A Major Project (CS8490) Report

On

FACIAL EMOTION DETECTION using Deep Learning

Submitted By

Md Ishtiyaque Ahmad – 1906146
Kundan Kumar – 1906148

Under the Guidance of

Dr. Somaraju Suvvari
Assistant Professor
CSE Dept., NIT PATNA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY PATNA
(An Institute of National Importance)
MAHENDRU, PATNA, BIHAR – 800005

NATIONAL INSTITUTE OF TECHNOLOGY PATNA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY PATNA
(An Institute of National Importance)
MAHENDRU, PATNA, BIHAR – 800005

CERTIFICATE

Certified that this project report "FACIAL EMOTION DETECTION using Deep Learning" the bonafide work of "Mr. MD ISHTIYAQUE AHMAD (1906146), Mr. KUNDAN KUMAR (1906148)" who carried out the project work under my supervision, during the academic year 2022-2023. It is certified that all the corrections/suggestions indicated for Internal Assessment have been incorporated in the project report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed.

Dr. Somaraju Suvvari
Assistant Professor
Supervisor

Dr. MP Singh
Professor
HOD

NATIONAL INSTITUTE OF TECHNOLOGY PATNA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY PATNA
(An Institute of National Importance)
MAHENDRU, PATNA, BIHAR – 800005

DECLARATION

We, **Md Ishtiyaque Ahmad** Roll No. **1906146** & **Kundan Kumar** Roll no. **1906148**, registered candidates for Undergraduate Programme (B.Tech.) under department of National Institute of Technology Patna, declare that this is our own original work and does not contain material for which the copyright belongs to a third party and it has not been presented and will not be presented to any other University/Institute for a similar or any other Degree award.

Signature of the members:

S. No.	NAME	SIGNATURE
1.	Md Ishtiyaque Ahmad (1906146)	…………………………….
2.	Kundan Kumar (1906148)	…………………………….

ACKNOWLEDGEMENT

Inspiration and motivation have always played a key role in the success of any venture. We feel to acknowledge our indebtedness and a deep sense of gratitude to our external guide **Dr. Somaraju Suvvari** Assistant Professor, CSE Dept., NIT PATNA. whose valuable guidance and kind supervision gave us throughout the course which shaped the present work as it shows.

We pay our deep sense of gratitude to **Prof. (Dr.) MP Singh**, H.O.D, Computer Science & Engineering Department, National Institute of Technology Patna to encourage us to the highest peak and to provide us with the opportunity to prepare for the project.

We would like to thank our Computer Science & Engineering Department, National Institute of Technology Patna and to all the faculty members for giving us continuous support and guidance that has helped us in completion of our project.

ABSTRACT

Human Emotion detection from image is one of the most powerful and challenging research tasks in social communication. Facial expression recognition is a highly relevant topic in various fields, such as artificial intelligence, customer service, gaming, marketing, and healthcare. In literature many different techniques proposed for extracting the human emotion. Recently Deep learning (DL) based emotion detection methods performing better than the traditional methods with image processing. The primary objective of this report is to classify images of human faces into seven basic emotions using deep learning techniques. Several models, including decision trees and neural networks, were explored before ultimately selecting a Convolutional Neural Network (CNN) model. CNNs are preferable for image recognition tasks as they can capture spatial features of the inputs due to their numerous filters. The final proposed model incorporates eight convolutional layers, three max pooling layers, two fully connected layers, and four dropout layers. After fine-tuning various hyperparameters, the model achieved a final accuracy of 77.74% which is better than the existing methods.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Emotion Detection

