

# **NEURAL NETWORK TECHNIQUE FOR DETECTING FACIAL EXPRESSIONS OF EMOTION**

**A Project Work Synopsis**

*Submitted in the partial fulfillment for the award of the degree of*

**BACHELOR OF ENGINEERING**

**IN**

**AITCSE-AIML**

**Submitted by:**

**KOPPULA PRAKASH**

**21BCS8824**

**Under the Supervision of:**

**SHUBHANGI MISHRA**



**CHANDIGARH  
UNIVERSITY**

Discover. Learn. Empower.

**CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,**

**PUNJAB**

**NOV & 2022**

## **Table of Contents**

Title Page	1
Abstract	3
Timeline / Gantt Chart	4
<b>1. INTRODUCTION</b>	<b>5</b>
1.1 Problem Definition	6
1.2 Project Overview/Specifications	6
1.3 Hardware Specification	7
1.4 Software Specification	7
<b>2. LITERATURE SURVEY</b>	<b>8</b>
2.1 Existing System	8
2.2 Proposed System	9
<b>3. PROBLEM FORMULATION</b>	<b>10</b>
<b>4. RESEARCH OBJECTIVES</b>	<b>11</b>
<b>5. METHODOLOGY</b>	<b>12</b>
<b>6. TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK</b>	<b>14</b>
<b>7. REFERENCES</b>	<b>15</b>

## **ABSTRACT**

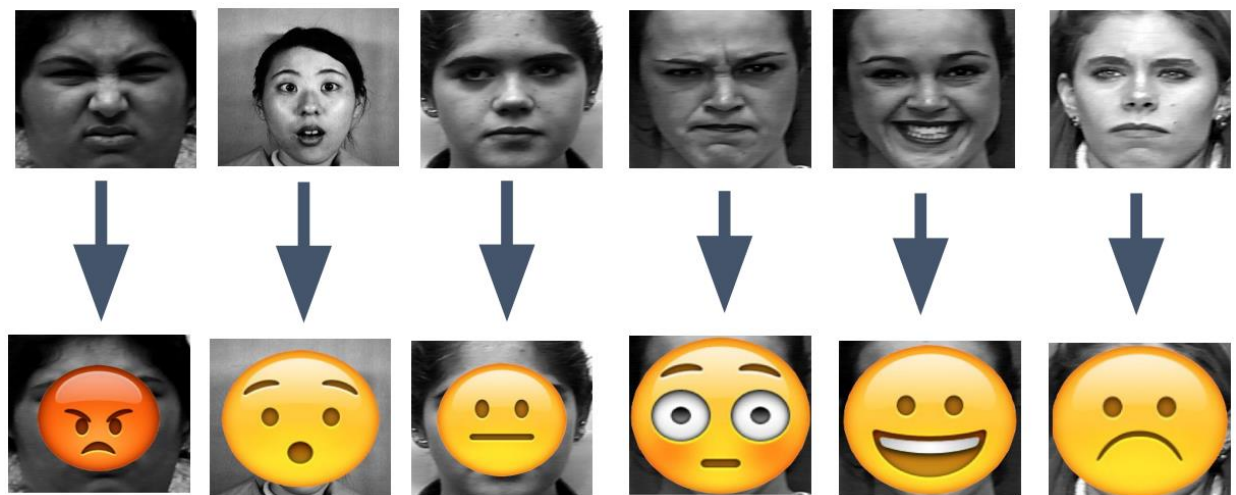
Emotion Recognition is a task to process a human facial expression and classify it into certain emotion categories. Such task typically requires the feature extractor to detect the feature, and the trained classifier produces the label based on the feature. The problem is that the extraction of feature may be distorted by variance of location of object and lighting condition in the image. In this project, we address the solution of the problem by using a deep learning algorithm called Convolutional Neural Network (CNN) to address the issues above. By using this algorithm, the feature of image can be extracted without user-defined feature-engineering, and classifier model is integrated with feature extractor to produce the result when input is given. In this way, such method produces a feature-location-invariant image classifier that achieves higher accuracy than traditional linear classifier when the variance such as lighting noise and background environment appears in the input image. The evaluation of the model shows that the accuracy of our lab condition testing data set is 94.63%, and for wild emotion detection it achieves only around 37% accuracy.

