036, 2021.

[2] Abdulrahman Alreshidi and Mohib Ullah. Facial emotion recognition using hybrid features. In *Informatics*, volume 7, page 6. MDPI, 2020.

[3] Shadi AlZu'bi, Raed Abu Zitar, Bilal Hawashin, Samia Abu Shanab, Amjed Zraiqat, Ala Mughaid, Khaled H. Almotairi, and Laith Abualigah. A novel deep learning technique for detecting emotional impact in online education. *Electronics*, 11(18), 2022.

[4] Tarun Kumar Arora, Pavan Kumar Chaubey, Manju Shree Raman, Bhupendra Kumar, Yagnam Nagesh, PK Anjani, Hamed MS Ahmed, Arshad Hashmi, S Balamuralitharan, and Baru Debtera. Optimal facial feature based emotional recognition using deep learning algorithm. *Computational Intelligence and Neuroscience*, 2022, 2022.

[5] Lisa Feldman Barrett. *How emotions are made: The secret life of the brain*. Pan Macmillan, 2017.

[6] Ross Buck. Nonverbal behavior and the theory of emotion: the facial feedback hypothesis. *Journal of Personality and social Psychology*, 38(5):811, 1980.

[7] Prateek Chhikara, Prabhjot Singh, Rajkumar Tekchandani, Neeraj Kumar, and Mohsen Guizani. Federated learning meets human emotions: A decentralized framework for human–computer interaction for iot applications. *IEEE Internet of Things Journal*, 8(8):6949–6962, 2021.

[8] M Kalpana Chowdary, Tu N Nguyen, and D Jude Hemanth. Deep learning-based facial emotion recognition for human–computer interaction applications. *Neural Computing and Applications*, pages 1–18, 2021.

[9] Charles Darwin and Phillip Prodger. *The expression of the emotions in man and animals*. Oxford University Press, USA, 1998.

[10] Shichuan Du, Yong Tao, and Aleix M Martinez. Compound facial expressions of emotion. *Proceedings of the national academy of sciences*, 111(15):E1454–E1462, 2014.

[11] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[12] Paul Ekman and Wallace V Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971.

[13] Paul Ekman and Wallace V Friesen. Facial action coding system. *Environmental Psychology & Nonverbal Behavior*, 1978.

[14] Paul Ekman, Wallace V Friesen, and Phoebe Ellsworth. *Emotion in the Human Face: Guide-lines for Research and an Integration of Findings: Guidelines for Research and an Integration of Findings*. Pergamon, 1972.

[15] Paul Ekman, E Richard Sorenson, and Wallace V Friesen. Pan-cultural elements in facial displays of emotion. *Science*, 164(3875):86–88, 1969.

[16] Zixiang Fei, Erfu Yang, David Day-Uei Li, Stephen Butler, Winifred Ijomah, Xia Li, and Huiyu Zhou. Deep convolution network based emotion analysis towards mental health care. *Neurocomputing*, 388:212–227, 2020.

[17] Mingxing Gong. A novel performance measure for machine learning classification. *International Journal of Managing Information Technology (IJMIT) Vol*, 13, 2021.

[18] Ian J Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, et al. Challenges in representation learning: A report on three machine learning contests. In *International conference on neural information processing*, pages 117–124. Springer, 2013.

[19] Jianzhu Guo, Zhen Lei, Jun Wan, Egils Avots, Noushin Hajarolasvadi, Boris Knyazev, Artem Kuharenko, Julio C. Silveira Jacques Junior, Xavier Baró, Hasan Demirel, Sergio Escalera, Jüri Allik, and Gholamreza Anbarjafari. Dominant and complementary emotion recognition from still images of faces. *IEEE Access*, 6:26391–26403, 2018.

[20] Jianzhu Guo, Zhen Lei, Jun Wan, Egils Avots, Noushin Hajarolasvadi, Boris Knyazev, Artem Kuharenko, Julio C Silveira Jacques Junior, Xavier Baró, Hasan Demirel, et al. Dominant and complementary emotion recognition from still images of faces. *IEEE Access*, 6:26391–26403, 2018.

[21] Kanika Gupta, Megha Gupta, Jabez Christopher, and Vasan Arunachalam. Fuzzy system for facial emotion recognition. In *Intelligent Systems Design and Applications: 20th International Conference on Intelligent Systems Design and Applications (ISDA 2020) held December 12-15, 2020*, pages 536–552. Springer, 2021.

[22] Selen Işık Ulusoy, Şeref Abdurrahman Gülseren, Nermin Özkan, and Cüneyt Bilen. Facial emotion recognition deficits in patients with bipolar disorder and their healthy parents. *General Hospital Psychiatry*, 65:9–14, 2020.

[23] Ping Jiang, Bo Wan, Quan Wang, and Jiang Wu. Fast and efficient facial expression recognition using a gabor convolutional network. *IEEE Signal Processing Letters*, 27:1954–1958, 2020.

[24] Dorota Kamińska, Kadir Aktas, Davit Rizhinashvili, Danila Kuklyanov, Abdallah Hussein Sham, Sergio Escalera, Kamal Nasrollahi, Thomas B Moeslund, and Gholamreza Anbarjafari. Two-stage recognition and beyond for compound facial emotion recognition. *Electronics*, 10(22):2847, 2021.