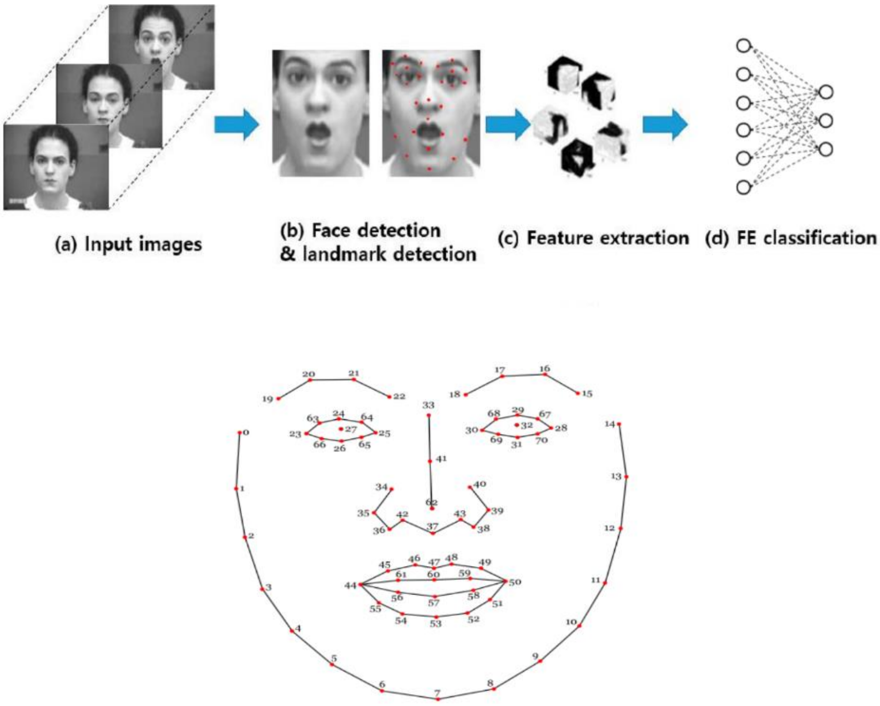
Emotion detection is the process of identifying human emotion. people vary widely in their accuracy at recognizing the emotions of others. Use of technology to help people with emotion recognition is a relatively nascent research area. Generally, the technology works best if it uses multiple modalities in context.

Emotion is a mental state associated with the nervous system associated with feeling, perceptions, behavioural reactions, and a degree of gratification or displeasure.

Emotion detection, also known as affect detection, refers to the process of identifying and analysing human emotions and emotional states using technology. This technology can be based on various methods such as machine learning, natural language processing, facial recognition, voice recognition, and physiological signals.

Emotion detection technology can be used to improve communication and understanding between humans and machines, as well as between humans themselves. For example, companies can use emotion detection to better understand their customers' needs and feelings, while healthcare professionals can use it to diagnose and treat mental health conditions more effectively

This emotion detection method is used for identifying seven emotions such as anger, disgust, neutral fear, happy, sad, and surprise using facial images.



1.1 Figure: FED procedure and Facial Landmarks

1.2 Advantages of emotion detection

1.2.1 Assess personality traits in interviews

A personal interview is a good way to interact with potential candidates and understand if they are fit for the position or not. However, it is not always possible to analyses a candidate's personality in such a short period. Moreover, many categories of discussion and judgment make it even more complex. Here emotion detection technology comes into the picture.

Emotion detection can assess and measure a candidate's emotions through facial expressions. It helps interviewers to understand a candidate's mood and personality traits.

Human resources can take advantage of this technology to create recruiting strategies and design policies to get the best performance out of employees.

1.2.2 Product testing and client feedback

Emotion detection technology can help the product industry understand the real emotions of clients when they try a product. Companies can arrange a product testing session, record it, and analyse it to detect and asses the facial emotions that arise during the session. Since emotion detection is powered by AI, it provides an ultimate method of assessing user response to new product launches.

1.2.3 Enhances customer service

Emotion detection enhances the user experience for almost every industry. Let us take examples of retail and healthcare industries to understand this. With this technology, retailers can craft better offers for their customers by analysing their browsing and buying patterns. And healthcare providers can use facial recognition to design better care plans and provide services much faster.

