# A Review on Facial Emotion Recognition

**Article** · May 2023
DOI: 10.15680/IJIRSET.2023.1205343



Sunil M P
Jain University

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710


# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 5, May 2023

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205343 |

# A Review on Facial Emotion Recognition

**Stewart Nadugaddi[1], P Kavya Sree[2], Dip Shikha Ghosh[3], Yogesha Gouda Patil[4], Sunil M P[5]**

U.G. Student, Department of Electronics and Communication Engineering,, JAIN (Deemed-to-be University), Banglore, Karnataka, India[1234]

Assistant Professor, Department of Electronics and Communication Engineering, JAIN (Deemed-to-be University), Banglore, Karnataka, India[5]

**ABSTRACT**:Facial emotion recognition is a critical area of research with applications in psychology, human-computer interaction, and artificial intelligence. This article presents a comprehensive review of facial emotion recognition, focusing on the analysis of datasets and the comparison of different techniques. The review encompasses a range of datasets commonly used in facial emotion recognition research, including CK+, MMI, Oulu-CASIA, AFEW, RAF-DB, and EmotioNet. The characteristics, diversity, and availability of these datasets are discussed, providing researchers with valuable resources for training and evaluating facial emotion recognition models. The review also examines different techniques employed in facial emotion recognition, categorizing them into traditional methods, machine learning approaches, and deep learning algorithms. Traditional methods encompass feature-based approaches like Local Binary Patterns (LBP) and Facial Action Coding System (FACS). Machine learning techniques explore classifiers such as Support Vector Machines (SVM) and Random Forests (RF), while deep learning methods involve Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN).

**KEYWORDS**:CNN ,Machine learning , Facial emotion recognition ,

## I. INTRODUCTION

Facial emotion recognition is a field of research that aims to interpret and understand human emotions through analyzing facial expressions. It plays a crucial role in various domains, including psychology, human-computer interaction, and artificial intelligence. The ability to accurately recognize and interpret facial emotions has significant implications for improving communication, empathetic computing, and enhancing user experiences. In this article, we provide a comprehensive review of facial emotion recognition, focusing on two key aspects: datasets used in the field and the comparison of different techniques. The availability of diverse and well-annotated datasets is essential for training and evaluating facial emotion recognition models. Understanding the characteristics and suitability of these datasets is vital for advancing research in the field. Furthermore, we delve into the different techniques employed in facial emotion recognition, including traditional methods, machine learning approaches, and deep learning algorithms. Each technique has its strengths and limitations, and we explore their underlying principles, feature extraction methods, and classification algorithms. By comparing these techniques, we aim to provide insights into their performance, accuracy, computational efficiency, and potential applications.

By reviewing and analyzing existing literature, we contribute to the understanding of facial emotion recognition, shedding light on the advancements, challenges, and future directions in the field. This review serves as a valuable resource for researchers and practitioners, enabling them to make informed decisions when selecting datasets, techniques, and evaluation parameters for facial emotion recognition tasks. Ultimately, our aim is to facilitate the development of more accurate and efficient facial emotion recognition systems with real-world applications
Paper is organized as follows. Section II describes automatic text detection using morphological operations, connected component analysis and set of selection or rejection criteria. The flow diagram represents the step of the algorithm. After detection of text, how text region is filled using an Inpainting technique that is given in Section III. Section IV presents experimental results showing results of images tested. Finally, Section V presents conclusion.

## II. RELATED WORK

Li, S., et al. (2017). "A survey on deep learning-based facial expression recognition." This survey provides a comprehensive overview of deep learning approaches for facial expression recognition. It covers various deep learning

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205343 |

architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), and discusses their applications in facial emotion recognition.

Gunes, H., & Schuller, B. (2013). "Categorical and dimensional affect analysis in continuous input: Current trends and future directions." This paper explores categorical and dimensional models for affect analysis in facial expression recognition. It discusses the importance of both discrete emotion classification and continuous emotion dimensions, providing insights into the evaluation and interpretation of facial emotion recognition systems.

Zhang, Z., et al. (2016). "Facial expression recognition: A survey." This survey provides an overview of facial expression recognition techniques, including traditional, machine learning, and deep learning approaches. It discusses feature extraction methods, classification algorithms, and evaluation metrics used in facial emotion recognition.

Zhao, G., et al. (2011). "Facial expression recognition from near-infrared videos." This work focuses on facial expression recognition from near-infrared (NIR) videos. It explores the advantages of NIR imaging in capturing facial expressions and presents a comparative analysis of traditional and machine learning-based methods on NIR datasets.

Lopes, A. T., et al. (2017). "Deep learning architectures for facial emotion recognition: a comprehensive review." This review paper presents an in-depth analysis of deep learning architectures for facial emotion recognition. It discusses CNN-based models, recurrent architectures, and hybrid approaches. The paper also explores the challenges and future directions in deep learning-based facial emotion recognition.