## Timeline / Gantt Chart

S.N	Strategies	1 <sup>st</sup> week	2 <sup>nd</sup> week	3 <sup>rd</sup> week	4 <sup>th</sup> week	5 <sup>th</sup> week	6 <sup>th</sup> week
1)	Problem Identification						
2)	Research & Analysis						
3)	Design						
4)	Coding						
5)	Implementation & testing						
6)	Project finalisation						
7)	Documentation						

# 1. INTRODUCTION

The facial expression of human emotion is one of the major topics in facial recognition, and it can generate both technical and everyday application beyond laboratory experiment. This project constructs a system of deep learning model to classify a given image of human facial emotion into one of the seven basic human emotions. The approach we take to build the model is through transfer learning of an existing pre-trained model, and the testing result will be evaluated based on accuracy of the model. The tasks in this project include preprocessing the image data, augmentation to enlarge the existing small dataset, test before training the model, training process, and predication with evaluation. The visual demonstration of the result will be similar to the figure below. The baseline of the test will be around 14.7% (1 in 7), and our objective is to achieve a result better than baseline accuracy.



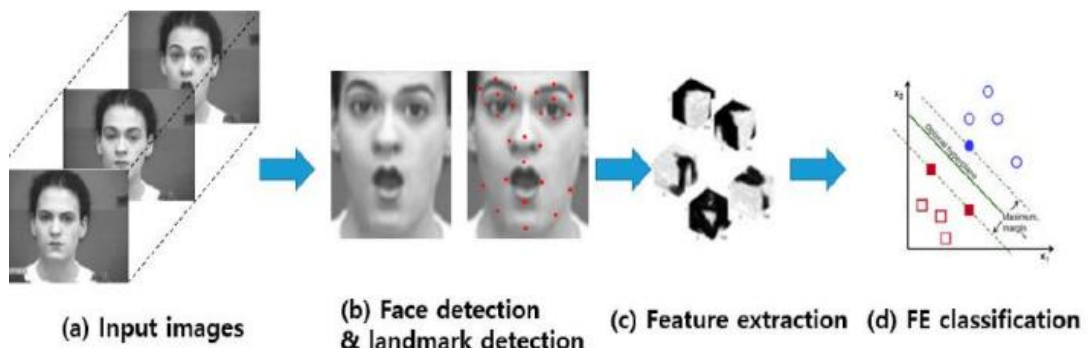
A demonstration: The emojis are correctly imposed on their corresponding faces. There are seven different emotions: happy, angry, sad, fear, surprise, neutral, and disgust. The input will be raw image of the expression, and output will be shown as above. The demonstration does not include disgust as its emoji is like angry.

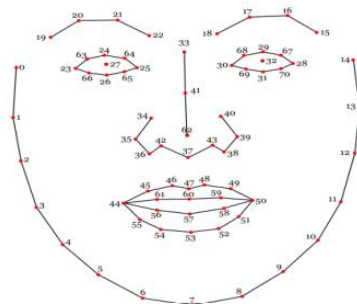
## 1.1 PROBLEM DEFINITION

In real life, people express their emotion on their face to show their psychological activities and attitudes in the interaction with other people. The primary focus of this project is to determine which emotion an input image that contains one facial emotion belongs to. Because human face is complex to interpret, emotion recognition can be specifically divided into classification of basic emotion and classification of compound emotion. For the goals of our project, the essential problem is to focus on the classification of 7 basic emotions.