1.3 Applications of Emotion Detection

Emotion detection technology has emerged as a promising tool for understanding human emotions and has found applications across various industries, including customer service, marketing, healthcare, education, gaming, law enforcement, human resources, and automotive. This technology utilizes deep learning algorithms to analyse and interpret human emotions from various data sources such as facial expressions, voice tone, and physiological signals. By understanding human emotions, businesses and organizations can provide better services, support, and products that meet customers' needs, enhance employee well-being, and improve overall human experience. The following are some of the most common ones.

1.3.1 Customer Service

In customer service, emotion detection technology can help companies quickly and accurately analyse customer feedback to understand their emotions, such as frustration or satisfaction, and provide better support. By understanding customer emotions, businesses can develop more personalized and effective support strategies that address customers’ needs and concerns, ultimately leading to improved customer retention and loyalty.

1.3.2 Marketing

In marketing, emotion detection technology can help companies analyse social media posts, emails, and customer reviews to understand how their customers feel about their products or services. This information can help businesses tailor their marketing strategies to better resonate with their target audience, leading to increased brand awareness and sales.

1.3.3 Healthcare

In healthcare, emotion detection technology can help doctors and therapists better understand their patients' emotional states, such as anxiety or depression, and provide appropriate treatments. By analysing patients' emotions, healthcare professionals can develop more personalized treatment plans that address patients' emotional needs and ultimately lead to better health outcomes.

1.3.4 Education

In education, emotion detection technology can help teachers better understand their students' emotional states and adjust their teaching strategies accordingly. By analysing students' emotions, teachers can create a more supportive learning environment that addresses student’s emotional needs and ultimately leads to improved academic performance and student engagement.

Emotion detection technology has numerous applications across various industries, offering unique benefits and opportunities for businesses and organizations to improve customer satisfaction, employee well-being, and overall human experience. However, the use of emotion detection technology also raises concerns related to privacy, ethics, and bias, which must be addressed to ensure its responsible and ethical use in various industries.

1.4 Types of Emotion Detection

There are several commonly used methods for emotion detection in artificial intelligence. These include facial expression recognition, speech and voice analysis, physiological signals, text analysis, multimodal analysis, and hybrid approaches. Each method has its own unique strengths and limitations, and the choice of method depends on the specific context and application in which it will be used. Some of the most common types are as follows.

1.4.1 Facial expression recognition

This type of emotion detection analyses facial expressions to identify emotions such as happiness, sadness, anger, disgust, fear, neutral and surprise. Facial expression recognition is often used in video-based emotion detection systems.

1.4.2 Speech and voice analysis

Emotion detection can also be performed through analysing speech and voice patterns. This can include factors such as pitch, tone, and speed of speech, as well as the use of specific words and phrases.

1.4.3 Physiological signals

Emotion detection can also be based on physiological signals such as heart rate, skin conductance, and body temperature. These signals can be used to identify emotional states such as stress and anxiety.

1.4.4 Text analysis

Emotion detection can be performed by analysing written text, such as emails, social media posts, and customer reviews. Text analysis can identify emotional states and sentiment by analysing language patterns and word usage.

1.4.5 Hybrid approaches

Emotion detection can also be performed using a combination of different methods, such as combining facial expression recognition with speech analysis or physiological signals. Each type of emotion detection has its own strengths and weaknesses, and the choice of method depends on the specific application and context in which it is used.