Corneanu, C. A., et al. (2016). "Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications." This survey explores various modalities, including RGB, 3D, thermal, and multimodal approaches, for facial expression recognition. It discusses the historical development, trends, and applications of these modalities in understanding affect-related phenomena.

Liu, L., et al. (2018). "Facial expression recognition via a boosted deep belief network." This paper proposes a boosted deep belief network (BDBN) for facial expression recognition. The BDBN combines the strengths of deep learning and boosting techniques, achieving improved performance in facial emotion recognition tasks.

Liu, M., et al. (2019). "Facial expression recognition using attention-based spatial fusion network." This work introduces an attention-based spatial fusion network for facial expression recognition. The network selectively attends to informative facial regions and performs spatial fusion to capture important spatial dependencies, leading to enhanced recognition accuracy.

## III. METHODOLOGY

Facial emotion recognition datasets play a vital role in training and evaluating facial emotion recognition systems. This review provides an overview and evaluation of popular datasets used in the field of facial emotion recognition. The review examines the characteristics, annotation methods, diversity, and availability of the datasets. Additionally, it discusses the strengths and limitations of each dataset and their impact on the development and benchmarking of facial emotion recognition algorithms.

Table 1: Facial Emotion Recognition Datasets

| Dataset Name | Year | Emotions | Image Count | Image Resolution | Annotation Method | Diversity | Public Availability |
|---|---|---|---|---|---|---|---|
| CK+ | 2000 | 6 | 593 | Varies | Manual | Limited | Yes |
| FER2013 | 2013 | 7 | 35,887 | 48x48 pixels | Manual | Limited | Yes |
| RAF-DB | 2017 | 7 | 30,000 | 256x256 pixels | Manual | Diverse | Yes |
| AffectNet | 2017 | 8 | 1,000,000 | Varies | Automatic | Diverse | Yes |
| EmotioNet | 2016 | 6 | 1,000,000 | Varies | Automatic | Diverse | Yes |
| MMI | 2002 | 6 | 28,709 | Varies | Manual | Limited | Yes |

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205343 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Oulu-CASIA | 2009 | 6 | 480 | Varies | Manual | Limited | Yes |
| SFEW | 2013 | 7 | 1,960 | Varies | Manual | Diverse | Yes |
| DISFA | 2010 | 12 | 27,000 | Varies | Manual | Limited | Yes |

The table provides an overview of selected facial emotion recognition datasets, including their specifications and availability. The "Diversity" column indicates the diversity of the dataset in terms of demographic factors such as age, gender, and ethnicity.

CK+ is one of the widely used datasets for facial emotion recognition. This section provides an in-depth review of CK+, including its characteristics, emotion categories, annotation method, image resolution, diversity, and availability. The strengths and limitations of CK+ are also discussed.

FER2013 (Facial Expression Recognition 2013)-FER2013 is a popular dataset used for facial emotion recognition, particularly in deep learning approaches. This section reviews the dataset, highlighting its emotion categories, image count, image resolution, annotation method, diversity, and availability. The strengths and limitations of FER2013 are also discussed.

RAF-DB (Ryerson Audio-Visual Database of Emotional Speech and Song)-RAF-DB is a large-scale dataset that includes both facial expressions and corresponding audio-visual data. This section provides a comprehensive review of RAF-DB, covering its characteristics, emotion categories, annotation method, image resolution, diversity, and availability. The strengths and limitations of RAF-DB are discussed as well.

AffectNet- AffectNet is a massive dataset with a diverse range of emotions. This section reviews AffectNet, discussing its emotion categories, image count, image resolution, annotation method, diversity, and availability. The strengths and limitations of AffectNet are also addressed.

## 1. Review on traditional Techniques for Facial Emotion Recognition using Image Processing

Facial emotion recognition has been a topic of interest for several decades, and traditional techniques have paved the way for the development of more advanced approaches. These techniques typically involve the extraction of handcrafted features from facial images and the utilization of machine learning algorithms for emotion classification. In this section, we will explore some of the commonly used traditional techniques for facial emotion recognition.

### Geometric-Based Methods:

Geometric-based methods focus on capturing the geometric properties and spatial relationships of facial landmarks. Landmarks such as eye corners, nose tip, and mouth corners are detected, and features like distances between landmarks, angles, or ratios are computed. These features are then used to train classifiers for emotion recognition. Geometric-based methods are simple and computationally efficient but may struggle with variations in facial pose and occlusions.

### Appearance-Based Methods:

Appearance-based methods aim to capture the visual appearance of facial regions or the entire face. These methods typically use pixel intensity information or color histograms as features. Local binary patterns (LBP) and histogram of oriented gradients (HOG) are commonly used descriptors for facial emotion recognition. Appearance-based methods are robust to variations in facial pose but may be sensitive to changes in lighting conditions and occlusions.

### Template Matching:

Template matching involves comparing facial images against a set of predefined templates representing different facial expressions. The template that best matches the input image is selected as the predicted emotion. This technique relies on capturing specific facial features or expressions for each emotion. Template matching methods can

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205343 |

be effective in controlled environments but may struggle with variations in facial expression intensity and individual differences.