## 1.2 PROJECT OVERVIEW OR SPECIFICATIONS

Facial Emotion Recognition typically has four steps. The first is to detect a face in an image and draw a rectangle around it and the next step is to detect landmarks in this face region. The third step is extracting spatial and temporal features from the facial components. The final step is to use a Feature Extraction (Facial Emotion) classifier and produce the recognition results using the extracted features. Figure 1.1 shows the Facial Emotion Recognition procedure for an input image where a face region and facial landmarks are detected. Facial landmarks are visually salient points such as the end of a nose, and the ends of eyebrows and the mouth as shown in Figure 1.2. The pairwise positions of two landmark points or the local texture of a landmark are used as features. Table 1.1 gives the definitions of 64 primary and secondary landmarks. The spatial and temporal features are extracted from the face and the expression is determined based on one of the facial categories using pattern classifiers.





A Short Introduction to Conventional Method Like every other classification problem, the emotion recognition problem requires an algorithm to complete feature extraction and categorical classification. In order to classify an emotion, we need to extract certain feature from data and build a model that can classify the input based on the feature. The procedure can be outlined as following:

1. Data Pre-processing: The data pre-processing is to standardize the data. The typical way is to set the mean of the data to 0 and to also divide the data by the standard deviation
2. Feature Extraction: The typical conventional method is to detect the face and extract the Action Units (AU)(shown in figure 5) from the face, and certain emotion contain the combination of AUs code as feature.
3. Model Construction: The conventional classifier can be either supervised or unsupervised algorithm. A typical example of supervised algorithm is Support Vector Machine, and the examples of unsupervised algorithm include Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA)
4. Label or Result Generation The typical way to generate label or result is to find which decision boundary has the minimum Euclidean distance from the data.

### 1.3 HARDWARE SPECIFICATIONS

Personal computer with keyboard and mouse maintained with uninterrupted power supply.

- Processor: Intel® core™ i5
- Installed Memory (RAM): 8.00 GB

### 1.4 SOFTWARE SPECIFICATIONS

Operating System : WINDOWS 7, 8.1,10,11  
Coding language : PYTHON  
Web Browser : GOOGLE CHROME

## 2. LITERATURE REVIEW

In recent years, facial emotion recognition has become a hot focus of research. To identify emotion from faces, most people utilize computer vision, machine learning, or deep learning technologies.

This study [1] gives a brief overview of FER research done over the last few decades. The traditional FER techniques are presented first, followed by a description of the typical FER system types and their major algorithms. The authors next describe deep-learning-based FER methods that use deep networks to enable "end-to-end" learning. This paper also looks at a new hybrid deep-learning technique that employs a convolutional neural network (CNN) for spatial characteristics of a single frame and a long short-term memory (LSTM) for temporal data of several frames. A brief overview of publicly accessible evaluation metrics is provided in the latter half of this work, as well as a comparison with benchmark findings, which constitute a standard for a quantitative comparison of FER investigations. Instead of minimizing the cross-entropy loss, learning reduces a margin-based loss.

Study of multi-level features in a convolutional neural network for facial emotion identification by Hai-Duong Nguyen [2]. They offer a model based on the data that purposely combines a hierarchy of characteristics to better the categorization job. The model was tested on the FER2013 dataset and found to be similar to existing state-of-the-art approaches in terms of performance. Using a feedforward learning model, the authors in [3] developed an instructor's face expression recognition technique

within a classroom. For successful high-level feature extraction, the face is first recognized from the obtained lecture videos and important frames are picked, removing all unnecessary frames. Then, using several convolution neural networks and parameter tweaking, deep features are retrieved and supplied to a classifier. A regularized extreme learning machine (RELM) classifier is used to classify five various expressions of the teacher within the classroom for quick learning and effective generalization of the method.

### 2.1 EXISTING SYSTEM

The existing conventional method requires a handcrafted feature extractor of facial Action Units (AUs) to extract feature from designated Facial Landmark regions, and these extracted AUs codes are processed through traditional machine learning algorithm such as Nearest Neighbors and SVM, which is a typical type of linear classifier. The problem with conventional method is that the lighting



variations and different position of object may corrupt the feature vector so that the accuracy is greatly reduced. Furthermore, it is typically difficult to conduct feature-engineering to fit the demand of facial recognition.