CHAPTER 2
LITERATURE REVIEW

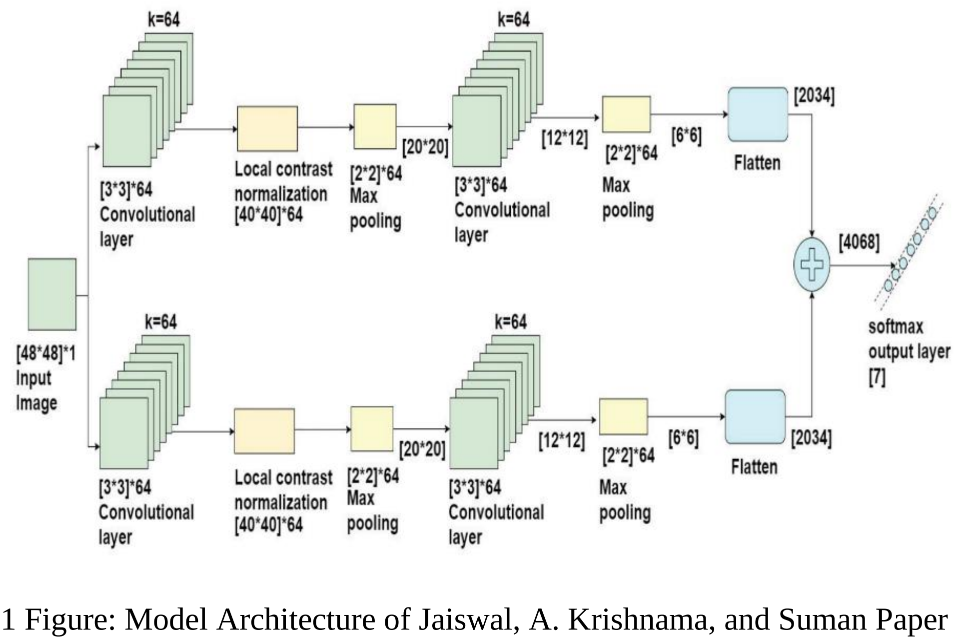
Facial emotion is a universal means for humans to communicate emotions. Facial emotion plays crucial role in various fields like customer service, marketing, healthcare, lie detection, and driving assistance systems, there have been numerous efforts to develop automated tools for facial expression analysis [7] due to its wide range of applications. The seven fundamental emotions - anger, fear, happiness, sadness, disgust, neutral and surprise - were identified by Ekman et al. [8] in the 20th century, independent of the cultural background in which an individual was raised. An image classification system typically comprises a feature extraction stage, followed by a classification stage. Fasel and Luettin conducted a comprehensive review of analytical feature extractors and neural network techniques for facial expression recognition, and concluded that both methods were roughly equally effective at the turn of the twenty-first century [9]. However, with the current abundance of training data and computational resources, it is anticipated that the performance of neural network-based models can be significantly enhanced. the following recent achievement have been established.

Krizhevsky and Hinton made a significant contribution to automatic image classification in general with their ground breaking publication [10]. They presented a deep neural network that mimics the functioning of the human visual cortex, which can be used to categorize objects in images. The researchers used their own labelled set of 60,000 images across ten categories from the CIFAR-10 dataset to develop the model. In addition, they visualized the filters in the network, enabling them to evaluate how the model analyses the images. This was a crucial outcome of the research.

A five-layer ensemble CNN was proposed by Yu and Zhang [11] for facial expression recognition. They first trained their models on FER-2013 dataset and then fine-tuned them on SFEW 2.0 dataset [12]. They used three different face detectors to extract faces from movie frames labelled with expressions. They also introduced a data augmentation and voting scheme to improve the CNN performance. They replaced max pooling layers with stochastic pooling layers to deal with the limited data size.

A deep neural network architecture for facial expression recognition was developed by Mollahosseini et al. [13]. Their network consisted of two convolutional layers followed by max pooling layers and four Inception layers. They evaluated their network on seven datasets, including the FER-2013 dataset. They also compared their network with an AlexNet [14] network on the same datasets. Their network achieved higher accuracy on the MMI and FER-2013 datasets and similar accuracy on the other five datasets. Their network reached an accuracy of 0.664 on the FER-2013 dataset.