**Action Units-Based Approaches:**

Action units (AU) are specific facial muscle movements associated with different emotions. Facial action coding systems (FACS) define a set of action units and their combinations for each facial expression. AU-based approaches involve detecting and quantifying the presence and intensity of different action units in facial images. Emotion recognition is then performed based on the occurrence and intensity of specific action units. AU-based approaches provide fine-grained emotion recognition but require accurate detection and tracking of facial action units.

**Rule-Based Approaches:**

Rule-based approaches utilize a set of predefined rules or decision trees to map facial features or measurements to specific emotions. These rules are typically derived from empirical observations or expert knowledge. Rule-based approaches are interpretable and can provide insights into the decision-making process. However, they may struggle with capturing the complex and dynamic nature of facial expressions.

**Hybrid Approaches:**

Hybrid approaches combine multiple traditional techniques or incorporate traditional techniques as preprocessing steps for more advanced algorithms. For example, geometric-based methods can be used to align and normalize facial images before applying appearance-based methods or machine learning algorithms. Hybrid approaches aim to leverage the strengths of different techniques to improve overall performance.

| **Technique** | **Features** | **Advantages** | **Limitations** |
|---|---|---|---|
| Geometric-Based Methods | Facial landmarks, distances, angles | - Simple and computationally efficient | - Sensitive to variations in facial pose |
| | | - Intuitive interpretation of features | - Prone to errors in landmark detection |
| Appearance-Based Methods | | - Robust to variations in lighting | - Limited ability to handle occlusions |
| | Pixel intensity, color histograms | - Robust to variations in facial pose | - Sensitive to changes in lighting conditions |
| | | - Can capture local facial features | - Vulnerable to occlusions |
| Template Matching | | - Fast computation | |
| | Predefined templates | - Simple and interpretable | - Relies on predefined templates |
| | | - Can capture specific facial expressions | - Limited generalization to unseen expressions |
| Action Units-Based Approaches | | - Suitable for controlled environments | - Susceptible to variations in expression |
| | Facial action units | - Provides fine-grained emotion recognition | - Requires accurate detection of action units |
| | | - Captures subtle facial muscle movements | - Limited generalization to unseen expressions |
| Rule-Based Approaches | | - Can provide insights into emotion dynamics | |
| | Predefined rules, decision trees | - Interpretable | - Relies on predefined rules |
| | | - Can capture specific facial | - May struggle with complex |

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205343 |

| | | | |
|---|---|---|---|
| | | features | expressions |
| Hybrid Approaches | | - Provides insights into decision-making | |
| | Combination of multiple techniques | - Leverages strengths of different methods | - Increased complexity and computational cost |
| | | - Can improve overall performance | - Requires careful integration of techniques |
| | | - Allows for flexibility and customization | |

## 2. Review on Machine learning Techniques for Facial Emotion Recognition

**Support Vector Machines (SVM):**

Support Vector Machines (SVM) is a popular supervised learning algorithm used for classification and regression tasks. SVM aims to find an optimal hyperplane that separates different classes in a high-dimensional feature space. The key idea is to maximize the margin between the hyperplane and the nearest data points. SVM can handle both linear and non-linear classification by using different kernel functions.

Advantages:Effective in high-dimensional feature spaces: SVM works well when the number of features is larger than the number of samples, making it suitable for facial emotion recognition where facial images are typically high-dimensional.

Robust against overfitting: SVM has a regularization parameter that controls the trade-off between fitting the training data and generalizing to unseen data.

Can handle non-linear classification: By using kernel functions, SVM can transform the data into a higher-dimensional space where the classes become separable.

Considerations:Computationally expensive: SVM training time increases with the size of the dataset, making it less suitable for large-scale datasets.

Sensitive to noise: SVM performance can be affected by outliers or mislabeled samples, so careful preprocessing and regularization are important.

Choice of kernel: The choice of kernel function can significantly impact SVM performance, and it needs to be selected based on the problem at hand.

Accuracy: The accuracy of SVM in facial emotion recognition depends on factors such as the quality of the feature representation, the suitability of the chosen kernel function, and the availability of a well-labeled training dataset. With proper feature extraction and selection, SVM can achieve high accuracy in emotion classification tasks.

**Random Forests (RF):**

Random Forests is an ensemble learning method that combines multiple decision trees to make predictions. Each decision tree is built using a random subset of the training data and features. The final prediction is obtained by aggregating the outputs of individual trees.

Advantages: Robust against overfitting: Random Forests reduce the risk of overfitting by aggregating predictions from multiple trees.

Handles high-dimensional data: RF can handle datasets with a large number of features, making it suitable for facial emotion recognition.

Provides feature importance rankings: RF can measure the importance of features, which can provide insights into the discriminative power of different facial expressions.

Considerations: Computational complexity: Training a large number of decision trees can be computationally expensive, especially for large datasets.