## **2.2 PROPOSED SYSTEM**

In this project, we approach the problem by taking deep-learning method of Convolutional Neural Networks (CNNs), which integrates the step of handcrafted feature extraction with training of classifier. This system is able to achieve the relatively most optimal solution through the process of backpropagation in which the algorithm learns the weights through modified stochastic gradient descent that can find the directions that best minimize the loss from the ground truth. The numerical result of the algorithm will show a probabilistic result of each labeled class. In order to reduce computational expense, the technique of fine-tuning is applied so that a pre-trained model can adapt the variance of our local dataset with benefit of reducing computational expense. As results, such method best resolves the issues of lighting variations and different orientation of object in the image and thus achieves a higher accuracy.

### **3. PROBLEM FORMULATION**

During software development, clones can occur in software intentionally or unintentionally. Developers tend to clone fragments of software during development to save efforts and expedite the development process.

From the literature review, it is observed that studies highlight the need of efficient and scalable approach for detecting code clones having software vulnerability. The existing techniques are not able to detect all types of vulnerable code clones. Different approaches suffer from high false negative rate and not scalable to large software systems due to high time complexity. So firstly, there is a need Haar cascade classifier. Second same subject systems should be used to compare the approaches which detect with this, the problems can be resolved by giving the better weights to model after repeating the steps of training. So, problems may arise in that. But it can be formulated and resolved by presenting correct conditions for different scenario.

## **4. RESEARCH OBJECTIVES**

The proposed work is aimed to carry out work leading to the development of an approach for Facial Emotion Recognition. The proposed aim will be achieved by dividing the work into following objectives.

1. Image data loading and visualization
2. Identify facial emotions using machine learning classifiers
3. Identify facial emotions using (convolutional) neural networks
4. Explore how sample diversity affects performance

The objective of face emotion recognition (FER) is identifying emotions of a human. The emotion can be captured either from face or from verbal communication. Psychological characteristics such as heartbeat and blood Pressure, speech, hand gestures, body movements, Facial expressions identify emotions of a person.

## 5. METHODOLOGY

Convolutional neural network (CNN) is the most popular way of analyzing images. CNN is different from a multi-layer perceptron (MLP) as they have hidden layers, called convolutional layers. The proposed method is based on a two-level CNN framework. The first level recommended is background removal [29], used to extract emotions from an image, as shown in Fig. 1. Here, the conventional CNN network module is used to extract primary expressional vector (EV). The expressional vector (EV) is generated by tracking down relevant facial points of importance. EV is directly related to changes in expression. The EV is obtained using a basic perceptron unit applied on a background-removed face image. In the proposed FERC model, we also have a non-convolutional perceptron layer as the last stage. Each of the convolutional layers receives the input data (or image), transforms it, and then outputs it to the next level. This transformation is convolution operation, as shown in Fig. 2. All the convolutional layers used are capable of pattern detection. Within each convolutional layer, four filters were used. The input image fed to the first-part CNN (used for background removal) generally consists of shapes, edges, textures, and objects along with the face. The edge detector, circle detector, and corner detector filters are used at the start of the convolutional layer 1. Once the face has been detected, the second-part CNN filter catches facial features, such as eyes, ears, lips, nose, and cheeks. The edge detection filters used in this layer are shown in Fig. 3. The second-part CNN consists of layers with  $3 \times 3$  kernel matrix, e.g.,  $[0.25, 0.17, 0.9; 0.89, 0.36, 0.63; 0.7, 0.24, 0.82]$ . These numbers are selected between 0 and 1 initially. These numbers are optimized for EV detection, based on the ground truth we had, in the supervisory training dataset. Here, we used minimum error decoding to optimize filter values. Once the filter is tuned by supervisory learning, it is then applied to the background-removed face (i.e., on the output image of the first-part CNN), for detection of different facial parts (e.g., eye, lips, nose, ears, etc.)

