# A SURVEY ON FACIAL EMOTION DETECTION SYSTEM USING CONVOLUTION NEURAL NETWORKS

**[1-4]Shalu Mishra, Shubhee Tiwari, Prachi Dubey, Oshi Dubey**
**Institute of Technology and Management GIDA, Gorakhpur**

**ABSTRACT**

Face detection system is a computer programme that can track, recognise, or categorise human faces from an image or video taken with a camera.Facial expression detection is a topic of great interest in most fields from AI. It is the method of how we differentiate human face using algorithm. This method is generally used for identification and security purpose. Facial expression recognition software aims to recognize basic human emotions using different algorithms like OPENCV, KERAS, CNN implemented by using Python. It involves two phases: Face detection and Face recognition. Here the software takes real-time input, capture it and classify it according to various emotions and produces output. Expression is key to understanding human thoughts, actions, and emotions. The manually compiled image datasets are used to train and evaluate the model.

**KEYWORDS**

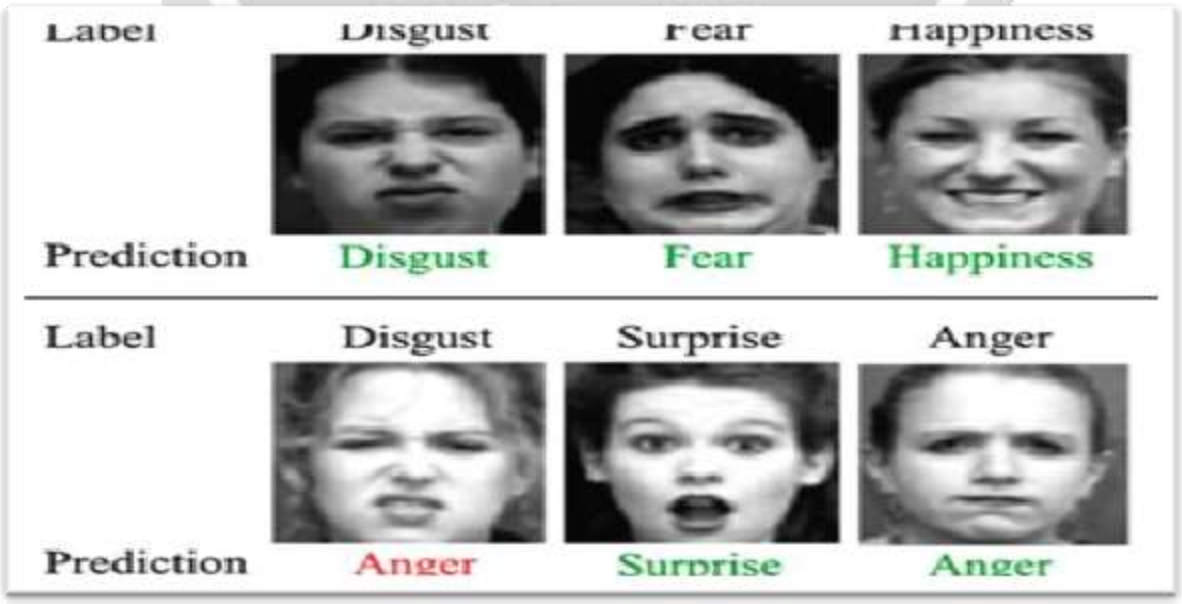
Facial Recognition- CNN, Deep Learning, Pooling, Soft max Activation, Scope, Artificial intelligence, ANN, RNN, Eigenfaces

---

**INTRODUCTION:**

Speaking, gesturing, and expressing emotions are all ways that people communicate with one another. If a machine is able to comprehend human emotion, it will be much easier for it to communicate with people naturally with the aid of AI. It operates by recognizing and quantifying facial features in images. It's important to understand how a user feels over time in a real-world setting [1].

Humans differentiate and recognize faces based on a variety of characteristics, including the placement, size, and shape of facial features including the nose, ears, lips, and eyes. While identifying utilizing different algorithms.Facial emotion recognition offers a variety of real-time applications. By evaluating the expression at different points in time, it is possible to create emotion-aware games in the gaming business.