[25] Asad Khattak, Muhammad Zubair Asghar, Mushtaq Ali, and Ulfat Batool. An efficient deep learning technique for facial emotion recognition. *Multimedia Tools and Applications*, 81(2):1649–1683, 2022.

[26] Imane Lasri, Anouar Riad Solh, and Mourad El Belkacemi. Facial emotion recognition of students using convolutional neural network. In *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)*, pages 1–6, 2019.

[27] Chong Li, Mingzhao Yang, Yongting Zhang, and Khin Wee Lai. An intelligent mental health identification method for college students: A mixed-method study. *International Journal of Environmental Research and Public Health*, 19(22), 2022.

[28] Shan Li, Weihong Deng, and JunPing Du. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2584–2593, 2017.

[29] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and

emotion-specified expression. In *2010 ieee computer society conference on computer vision and pattern recognition-workshops*, pages 94–101. IEEE, 2010.

[30] Michael Lyons, Miyuki Kamachi, and Jiro Gyoba. The Japanese Female Facial Expression (JAFFE) Dataset. April 1998.

[31] Ján Magyar, Gergely Magyar, and Peter Sincak. A cloud-based voting system for emotion recognition in human-computer interaction. In *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, pages 109–114, 2018.

[32] Abigail A Marsh, Reginald B Adams Jr, and Robert E Kleck. Why do fear and anger look the way they do? form and social function in facial expressions. *Personality and Social Psychology Bulletin*, 31(1):73–86, 2005.

[33] David Matsumoto. Culture and nonverbal behavior. *The SAGE handbook of nonverbal communication*, pages 219–235, 2006.

[34] Golam Morshed, Hamimah Ujir, and Irwandi Hipiny. Customer's spontaneous facial expression recognition. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(3):1436–1445, 2021.

[35] Madhuka Nadeeshani, Akash Jayaweera, and Pradeepa Samarasinghe. Facial emotion prediction through action units and deep learning. In *2020 2nd International Conference on Advancements in Computing (ICAC)*, volume 1, pages 293–298, 2020.

[36] Shaldon Wade Naidoo, Nalindren Naicker, Sulaiman Saleem Patel, and Prinavin Govender. Computer vision: The effectiveness of deep learning for emotion detection in marketing campaigns. *International Journal of Advanced Computer Science and Applications*, 13(5), 2022.

[37] Heenakausar Pendhari, Sushma Nagdeoti, Sandeep Rathod, Lubna Khan, and Saurabh Vishwakarma. Compound emotions: A mixed emotion detection. *Available at SSRN 4120265*, 2022.

[38] Robert Plutchik and Henry Kellerman. *Emotion, theory, research, and experience*, volume 3. Academic press, 1980.

[39] Lance M. Rappaport, Nicole Di Nardo, Melissa A. Brotman, Daniel S. Pine, Ellen Leibenluft, Roxann Roberson-Nay, and John M. Hettema. Pediatric anxiety associated with altered facial emotion recognition. *Journal of Anxiety Disorders*, 82:102432, 2021.

[40] James A Russell. Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies. *Psychological bulletin*, 115(1):102, 1994.

[41] Khadija Slimani, Khadija Lekdioui, Rochdi Messoussi, and Raja Touahni. Compound facial expression recognition based on highway cnn. In *Proceedings of the New Challenges in Data Sciences: Acts of the Second Conference of the Moroccan Classification Society*, pages 1–7, 2019.

[42] Iryna Spivak, Svitlana Krepych, Vasyl Faifura, and Serhii Spivak. Methods and tools of face recognition for the marketing decision making. In *2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T)*, pages 212–216, 2019.

[43] A Swaminathan, A Vadivel, and Michael Arock. Ferce: facial expression recognition for combined emotions using ferce algorithm. *IETE Journal of Research*, 68(5):3235–3250, 2022.

[44] Selvarajah Thuseethan, Sutharshan Rajasegarar, and John Yearwood. Complex emotion profiling: An incremental active learning based approach with sparse annotations. *IEEE Access*, 8:147711–147727, 2020.

[45] Güray Tonguç and Betul Ozaydın Ozkara. Automatic recognition of student emotions from facial expressions during a lecture. *Computers & Education*, 148:103797, 2020.

[46] Ni Wayan Surya Wardhani, Masithoh Yessi Rochayani, Atiek Iriany, Agus Dwi Sulistyono, and Prayudi Lestantyo. Cross-validation metrics for evaluating classification performance on imbalanced data. In *2019 International conference on computer, control, informatics and its applications (IC3INA)*, pages 14–18. IEEE, 2019.

[47] Yuanlun Xie, Wenhong Tian, and Tingsong Ma. A transfer learning approach to compound facial expression recognition. In *2020 4th International Conference on Advances in Image Processing*, pages 95–101, 2020.

[48] Tian Xu, Jennifer White, Sinan Kalkan, and Hatice Gunes. Investigating bias and fairness in facial expression recognition. In *European Conference on Computer Vision*, pages 506–523. Springer, 2020.

[49] Milad Mohammad Taghi Zadeh, Maryam Imani, and Babak Majidi. Fast facial emotion recognition using convolutional neural networks and gabor filters. In *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*, pages 577–581. IEEE, 2019.