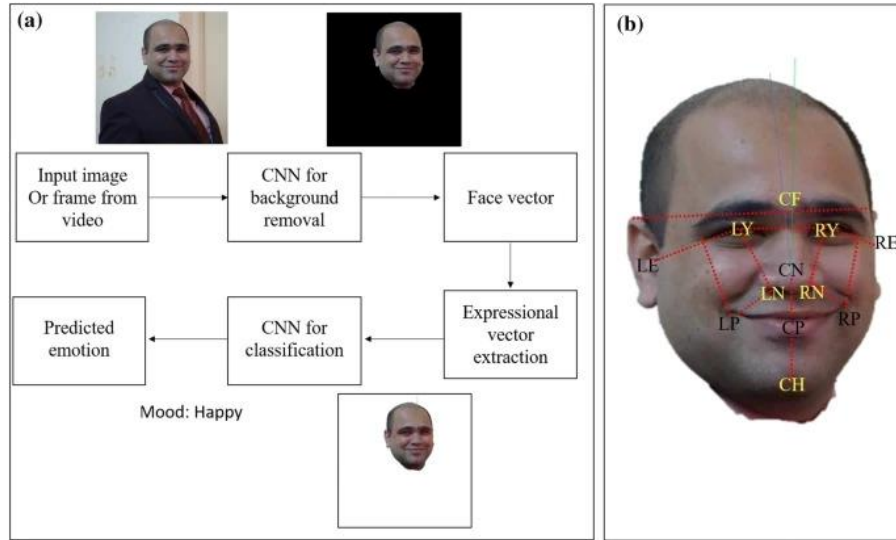


Fig. 1

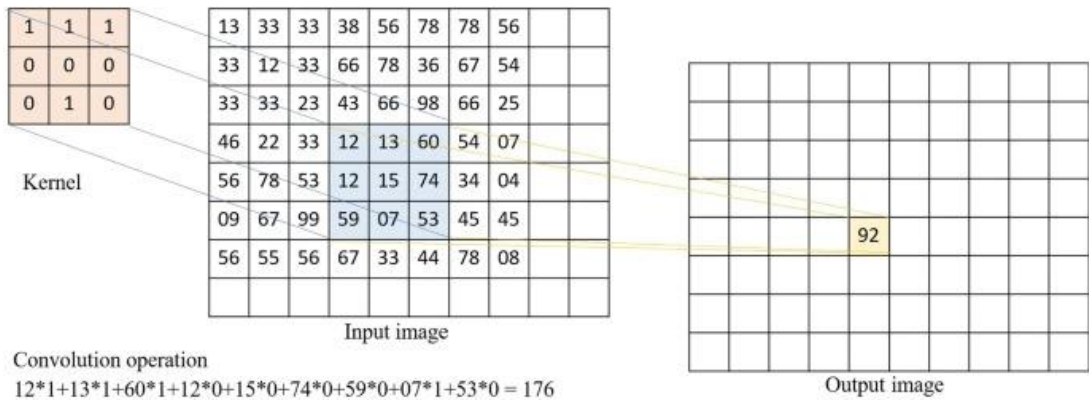


Fig. 2

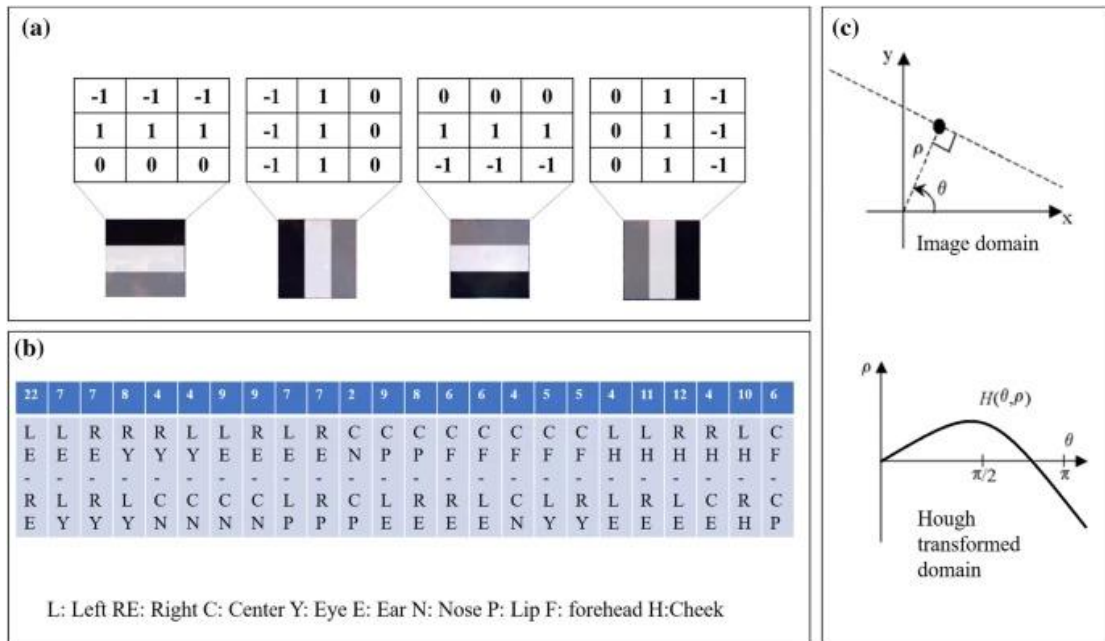


Fig. 3

## **6. TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK**

### **CHAPTER 1: INTRODUCTION**

This chapter will cover the overview of Facial Emotion Recognition.

### **CHAPTER 2: LITERATURE REVIEW**

This chapter include the literature available for Facial Emotion Recognition. The findings of the researchers will be highlighted which will become basis of current implementation.

### **CHAPTER 2: BACKGROUND OF PROPOSED METHOD**

This chapter will provide introduction to the concepts which are necessary to understand the proposed system.

### **CHAPTER 4: METHODOLOGY**

This chapter will cover the technical details of the proposed approach.

### **CHAPTER 5: EXPERIMENTAL SETUP**

This chapter will provide information about the subject system and tools used for evaluation of proposed method.

### **CHAPTER 6: RESULTS AND DISCUSSION**

The result of proposed technique will be discussed in this chapter.

### **CHAPTER 7: CONCLUSION AND FUTURE SCOPE**

The major finding of the work will be presented in this chapter. Also, directions for extending the current study will be discussed.