**ARTIFICAL INTELLIGENCE:**

The replication of human intelligence by machines, particularly computer systems, is known as artificial intelligence. Expert systems, machine learning, speech recognition, natural language processing, and vision are a few examples of specific AI applications.

**HISTORY OF NEURAL NETWORK:**

An artificial intelligence technique known as a neural network teaches computers to interpret data in a manner that is inspired by the human brain. In order to simulate a layered architecture of interconnected neurons or nodes is used in the human brain by a machine learning technique known as deep learning. [10].

Frank Rosenblatt, a psychologist, created the first artificial neural network in 1958. Its name, perceptron, was chosen to represent how the human brain interpreted visual information and developed object recognition skills. Similar ANNs have since been employed to examine human cognition by other researchers. Artificial neural networks (or ANNs) are a type of neural network technology that aims to mimic the functions of the human brain in a computer.

It is created with the ultimate goal of its development is to enable a computer to learn from experience much like humans do.

**Neural network types:**

**Artificial Neural Network(ANN)**: Multiple perceptron or neurons are grouped together to form an artificial neural network (ANN) at each layer. Because all inputs are processed in the forward direction, ANN is frequently referred to as a feedforward neural network. These kinds of neural networks are among the most fundamental. Information is transmitted in a single direction, travelling via a number of input nodes before it is received at the output node. The network may or may not have hidden node levels, which helps to explain how the network works.

Disadvantages:

1. hardware dependence.
2. Network behaviour that is not understandable
3. Establishing the appropriate network structure.
4. Tabular and text-based data types are available.
5. ANN is thought to be less effective than CNN and RNN.
6. Hardware dependence and unexpected network behaviour.

**Recurrent Neural Network(RNN)**:The complexity of recurrent neural networks is higher. In order to teach the model how to predict the outcome of a layer, the output of processing nodes is preserved and handed to it later (they did not transfer the information in one direction only). The RNN model's nodes function as neural cells, carrying out computation and implementation tasks at each node. if the networks' predictions are incorrect. After that, during the back propagation, the system continues to learn for itself and try for the correct estimates.

Disadvantages:

1. Issues with gradient fading and explosions are a drawback.
2. An RNN's training process could be very challenging.
3. Tanh or relu cannot parse very long sequences when employed as the activation function.
4.Data that is in sequence.
5.In comparison to CNN, RNN has fewer feature compatibility.
6.Gradient disappearing.

**Convolution Neural Networks (CNN)**: CNNs are among the most widely-used models currently in use. This computational model of a multilayer perceptron version is used by a neural network, and it contains one or more convolutional layers that are either fully coupled or pooled. As a result of these convolutional layers, features maps are produced, which take a portion of the image and send it for nonlinear processing after being separated into rectangles.

Advantages:

1. Extremely high accuracy in issues involving image recognition.
2. Detects key properties automatically and without human intervention.
3. Weight distribution
4. Compared to ANN and RNN, CNN is considered to be more strong.
5. Image-based data.

The first sincere attempts to create a face were made in the 1980s and 1990s using a technique known as Eigenfaces. A face recognizer assumes that each face is composed of several Eigenfaces images that have been layered one pixel at a time on top of each other in order create a blur image that resembles a face. However, this approach didn't really succeed. The following generation of face recognizers would then take each image of a face and identify significant features like a mouth corner or an eyebrows. These points' coordinates are referred to as feature points. Although this method is superior to the Eigenfaces method, it is still not perfect. We waste a great deal of information that may be useful: Hair colour, eye colour, and other face structures that a feature point does not capture etc. All of this is ignored by the current generation of face recognition software. Convolutional neural networks were utilised in this technique (CNNs).