In 2020 INCET Jaiswal, A. Krishnama, and Suman proposed two model (A and B) architecture [3] which utilizes a deep learning architecture based on convolutional neural networks (CNN) for emotion detection from images. the performance of the proposed method is evaluated using two datasets, the Facial Emotion Recognition Challenge (FERC-2013) and Japanese Female Facial Emotion (JAFFE), achieving accuracies of 70.14% and 98.65%, respectively. Figure 2.1 model architecture of this report.



2.1 Figure: Model Architecture of Jaiswal, A. Krishnama, and Suman Paper [3]

A paper by Saravanan, Perichetla, and Dr. Gayathri[1], in this paper they tested various models, such as decision trees and neural networks, before choosing a Convolutional Neural Network (CNN) model for image recognition. CNNs can extract spacial features from the images using many filters. They used FER-2013 dataset and a model that has six convolutional layers, two max pooling layers and two fully connected layers. After adjusting the hyperparameters, they obtained an accuracy of 0.60.

Proposed Convolutional Neural Network
CONV2D-64 RELU
CONV2D-64 RELU MAXPOOL2D DROPOUT
CONV2D-128 RELU
CONV2D-128 RELU
CONV2D-256 RELU
CONV2D-256 RELU MAXPOOL2D DROPOUT
FLATTEN
FULLY CONNECTED RELU DROPOUT
FULLY CONNECTED SOFTMAX

2.2 Figure: Model Architecture of Saravanan, Gurudutt, and Dr. Gayathri Paper [1]

CHAPTER 3
DATASET PREPARATION

3.1 Dataset Collection

In general, neural networks, especially deep neural networks, tend to perform better when larger amounts of training data set are present. In addition, the choice of images used for the training is responsible for a large part of the eventual model's performance. It means the need for a data set that is both high quality and quantitative. Several datasets are available for research to recognize emotions, ranging from a few hundred high resolution photos to tens of thousands of smaller images. With this in mind, we chose Facial Expression Recognition (FER-2013).

The FER-2013 dataset was introduced in the ICML 2013 Challenges in Representation Learning [6]. It contains 35,887 images with the following basic expressions: angry, disgusted, fear, happy, sad, surprised, and neutral. Each image is a frontal view of a subject, taken from the wild and annotated to one of the seven expressions. the datasets primarily vary in the amount, consistency, and cleanness of the images. For example, the FERC-2013 collection has about 32,000 low-resolution images. while the FERC-2013 set displays "in the wild" emotions. This makes it harder to interpret the images from the FERC 2013 set, but given the large size of the dataset, it increases robustness of the model and it also can be beneficial for the diversity of dataset.

3.2 Pre-processing

The dataset consisted of a number of images, each of which was represented as a 48×48 matrix. where each number represented the value of a pixel.

The original dataset of 35,887 images was split into a training set of 28,709 images and a testing set of 7,178 images, in an 80:20 split.

from testing set taken 70% of image for validation and rest are for testing purpose.

In deep learning, the size of the training set is the most important factor in determining the accuracy of the model. A larger training set will generally lead to a more accurate model. However, it is also important to have a testing set to evaluate the model's performance on unseen data. If the testing set is too small, it may not be representative of the overall dataset, and the model may not generalize well to new data. In this case, the testing set was 20% of the total number of images. This is a relatively small testing set, but it is sufficient to evaluate the model's performance. A larger testing set would have been preferable, but it would have required more time and computational resources.

A one-hot encoding scheme was used for the labels. This means that each label was represented as a vector of 7 elements, where each element was either 0 or 1. This allowed the model to learn the relationship between the images and the labels.

It is to be noted that the number of disgusted expressions (547) is much lower in comparison to the other expressions. There was also an obvious bias towards happy expressions due to the sheer number of sample data present for the expression. To remove this biasness of model towards the happiness we first balance the dataset using augmentation technique.

During the live testing, Haar Cascades [15] were used to identify a face. Haar Cascades [15] are a type of machine learning algorithm that can be used to detect objects in images. They are relatively fast and efficient, and they are able to detect faces with a high degree of accuracy.