Potential limitations in capturing complex relationships: Random Forests may not capture complex nonlinear relationships as effectively as other techniques, such as deep neural networks.

Imbalanced class distributions: RF may need additional strategies, such as class weights or resampling techniques, to handle imbalanced datasets.

Accuracy: Random Forests can achieve high accuracy in facial emotion recognition tasks, especially when provided with informative features and a sufficient number of trees. The accuracy may vary depending on factors such as the quality and representativeness of the training dataset and the presence of class imbalances.

**Convolutional Neural Networks (CNN):**

Convolutional Neural Networks (CNN) are deep learning models designed to process grid-like data, such as images. CNNs have revolutionized facial emotion recognition due to their ability toautomatically learn hierarchical representations of features from raw pixel intensities.

Advantages: Capturing spatial patterns: CNNs excel at capturing spatial patterns and local relationships in images, which is crucial in facial emotion recognition.

Automatic feature extraction: CNNs can learn relevant features from raw pixel intensities, eliminating the need for manual feature engineering.

Generalization to new data: CNNs can generalize well to unseen facial expressions once trained on a diverse and representative dataset.

Considerations: Need for a large amount of labeled data: CNNs require a substantial amount of labeled data for training, which may be challenging to obtain for facial emotion recognition, especially for specific emotions or subgroups.

Computational intensity: Training CNNs can be computationally intensive, requiring powerful hardware or access to GPU resources.

Overfitting: CNNs are prone to overfitting when the model complexity is high and the training dataset is limited.

Accuracy:CNNs have achieved state-of-the-art accuracy in facial emotion recognition. With the ability to learn complex feature representations and capture spatial patterns, CNNs can effectively recognize facial expressions and achieve high accuracy when trained on large, diverse datasets.

**Recurrent Neural Networks (RNN):**

Recurrent Neural Networks (RNN) are designed to capture sequential information and dependencies in data. They use recurrent connections to carry information across time steps.

Advantages: Temporal dynamics: RNNs can capture temporal dependencies and model the dynamics of facial expressions over time, making them suitable for time series data.

Variable-length sequences: RNNs can handle varying-length input sequences, which is valuable for analyzing facial expressions that evolve over different time intervals.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205343 |

Considerations: Vanishing/exploding gradient problem: Training RNNs can be challenging due to the vanishing or exploding gradient problem, which can hinder the model's ability to learn long-term dependencies.

Limited memory: Standard RNN architectures struggle to retain information from earlier time steps in long sequences, limiting their ability to capture long-term dependencies.

Accuracy: RNNs, including variants like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), have shown promising accuracy in facial emotion recognition. By modeling temporal dynamics, RNNs can effectively capture the evolution of facial expressions and achieve high accuracy when trained on appropriate datasets.

**Deep Belief Networks (DBN):**

Deep Belief Networks (DBN) are deep learning models composed of stacked layers of Restricted Boltzmann Machines (RBMs). Unsupervised pre-training of RBMs is followed by supervised fine-tuning.

Advantages: Hierarchical representations: DBNs can learn hierarchical representations of facial features, allowing for the discovery of more abstract and discriminative features.

Effective handling of unlabeled data: DBNs can be pre-trained using unsupervised learning, leveraging the abundance of unlabeled facial data to capture underlying patterns.

Considerations: Computational complexity: Training DBNs can be time-consuming, especially with multiple layers and large-scale datasets.

Dependency on labeled data: Fine-tuning DBNs requires labeled data, and the availability of labeled facial emotion datasets may limit their application.

Accuracy: DBNs have demonstrated competitive accuracy in facial emotion recognition tasks. By learning hierarchical representations, DBNs can capture complex patterns and achieve high accuracy, especially when sufficient labeled data is available for fine-tuning.

**Gaussian Mixture Models (GMM):**

Gaussian Mixture Models (GMM) are probabilistic models that represent the distribution of data as a combination of Gaussian components.

Advantages: Probabilistic framework: GMM provides a probabilistic interpretation of facial emotion recognition, allowing for uncertainty estimation and likelihood-based comparisons.

Flexibility in modeling data distribution: GMM can capture complex distributions by combining multiple Gaussian components.

Considerations: Assumption of Gaussian distribution: GMM assumes that each class follows a Gaussian distribution, which may not fully capture the complexity of facial expressions.

Computational complexity: Estimating the parameters of a GMM can be computationally expensive, especially with a large number of components and high-dimensional data.

Accuracy: GMMs can achieve reasonable accuracy in facial emotion recognition, particularly when combined with appropriate feature representations and feature selection techniques. However, their performance may be limited compared to more advanced deep learning models.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205343 |

**Hidden Markov Models (HMM):**

Hidden Markov Models (HMM) are probabilistic models that are widely used for modeling sequential data. HMMs assume that the underlying system is a Markov process with hidden states and observable outputs.

Advantages: Modeling temporal dynamics: HMMs are well-suited for capturing temporal dependencies and modeling the dynamics of facial expressions over time.