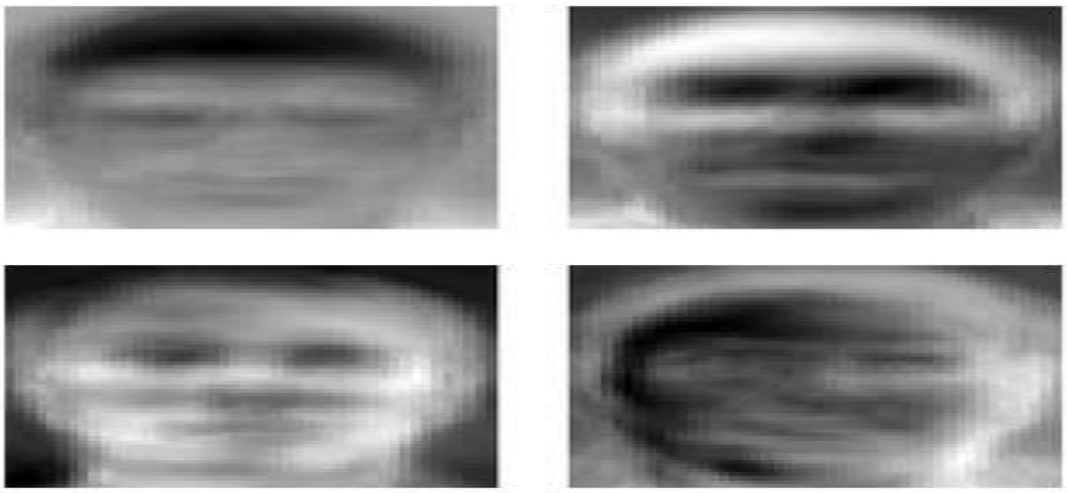


Fig 1: Some Eigenfaces

An artificial neural network commonly utilised in deep learning for object/image recognition and categorization is the convolutional neural network. [6]. Convolutional neural networks were first called Le Net after Yann Le Cun's conception of them. The majority of its development took place between 1989 and 1988 for the purpose of handwritten digit recognition. A convolutional algorithm is a type of deep learning that can take in an input image, evaluate various components and objects within it, and distinguish between them. CNNs are the preferred specification for identifying and recognising objects, despite the fact that there are other forms of neural networks for deep learning.The four different CNN layers the convolutional, pooling, totally linked, Re LU correction, and other layers are examples. CNN makes use of the geographic correlations in the input data. A few input neurons are connected to each concurrent layer of the neural network. It is known as a local receptive field for this reason. On-site neurons are the main focus of the receptive field. CNN has an edge over its forerunners in that it recognises significant aspects without human supervision. For instance, given numerous images of cats and dogs, it discovers on its own the specific characteristics for each class.

Convolution, 5x5 filters
Downsampling 2x2, batch normalization, rectification

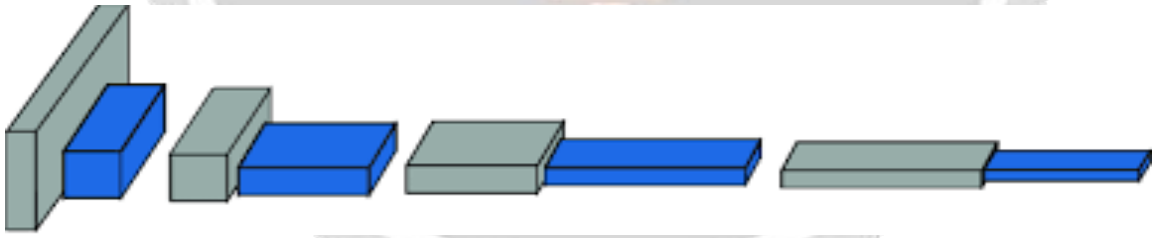


Fig:2 CNN Architecture

**Working of CNN:**

The following list includes the numerous modules that make up the convolution neural network and are structured according to certain processes:

• Insert pictures
• Layer Convolution
• Layering Pools
• Classification
• Architecture

**Insert images**: CNN receives photos as input and recognises various colour spaces while classifying items based on three different colour planes. Additionally measured is the image dimension.

LAYER KERNEL CONVOLUTIONAL:

The CNN Kernel uses the following equation to function.