CHAPTER 4
FACIAL EMOTION DETECTION

There are several different models were used in order to select the best foundation to work on. From the literature review tried few models and got approximately same result. After several attempt we decided to merge two models together but result was not quite well. then we want to implement augmentation on the dataset but problem was dataset disbalanced. Due to disbalance we could not make it balance so we balance dataset individually and saved in the folders. then we run on the proposed CNN model of Saravanan got higher accuracy (74%). Further we experimented with model, we got below model architecture. All convolution layer kernel size is 3x3 and maxpooling2d kernel size 2x2.

4.1 Proposed Model Architecture

INPUT LAYER
CONV2D-64 RELU CONV2D-64 RELU DROPOUT 0.25
CONV2D-128 RELU CONV2D-128 RELU MAXPOOL2D
CONV2D-256 RELU CONV2D-256 RELU MAXPOOL2D DROPOUT 0.25
CONV2D-512 RELU CONV2D-512 RELU MAXPOOL2D DROPOUT 0.25
FLATTEN FULLY CONNECTED RELU DROPOUT 0.2 FULLY CONNECT SOFMAX
OUTPUT LAYER

Table 4.1 Proposed model architecture

4.2 Model Training

The dataset was initially split into an 80% training set and a 20% testing set. 70% of the testing data was used as a validation dataset.

The network was trained using a GPU for 50 epochs to ensure that the precision converges to the optimum. The network was trained on a larger dataset than the one previously described, to improve the model even more. Training took place with 28,709 images from the FERC-2013 dataset. The FERC-2013 database also uses verification dataset (5024 images). The accuracy was higher on all validation and test sets than in previous runs, emphasizing that emotion detection using deep convolutional neural networks can improve the performance of a network. After training over, training accuracy was 95.7% and validation accuracy was 77.6%.

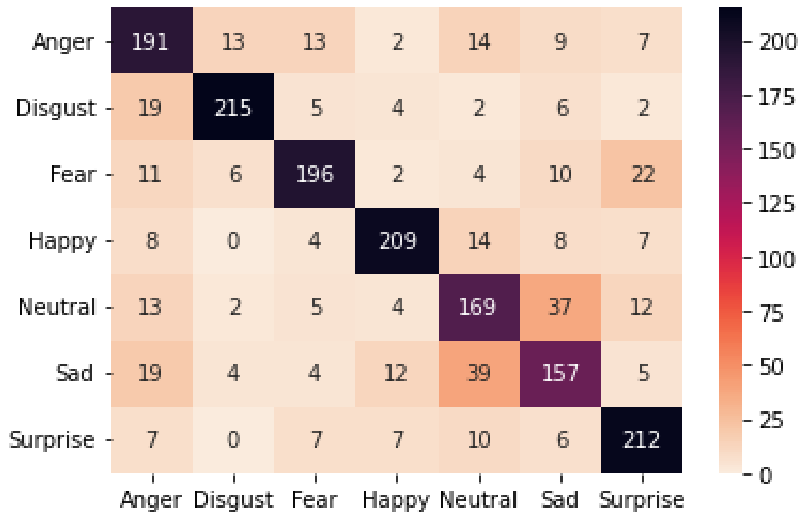
CHAPTER 5
RESULTS

After extensive training, the model was ready for the testing phase, the trained networks were loaded and fed the entire testing set, one image at a time. Each image was new to the model, and had not been seen during training. The image was pre-processed in the same way as described in the previous section. The model did not know the correct output for each image, and had to accurately predict it based on its own training. It attempted to classify the emotion shown on the image simply based on what it had already learned, along with the characteristics of the image itself. In the end, it gave a list of classified emotion probabilities for each image. The highest probability emotion for each image was then compared with the actual emotions associated with the image and calculated the accuracy of model. And finally, we got accuracy 61.31% without augmentation and with augmentation 77.74%.