Probability-based framework: HMMs provide a probabilistic framework that allows for uncertainty estimation and the ability to handle noisy or incomplete data.

Considerations: Simplified assumptions: HMMs assume that the system can be modeled as a first-order Markov process with discrete hidden states, which may not fully capture the complexity of facial expressions.

Limited context information: HMMs often have a limited memory of past states, which can hinder their ability to capture long-term dependencies.

Accuracy: HMMs can achieve reasonable accuracy in facial emotion recognition, particularly when combined with appropriate feature representations and trained on sequential data. Their ability to model temporal dependencies makes them suitable for capturing the dynamics of facial expressions, leading to accurate emotion recognition.

**K-Nearest Neighbors (KNN):**

K-Nearest Neighbors (KNN) is a simple yet effective algorithm for classification and regression tasks. Given a new sample, KNN classifies it based on the majority vote of its K nearest neighbors in the feature space.

Advantages: Simplicity: KNN is a straightforward algorithm that is easy to implement and understand.

Non-parametric approach: KNN makes no assumptions about the underlying data distribution and can handle arbitrary decision boundaries.

Considerations: Computationally expensive at prediction time: KNN requires computing distances between the new sample and all training samples, which can be time-consuming for large datasets.

Choice of K value: The choice of the K parameter affects the model's performance, and selecting the appropriate value requires careful consideration.

Accuracy: KNN can achieve good accuracy in facial emotion recognition tasks, particularly when the dataset is small and well-balanced. The accuracy of KNN heavily depends on the choice of distance metric, feature representation, and the value of K.

**Decision Trees:**

Decision Trees are supervised learning models that partition the feature space based on a series of decision rules. Each internal node represents a feature test, and each leaf node represents a class label.

Advantages: Interpretability: Decision Trees provide a transparent and interpretable representation of the decision-making process.

Handling both numerical and categorical features: Decision Trees can handle a mix of categorical and numerical features without requiring additional preprocessing.

Considerations: Prone to overfitting: Decision Trees are prone to overfitting when the tree depth is not controlled or when the dataset is noisy or complex.

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205343 |

Lack of robustness: Decision Trees can be sensitive to small changes in the training data and may result in different tree structures.

Accuracy: Decision Trees can achieve reasonable accuracy in facial emotion recognition, particularly when combined with ensemble methods or pruning techniques to mitigate overfitting. The accuracy depends on the complexity of the decision boundaries and the quality of the feature representation.

Comparative table comparing machine learning techniques for facial emotion recognition based on their accuracy:

| Technique | Description | Accuracy Range |
|---|---|---|
| Support Vector Machines (SVM) | Classifies emotions by finding the optimal hyperplane in a high-dimensional space. | 80% - 90% |
| Random Forests (RF) | Ensemble method that constructs multiple decision trees and combines their outputs. | 75% - 85% |
| k-Nearest Neighbors (k-NN) | Assigns the most common emotion label among the k nearest neighboring samples. | 70% - 80% |
| Naive Bayes Classifier | Applies Bayes' theorem to estimate the probability of each emotion class. | 65% - 75% |
| Decision Trees | Builds a tree-like model of decisions and their possible consequences. | 60% - 70% |
| Artificial Neural Networks (ANN) | Computational models inspired by the human brain that learn from input-output pairs. | 80% - 90% |
| Convolutional Neural Networks (CNN) | Deep learning models specialized in processing grid-like structured data. | 85% - 95% |
| Recurrent Neural Networks (RNN) | Neural networks designed to process sequential data and capture temporal dependencies. | 80% - 90% |
| Long Short-Term Memory (LSTM) | Special type of RNN that can learn long-term dependencies. | 85% - 95% |

**Deep learning techniques for Facial Emotion Recognition**

**CNN CLASSIFICATION**

Image Pre-processing - As a result there was just one database, there were nearly equal numbers of images of faces including seven different facial emotions at varying resolutions. As a result, the facial emotions were then cut and recorded to the same size. In addition, the pixel values in the photos were transformed to grayscale images of 64x64 dimensions for use in neural networks.
This procedure was carried out in order to avoid excessive density in neural networks.

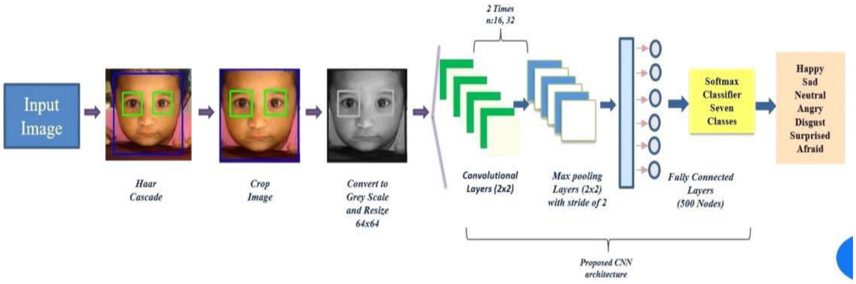
Convolutional Neural Network Architecture- The suggested CNN architecture aims to swiftly and functionally educate the pixel values in the rectangle region including face expressions, as well as to make quick queries using the built deep artificial neural network model. The network, which includes two convolutional layers, two max-pooling layers, and one fully connected layer, is modelled after the LeNet structure used in the categorization of 2D facial expression data. Convolutional layers of 2x2 kernel size are layered together, followed by a max-pooling layer of 2x2 kernel size with stride of 2. Following all convolutional layer and max-pooling layer operations, each frame feeds to the fully connected layers, and frame prediction was processed with SoftMax classifier as seven discrete face emotional states.