Size of image: n1*n2*1.

where 1 is the number of channels, and n1 and n2 are the height and breadth, respectively.

**POOLING LAYER**: Quantityof computational resources required to process the data is decreased by using the data preparation technique.

Additionally, it is advantageous thus, the process must be kept up to par for the extraction of the dominant features, which are fundamentally rotational and positional invariant.
POOLING Forms: Following are the two primary types of pooling:

Max Pooling: In terms of value within the kernel-covered image, Max Pooling essentially gives the most value.
Average Pooling: The kernel covers an image, and the average pooling gives and returns by value inside that picture.
CLASSIFICATION: LAYER FULLY CONNECTED
As indicated by the convolutional layer's output, the inclusion of FC layer is often the simplest method for learning the non-linear combination of abstract level structures. The spaces for learning non-linear functions are provided by the FC layer.
As we have successfully completed the task of converting our output image into a particular type of multilayer perceptron, now we must turn the resulting image into a directed graph by smoothing it.The model has essentially been successful in identifying between dominant and low-level properties across various periods.
Here are some interesting CNN architectural examples:

- Alex Net
- Google Le Net
- ZF Net
- Le Net
- Res Net [6].

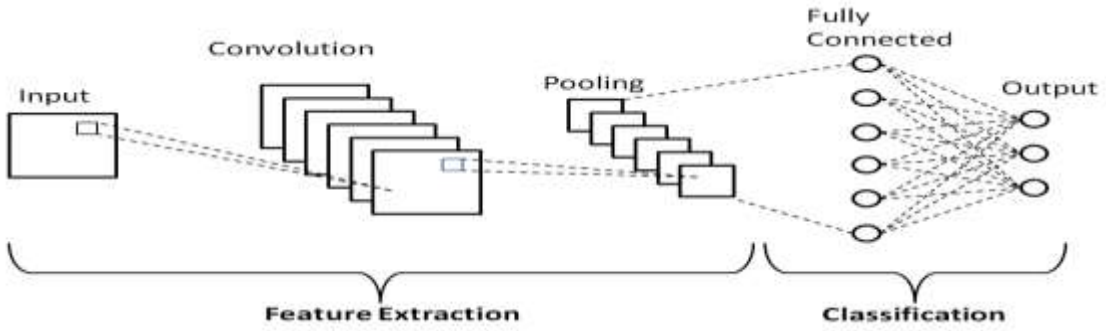



Fig 3: CNN Working

**SCOPE:**

A. Marketing Research
B. Crowd Testing
C. AI Robots
D. Helps to identify missing children
E. Banking and Telecom
F. Entertainment
G. Warehouse
H. Aviation
I. Facial Expression Detection in Interviews
J. Video Games Beta Testing [2].

**LITERATURE REVIEW:**

Humans have always found it simple to identify emotions from facial expressions, but applying a computer programme to do the same thing is quite difficult.
The classification of facial expressions using OpenCV, Keras, and Convolutional Neural Network is the main goal of the work presented in this study. The goal of facial expression recognition software is to distinguish between fundamental human emotions including happiness, sadness, anger, surprise, and neutrality.
The goal of this paper is to bring advancement and development in the field of technology [3].
This includes extraction of the feature of image which becomes easy using deep learning algorithm, and classifier model to produce output when input is given. It achieves higher accuracy as compared to traditional classifier model. It was found that the model almost predicts the emotions like happy, sad but it rarely predicts disgust emotion. Model gives the highest accuracy while predicting happiness as compared to other emotions having lower accuracy. The result for surprise is almost good [8].
In the Facial Emotion Recognition Model (FERC) that NinadMehendale suggested, they used a method that was based on a two-level CNN architecture. The first level describes how to remove the background from an input image while still preserving the core expressional vector using a normal CNN network module (EV). Here, the

EV is directly proportional to the changes in expressions on face. And the second level mainly concentrates on facial feature vector extraction. To achieve highest accuracy, they worked on a large dataset with the sample size of 10,000 images. Accuracy of this model is 96% [1].