S.No.	MODEL	ACCURACY
1.	Forward NN	17.38%
2.	CNN model by Jaiswal [3]	50%
3.	CNN model by Saravanan [1]	60%
4.	CNN model by Saravanan with augmentation	74%
5.	Proposed Model	61.31%
6.	Proposed Model with augmentation	77.74%

Table 5.1 Comparison of models

5.1 Confusion Matrix



5.1 Figure: Confusion Matrix of final model

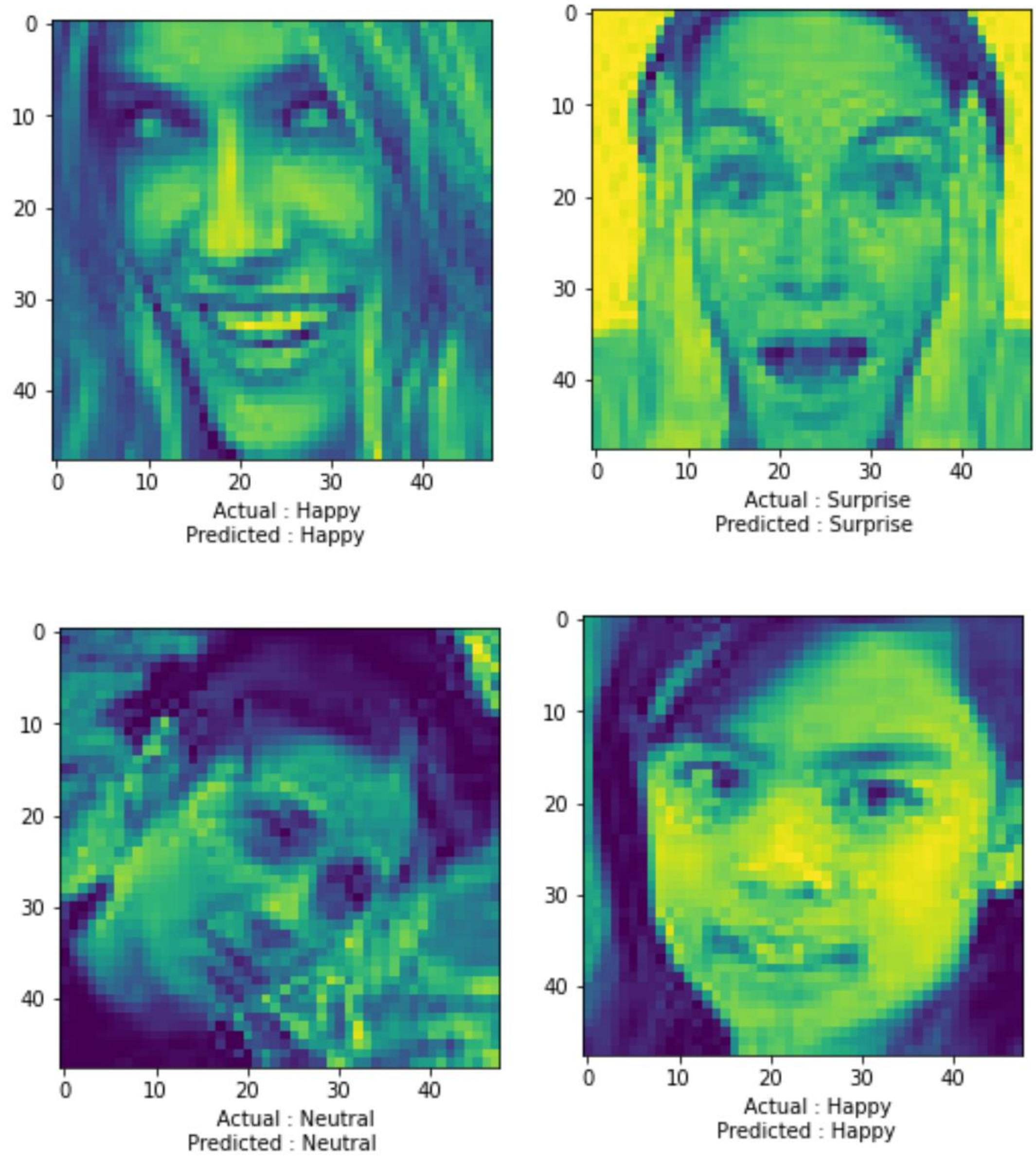
5.2 Classification Report

Classification Report

	precision	recall	f1-score	support
Anger	0.71	0.77	0.74	249
Disgust	0.90	0.85	0.87	253
Fear	0.84	0.78	0.81	251
Happy	0.87	0.84	0.85	250
Neutral	0.67	0.70	0.68	242
Sad	0.67	0.65	0.66	240
Surprise	0.79	0.85	0.82	249
accuracy			0.78	1734
macro avg	0.78	0.78	0.78	1734
weighted avg	0.78	0.78	0.78	1734

5.2 Figure: Classification report of final model

5.3 Predicted Images



5.3 Figure: Predicted results of the model

5.4 Figure: FED App Screen shot

CHAPTER 6

CONCLUSION

In this report, we present a novel approach to facial expression recognition (FER) using deep learning. We evaluated several different models on the FER-2013 dataset, including feed-forward neural networks, and smaller convolutional networks. We found that a deep convolutional neural network (CNN) with modified hyperparameters achieved the best performance, with an accuracy of 77.74%. The performance of the proposed facial emotion detection model is evaluated in terms of validation accuracy, learning rate, and validation loss. The model is trained and tested on a dataset of images, and its performance is compared to that of previous models. The results show that the proposed model outperforms previous models in terms of validation accuracy, and learning rate. And, next we built a web app that detect facial emotion in nextjs and python.

REFERENCES

[1] Saravanan, Perichetla, Dr. Gayathri, Facial Emotion Recognition using Convolutional Neural Networks, https://arxiv.org/pdf/1910.05602v1.pdf , [cs.CV] 12 Oct 2019