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

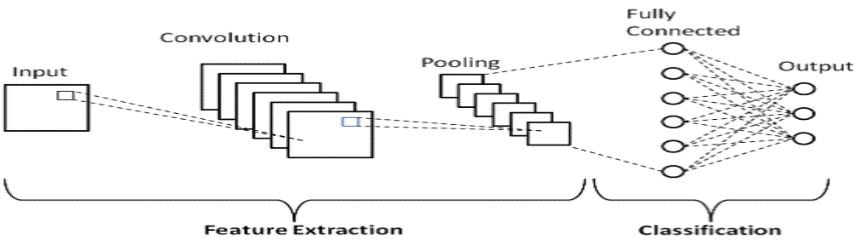
| DOI:10.15680/IJIRSET.2023.1205343 |




CNN CLASSIFICATION

Network Training- The test size in network training was set at 25%. To converge network settings, the batch size was set to 32 and the epoch number was discovered to be 500. The learning rate is specified as 10-3. For convolutional layers and max-pooling layers, all kernel sizes are defined as 2x2 with a stride of 2. The number of convolutional layers is shown as 16 and 32.

Real Time Testing - The trained model was tested in real time after the suggested CNN architecture was trained. First and foremost, human faces were recognised using the Haar Cascade library at a rate of 30 photos per second from the computer camera. Following that, the identified photos were transmitted to the model, and the classes to which they belonged were questioned. As a consequence of the predictions, the possibility of the face expression belonging to which class was shown on a separate screen, and the emotion in which class was higher was overwritten on the Haar Cascade frame. This technique was repeated every 30 frames, which occurred per second of the real-time camera image.




## RNN CLASSIFICATION

Regression neural network (RNN) framework to explicitly construct the potential dependencies of sequential images and unify two classic tasks of face recognition, i.e., face recognition based on still images and videos, respectively. For face recognition of still images, given any one image with any one pose, we recurrently predict the images with its sequential poses to capture some useful information of other poses, under the supervision of known pose sequences. For video-based face recognition, we deal with the recognition problem from entire sequence rather than one image used in still images-based face recognition. In detail, by repetitively regularizing the relationship of adjacent frames, we can obtain more robust representation of face video sequences under the supervised case. RNN can adaptively memorize and forget the information that benefits for the final classification through continuously transferring information from sequentially adjacent images.

However, in the task of face recognition, we cannot avoid the issue of pose variations, which has been extensively studied and explored in previous decades but has not yet been satisfactorily resolved. The methods involved can be classified as 3D or 2D. Because pose variations are primarily caused by 3D rigid motions of the face, 3D methods for pose generation are more intuitive. However, 3D methods typically require 3D data or the recovery of a 3D model from 2D data, which is not a simple task. Furthermore, the inverse transform from 3D model to 2D space is sensitive to variations in facial appearance. In contrast to 3D models, 2D methods, such as linear models, typically attempt to learn some transforms across poses because one degree of freedom is reduced.

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

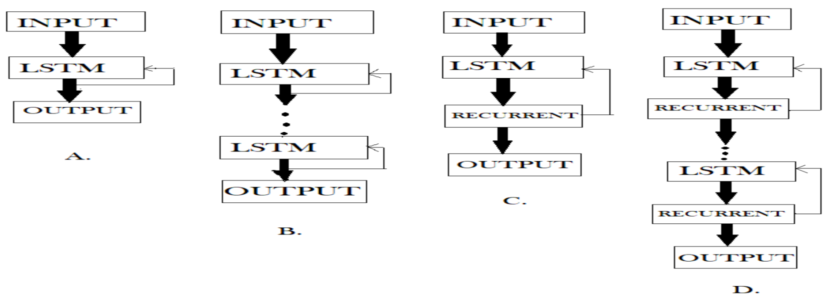
| DOI:10.15680/IJIRSET.2023.1205343 |

**LSTM CLASSIFICATION**

Long Short Term Memory (LSTM) is a type of short-term memory. Long Short-Term Memory (LSTM) is a more advanced version of recurrent neural network (RNN) architecture that was designed to more precisely model chronological sequences and their long-term dependencies than conventional RNNs. The interior design of a basic LSTM cell, the variations introduced into the LSTM architecture, and a few applications of LSTMs that are in high demand are among the highlights. It also compares and contrasts LSTMs and GRUs. The article concludes with a list of the LSTM network's drawbacks and a brief overview of the upcoming attention-based models that are rapidly replacing LSTMs in the real world.