In this model Ruhi Jaiswal used Depth- wise seperable convolutions composing of two different layers in which first layer is depth- wise convolutions which seperates the spatial cross- correlations from the channel cross-correlations. And the second layer is point- wise convolutions which generates a prediction by using a soft-max activation function and global average pooling. They worked on FER2013 datasets which is a cleaned dataset with 28,709 sets. Disgust expression is hardly understood by the trained model and, currently it has an accuracy of 66% [2].

This architecture was proposed by Dr. D. Dhanya and their team which consist of five layers for facial emotion recognition. This model was prepared by using FER2013 datasets provided by Kaggle. The first layer accepts input of size 48x48 black and white and convolved with 5x5 kernel which reduces the spatial dimensions. The second layer convolves a 3x3 kernel with 64 pixels as input from the first layer's output. The third layer now receives input from the second's output and convolves it with a 3x3 kernel and 128 filters. Finally, the output is produced by the two dense layers using the soft-max activation function. This model has a 98.7666% accuracy rate. [3].

Arvind R and the team members proposed a model Facial Emotion Recognition Using CNN, in this model image is processing using Gabor Filter, Model training using CNN, saved model using json and use it for testing, virtualisation, with Metplotlib, Gabor filters and CNN. Accuracy of this model is best in LDA (96.25%) and lowest in CNN (93%) [5].

**CONCLUSION:**

Reviews of all the related work, using traditional methods for face detection is failed because the model is not capable of classifying emotion. A deep CNN, on the other hand, can learn to identify more complex facial features, such as the texture of skin or the shape of the chin. Once a CNN has been trained on dataset of facial images, it can be used to identify faces in new images. This process is called facial recognition.

In pattern recognition and image processing, CNN is a popular and effective recognition technique. It has many benefits, including adaptability, a simple framework, and a lack of extensive training.

CNN is the greatest for facial recognition since it modified how images were previously learned. It became quite simple. CNN mimics how people view images by focusing on one section of the image at a time and scanning the complete image.

In summary, this paper has elaborated on the technique used in recent times for facial recognition and, the current trends. Along with it, discussed the scope for future directions. It is concluded that CNN in deep learning is the best method for classifying emotions.

**REFERENCES:**


1. NinadMehendale, “Facial emotion recognition using convolutional neural networks (FERC)”, SN Applied Sciences (2020), https://doi.org/10.1007/s42452-020-2234-1.
2. Ruhi Jaiswal, “Facial expression classification using CNN and its applications”, 15th (IEEE) 2020,
3. Dr. D. Dhanya et.al, “Emotion analysis using CNN”, ICCIDT (2022), ISSN: 2278-0181
4. N. Swapna Gond et.al, “Facial emoji recognition”, IJTSRD (2019), ISSN: 2456-6470
5. Arvind R et.al, “Facial Emotion Recognition Using CNN, IJRASET (2022), http://doi.org/10.22214/ijraset.2022.41536.
6. SoadAlmabdy and LamiaaElrefaei, “Deep Convolutional Neural Network-Based Approaches For Face Recognition, MDPI Applied Sciences (2019), 9, 4397; http://doi:10.3390/app9204397.
7. Li, Chieh-En James et.al, “Emotion Recognition Using CNN (2019), https://docs.lib.purdue.edu/purc/2019/Posters/63.
8. Akash Saravanan et.al, “Facial Emotion Recognition using CNN, 12 oct 2019, arXiv:1910.05602vl.
9. ByoungChul Ko, “A Brief Review of Facial Emotion Recognition Based on Visualisation”, 2018, http://doi:10.3390/s18020401.
10. Shivam Singh and Prof s Graceline Jasmine, “Facial Recognition System”, IJERT (2019), ISSN:22780181.
11. Steve Lawrence et.al, “Face Recognition: A Convolutional Neural Network Approach, IEEE 1997
12. Sameer Aqib Hashmi, “Face Detection in Extreme Conditions: A Machine-Learning Approach, email: Sameer.aqib@northsouth.edu