[2] P. Ekman, & W. V. Friesen. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17(2), 124-129. (1971)

[3] Jaiswal, A. Krishnama Raju, Suman Deb, Facial Emotion Detection Using Deep Learning https://ieeexplore.ieee.org/document/9154121, 2020 International Conference for Emerging Technology (INCET)

[4] S. Li and W. Deng, "Deep facial expression recognition: A survey," arXiv preprint arXiv:1804.08348, 2018.

[6] Challenges in representation learning: Facial expression recognition challenge. (2013) http://www.kaggle.com/c/challenges-in-representation learning-facial-expression-recognition-challenge

[7] Zafar B, Ashraf R, Ali N, Iqbal M, Sajid M, Dar S, Ratyal N (2018) A novel discriminating and relative global spatial image representation with applications in CBIR. Appl Sci 8(11):2242

[8] Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. J Personal Soc Psychol 17(2):124

[9] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. Pattern recognition, 36(1):259–275, 2003.

[10] A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images, 2009.

[11] Zhiding Yu & Cha Zhang. (2015). Image based Static Facial Expression

Recognition with Multiple Deep Network Learning. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction . 435- 442

[12] Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon. (2011). Static Facial Expressions in Tough Conditions: Data, Evaluation Protocol And Benchmark, First IEEE International Workshop on Benchmarking Facial Image Analysis Technologies BeFIT, IEEE International Conference on Computer Vision ICCV2011, Barcelona, Spain, 6-13 November 2011

[13] A. Mollahosseini, D. Chan and M. H. Mahoor. (2016). Going deeper in facial expression recognition using deep neural networks. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, 1-10.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 10971105

[15] Open Source Computer Vision. Face Detection using Haar Cascades: https://docs.opencv.org/3.4.1/d7/d8b/tutorial_py_face_detection.html