## 7. REFERENCES

- [1] Guo, J., Zhou, S., Wu, J., Wan, J., Zhu, X., Lei, Z., & Li, S. Z. (2017). Multi-modality Network with Visual and Geometrical Information for Micro Emotion Recognition. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 814–819.  
<https://doi.org/10.1109/FG.2017.103>
- [2] Liu, K., Zhang, M., & Pan, Z. (2016). Facial Expression Recognition with CNN Ensemble. Proceedings - 2016 International Conference on Cyberworlds, CW 2016, 163–166. <https://doi.org/10.1109/CW.2016.34>
- [3] Nour, N., Elhebir, M., & Viriri, S. (2020). Face Expression Recognition using Convolution Neural Network (CNN) Models. International Journal of Grid Computing & Applications, 11(4), 1–11.  
<https://doi.org/10.5121/ijgca.2020.11401>
- [4] Liu, Y., Yuan, X., Gong, X., Xie, Z., Fang, F., & Luo, Z. (2018). Conditional convolution neural network enhanced random forest for facial expression recognition. Pattern Recognition, 84, 251–261.  
<https://doi.org/10.1016/j.patcog.2018.07.016>
- [5] Rajendra Kurup, A., Ajith, M., & Martínez Ramon, M. (2019). Semi-supervised facial expression recognition using reduced spatial features and Deep Belief Networks. Neurocomputing, 367, 188–197.  
<https://doi.org/10.1016/j.neucom.2019.08.029>