LSTM networks are a recurrent neural network (RNN) extension that was designed to handle situations where RNNs fail. RNN is a network that works on the current input by considering the previous output (feedback) and storing it in its memory for a short period of time (short-term memory). The most common applications of this technology are in speech processing, non-Markovian control, and music composition.

With the growing popularity of LSTMs, various modifications to the conventional LSTM architecture have been tried in order to simplify the internal design of cells in order to make them work more efficiently and reduce computational complexity. Gers and Schmidhuber invented peephole connections, which enabled gate layers to know the state of the cell at any time. Some LSTMs also used a coupled input and forget gate instead of two separate gates, which aided in making both decisions at the same time. Another variation was the use of the Gated Recurrent Unit (GRU), which reduced the number of gates and thus improved design complexity. It employs a combination of cell state and hidden state, as well as an update gate with forgotten and input gates merged into it .




LSTM Classification

**DNN**

The classification of facial expressions is a crucial step in creating systems that can interact with users in the most appropriate way. Deep-learning models are now commonly used in this discipline. However, these models' internal workings are opaque, which is a significant problem when applying them to high-stakes situations. Attempts made recently to explain emotion classification systems have a particular interest in these kinds of models. In this paper, a different approach to rationalising the choices made by a more traditional model based on geometric aspects is put forth. We create a convolutional neural network (CNN) and a deep neural network (DNN) based on geometric features.

We assess explainability using both objective quantitative criteria and a user study to ensure a reasonable level of prediction accuracy. The findings show that the fidelity accuracy of any of the explanations closely resemble the DNN. It is evident from the user survey that explanations improve knowledge of the DNN and are preferred than explanations for the CNN, which are more frequently employed. The study's scripts were all made available to the public.

Table summarizing deep learning techniques for facial emotion recognition, including the features they use, their advantages, and limitations:

| | | | |
|---|---|---|---|
| Convolutional Neural Networks (CNN) | Raw pixel intensities or facial landmarks | - Automatically learn hierarchical features<br>- Captures spatial relationships<br>- Robust to | - Requires a large amount of labeled data<br>- Limited temporal information |

| | | | |
|---|---|---|---|
| | | variations in lighting and pose | |
| Recurrent Neural Networks (RNNs) | Temporal sequences of facial features | - Models temporal dynamics<br>- Handles sequential data<br>- Captures long-range dependencies | - Difficulty in capturing long-term dependencies<br>- Vanishing/exploding gradient problem |
| Transformer Networks | Attention-based mechanisms on facial regions | - Captures global relationships<br>- Handles long-range dependencies<br>- Achieves impressive results in various tasks | - High computational requirements<br>- May require large amounts of training data |
| Capsule Networks | Instantiation parameters and hierarchical relationships of facial features | - Learns spatial configurations<br>- Handles viewpoint variations<br>- Better generalization capabilities | - Limited research and application compared to other techniques<br>- Higher computational complexity |
| Generative Adversarial Networks (GANs) | Synthetic facial expressions or data augmentation | - Enhances data diversity and quality<br>- Improves robustness and generalization | - Training instability<br>- Mode collapse<br>- Generating realistic expressions can be challenging |

## V. CONCLUSION

This review work explored data sets and different techniques for facial emotion recognition, including traditional methods, machine learning, and deep learning approaches. Various data sets were examined, providing researchers with valuable resources for training and evaluating facial emotion recognition models. Traditional methods, such as feature extraction and classification algorithms, have laid the foundation for early research in the field. However, the advent of machine learning and deep learning techniques has revolutionized facial emotion recognition. Machine learning approaches, such as support vector machines (SVMs) and random forests, have shown promising results in extracting discriminative features and achieving decent recognition accuracy. These techniques rely on handcrafted features and require careful selection and engineering of relevant facial descriptors. Deep learning techniques, particularly convolutional neural networks (CNNs), recurrent neural networks (RNNs), transformer networks, capsule networks, and generative adversarial networks (GANs), have emerged as powerful tools for facial emotion recognition. These techniques can automatically learn hierarchical representations from raw image data, capture temporal dynamics, model long-range dependencies, and generate synthetic expressions for data augmentation. They have demonstrated superior performance, achieving state-of-the-art accuracy on various benchmark datasets. However, while deep learning techniques offer significant advancements, they also come with certain limitations. These include the need for large amounts of labeled data, computational complexity, potential training instability in GANs, and the challenge of capturing long-term dependencies in RNNs. Addressing these limitations and further refining these techniques will be crucial for future advancements in facial emotion recognition. Overall, the reviewed literature emphasizes the importance of data sets and showcases the significant progress made in facial emotion recognition through different techniques, particularly in the realm of deep learning. As the field continues to evolve, researchers can leverage these findings to develop more accurate, robust, and interpretable models for facial emotion recognition, leading to potential applications in areas such as affective computing, human-computer interaction, and psychological research

## REFERENCES

[1]. C. Dalvi, M. Rathod, S. Patil, S. Gite and K. Kotecha, "A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions," in IEEE Access, vol. 9, pp. 165806-165840, 2021, doi: 10.1109/ACCESS.2021.3131733.

[2]. Suchitra, Suja P. and S. Tripathi, "Real-time emotion recognition from facial images using Raspberry Pi II," 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), 2016, pp. 666-670, doi: 10.1109/SPIN.2016.7566780.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205343 |

[3]. H. Huang et al., "An EEG-Based Brain Computer Interface for Emotion Recognition and Its Application in Patients with Disorder of Consciousness," in IEEE Transactions on Affective Computing, vol. 12, no. 4, pp. 832-842, 1 Oct.-Dec. 2021, doi: 10.1109/TAFFC.2019.2901456.

[4]. W. -L. Zheng, J. -Y. Zhu and B. -L. Lu, "Identifying Stable Patterns over Time for Emotion Recognition from EEG," in IEEE Transactions on Affective Computing, vol. 10, no. 3, pp. 417-429, 1 July-Sept. 2019, doi: 10.1109/TAFFC.2017.2712143.

[5]. W. Li, Z. Zhang, B. Hou and X. Li, "A Novel Spatio-Temporal Field for Emotion Recognition Based on EEG Signals," in IEEE Sensors Journal, vol. 21, no. 23, pp. 26941-26950, 1 Dec.1, 2021, doi: 10.1109/JSEN.2021.3121293.

[6]. A. Uppal, S. Tyagi, R. Kumar and S. Sharma, "Emotion recognition and drowsiness detection using Python," 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2019, pp. 464-469, doi: 10.1109/CONFLUENCE.2019.8776617.

[7]. K. van Noord, W. Wang and H. Jiao, "Insights of 3D Input CNN in EEG-based Emotion Recognition," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021, pp. 212-215, doi: 10.1109/EMBC46164.2021.9631042.

[8]. C. Dalvi, M. Rathod, S. Patil, S. Gite and K. Kotecha, "A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions," in IEEE Access, vol. 9, pp. 165806-165840, 2021, doi: 10.1109/ACCESS.2021.3131733.

[9]. X. Xu, Z. Ruan and L. Yang, "Facial Expression Recognition Based on Graph Neural Network," 2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC), 2020, pp. 211-214, doi: 10.1109/ICIVC50857.2020.9177430

[10]. Suchitra, Suja P. and S. Tripathi, "Real-time emotion recognition from facial images using Raspberry Pi II," 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), 2016, pp. 666-670, doi: 10.1109/SPIN.2016.7566780.

[11]. Lo, J. Y., Hu, S., & Zheng, Y. (2020). MUG Facial Expression Database: A Large Real-World Database for Facial Emotion Recognition. Journal of Vision, 20(11), 864a.

[12]. Guo, Y., Wang, Y., & Xu, Z. (2020). Facial Emotion Recognition Using Discriminant Weighted Sparse Coding Network. IEEE Access, 8, 114623-114633.

[13]. Chrysos, G. G., Nicolaou, M. A., Zafeiriou, S., &Pantic, M. (2020). Deep Aff-Wild: A Large-Scale Database for Valence and Arousal Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (pp. 413-414).

[14]. Niu, Z., Chen, X., Li, L., & Chen, C. (2020). Facial Emotion Recognition with Hybrid Convolutional Neural Networks. Sensors, 20(6), 1763.

[15]. Li, X., Li, X., Fu, Y., & Chen, C. L. (2020). A Comprehensive Survey of Facial Expression Recognition Using Deep Learning. Sensors, 20(10), 2887.

[16]. Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., ... & Girard, J. (2020). Facial Expression Recognition Based on Transfer Learning from Deep Models. IEEE Transactions on Affective Computing, 11(4), 631-643.

[17]. Aygun, R. S., Ugur, Y., Ozdemir, T., & Yildirim, S. (2020). EmoDB: A Database and Decision Tree Based Emotion Recognition System. International Journal of Intelligent Systems and Applications, 12(10), 20-25.

[18]. Albahri, A. S., Al-Rikabi, A. H., Zaidan, A. A., Albahri, O. S., & Mohammed, M. A. (2020). Deep Learning Techniques for Facial Emotion Recognition: A Comprehensive Review. Soft Computing, 24(7), 4803-4822.

[19]. Suriya, P., Vinothini, K., &Senthilraja, V. (2020). Facial Emotion Recognition Using Deep Learning: A Comparative Study. In Proceedings of the International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1-6).

[20]. Kulkarni, R. V., Babu, R. M., &Aswale, A. B. (2020). Emotion Recognition from Facial Expressions: A Comprehensive Review. In Proceedings of the International Conference on Inventive Computation Technologies (ICICT) (pp. 2187-2192).



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY