FACIAL EMOTION RECOGNITION USING DEEP LEARNING

A PROJECT REPORT

Submitted by

LISMY THOMAS [SCM21MCA-2012]

to

The APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree

of

Master of Computer Application

Department of Computer Science and Engineering

SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY

(*Affiliated to APJ Abdul Kalam Technological University*)

VIDYA NAGAR, PALISSERY, KARUKUTTY

ERNAKULAM - 683582

DECEMBER 2023

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY

(*Affiliated to APJ Abdul Kalam Technological University*)

VIDYA NAGAR, PALISSERY, KARUKUTTY

ERNAKULAM - 683582

CERTIFICATE

This is to certify that the report entitled **'FACIAL EMOTION RECOGNITION USING DEEP LEARNING'** submitted by **LISMY THOMAS** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Master of Computer Application is a bonafide record of the project work carried out by him/her under my guidance and supervision.

Project Guide	Project Coordinator	Head of the Department
Ms. Greshma P Sebastian	Ms. Deepa K	Dr. Varun G Menon

ACKNOWLEDGEMENT

I am greatly indebted to Dr.Anitha G.Pillai, Principal, SSET, Ernakulam and Dr. Varun G Menon, Head of department, Department of Computer Science and Engineering, SSET, who whole heartedly granted me the permission to do the project.

I would like to thank my guide, Ms. Greshma P Sebastian , and project coordinator, Ms. Deepa K, Department of Computer Science and Engineering, SSET who has given me valuable guidance and support all the way.

I would like to express my sincere gratitude to all the teachers of Computer Science Department who gave me moral and technical support. I would like to thank the supporting staff in the Computer lab whose dedicated work kept the lab working smoothly, thus enabling me to have access to various resources which helped me understand more about the project topic. I would also like to thank friends and family members for providing me with necessary resources and support. Last but not the least, I would like to thank God Almighty for helping me to do my project hassle Free.

ABSTRACT

Facial emotion recognition is a crucial aspect of human-computer interaction that involves detecting and classifying emotions based on facial expressions. This field has applications in a variety of areas including psychology, market research, and security. With advancements in computer vision and deep learning, Convolutional Neural Networks (CNNs) have become a popular tool for automated facial emotion recognition. In this project, we aim to develop a robust facial emotion recognition system based on CNNs. The dataset used for training and evaluation will consist of images of faces with corresponding emotions. We will explore various architectures and hyper-parameters to identify the best model for our task. The performance of the model will be evaluated using standard metrics such as accuracy, precision, and recall. The results of this project will demonstrate the potential of deep learning in the field of facial emotion recognition and provide insights into the development of more advanced systems in the future.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER-1

INTRODUCTION

1.1. GENERAL BACKGROUND

Machine learning is a subset of artificial intelligence that allows algorithms to learn from data, discover patterns and make predictions without explicit programming. There are three main types: supervised, unsupervised and reinforcement learning, and it is used in diverse fields such as image and speech recognition, natural language processing and predictive analytics.

Facial emotion recognition refers to the task of detecting and classifying emotions based on facial expressions. It is a crucial aspect of human-computer interaction and has applications in various fields such as psychology, market research, and security. The recognition of facial emotions is based on the analysis of changes in facial features, such as eye shape, mouth shape, and wrinkles that are associated with different emotions. Traditional methods for facial emotion recognition relied on hand-engineered features, but recent advancements in computer vision and deep learning have led to the development of automated facial emotion recognition models using Convolutional Neural Networks (CNNs). These models have shown remarkable accuracy and are being used in a variety of applications.

1.2. OBJECTIVE

Our emotion is revealed by the expressions in our face. . The purpose of this project is to create a system that can identify and categorize facial expressions into one of seven emotions: angry, disgust, fear, happy, sad, surprise, or neutral, by analyzing human facial images as input. This will be done through the application of computer vision techniques, specifically by training a model using deep learning on a dataset of facial expressions that have been labeled with the corresponding emotion.

1.3. SCOPE

Facial expression recognition has a wide range of potential applications. One area of application is human-computer interaction, where the technology can be used to make the interactions between humans and computers more natural and intuitive. Additionally, affective computing is another field that can benefit from facial expression recognition, allowing systems to sense, interpret and respond to human emotions. In the medical field, facial expression recognition can be used to monitor and diagnose mental health conditions such as depression and anxiety. Other areas of application include marketing and advertising, where the technology can be used to analyze consumer reactions to products and advertisements, security and surveillance, Robotics and animatronics, gaming, and personal development and self-care. The possibilities of facial expression recognition are endless and with the development of technology and research in this field, it is expected to be more widely adopted in different industries.

1.4 ORGANIZATION OF REPORT

The report is organized into five chapters. The first chapter of the report deals with the general background, objective and scope of the project. The second chapter contains various literature reviews related to the project. The detailed architecture and working of the proposed system is discussed in the third chapter .The fourth chapter contains the experimental results and discussions. The conclusions are summarized in the final chapter. The future scope of the given project is also added in the last chapter. Finally the references are given in the last pages.

CHAPTER-2

LITERATURE SURVEY

Facial expression recognition has been an active area of research in the field of computer vision and machine learning for several decades. There are various approaches that have been proposed in the literature to recognize facial expressions.

The author Ninad Mehendale[2] trained a CNN model on the FER2013 dataset, which consists of 35,887 images of faces labeled with one of seven emotions. The model achieved an accuracy of 70.14% on the test set of the FER2013 dataset. The model was also tested on other datasets, such as CK+ and JAFFE. The paper concludes that the proposed CNN-based model is an effective approach for facial expression recognition.

"Facial emotion recognition using deep convolutional neural network" is a research paper [3] that proposes the use of deep convolutional neural networks (CNNs) for facial expression recognition. The authors trained their model on the FER2013 dataset, which consists of 35,887 images of faces labeled with one of seven emotions. The model achieved an accuracy of 68.48% on the test set of the FER2013 dataset, which is comparable to the state-of-the-art results on the same dataset.

Mohammadpour, Mostafa [4] presents a method that uses deep CNNs to recognize facial emotions. The method is trained and tested on CK+ and FER2013 datasets and compared with traditional machine learning algorithms. The effect of different CNN architectures on performance is also evaluated and it is shown that the proposed method outperforms traditional algorithms.

CHAPTER-3

PROPOSED SYSTEM

To identify and categorize human emotions from facial expressions. Face emotion recognition system must go through several phases, including Pretreatment, feature extraction, classifier training and classification. Preprocessing images reduces the size of all images to a same size. Then the features of the preprocessed image are extracted. Convolutional Neural Network is the algorithm is used (CNN).

3.1 CONVOLUTIONAL NEURAL NETWORK

3.1.1 INTRODUCTON

Convolutional Neural Networks (CNNs) are a type of deep learning algorithm commonly used for image and visual recognition tasks. They are composed of multiple layers including convolutional layers, activation functions, pooling layers and fully connected layers. CNNs work by sliding a filter over an input image to extract features and perform convolutional operations on those features. The extracted features are then passed through activation functions, pooling layers and fully connected layers to classify the input image into various classes. CNNs have shown remarkable success in various computer vision applications including image classification, object detection, segmentation, and generation.

3.1.2 OVERVIEW OF METHOD

The three fundamental layers of the convolutional neural network are as follows:

- Convolutional Layer
- Pooling Layer

- Fully Connected Layer

Convolutional Layer: A convolutional layer is a key component of a Convolutional Neural Network (CNN). It performs a mathematical operation called "convolution" on the input image, sliding a filter (also called a kernel) over the image to extract features and create a feature map. The extracted features are then processed by activation functions and pooling layers to reduce the spatial dimensions and increase the invariance of the features. Convolutional layers play a crucial role in the overall performance of a CNN and are responsible for learning and extracting useful features from input images.

Pooling Layer: A pooling layer is a component in Convolutional Neural Networks (CNNs) used to down-sample the spatial dimensions of the feature map produced by a convolutional layer. It is applied after one or multiple convolutional layers to reduce the computational cost and control over-fitting. Pooling operates by taking the maximum or average of a small region of the feature map, resulting in a new, smaller feature map. Common types of pooling include max pooling and average pooling. Pooling helps to preserve the most important features in the image, while reducing the number of parameters and computational cost.

i. **Max Pooling -** Returns the largest value from the picture covered by the Kernel's array.

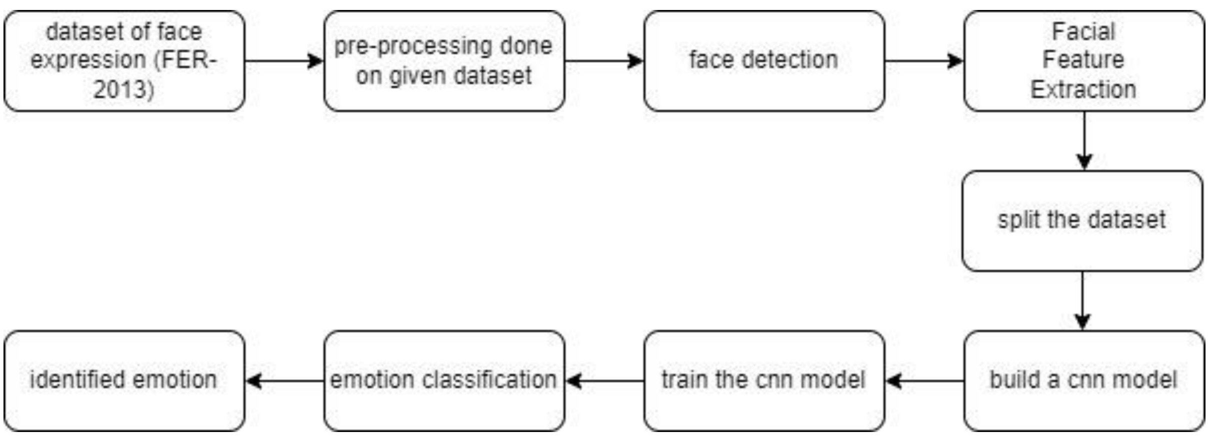
ii. **Average Pooling**- Returns the average of all the values in the image that the Kernel's array covers.

Fully Connected Layer (FC): A fully connected layer in a CNN is used for the final classification of input images. It is connected to all neurons in the previous layer, performs matrix multiplication and activation to produce a prediction, and is typically the last layer in a CNN.

3.1.3 OVERVIEW OF RESULT

Facial emotion recognition is a field of computer vision and machine learning that aims to classify human emotions based on facial expressions. Results have shown promising accuracy in recognizing emotions such as happiness, sadness, anger, fear, and surprise. However, challenges remain, including recognizing emotions in diverse individuals and lighting conditions. Despite these challenges, facial emotion recognition technology is used in various applications such as human-computer interaction and mental health assessment. The field is expected to continue growing and evolving with the development of advanced algorithms and annotated datasets.

3.2. Architecture



This process involves collecting face images, preprocessing them, training a CNN model on the processed images with corresponding emotion labels, evaluating the model on unseen data, and finally using the trained model to predict emotions in new face images. The interface was created using gradio for test data

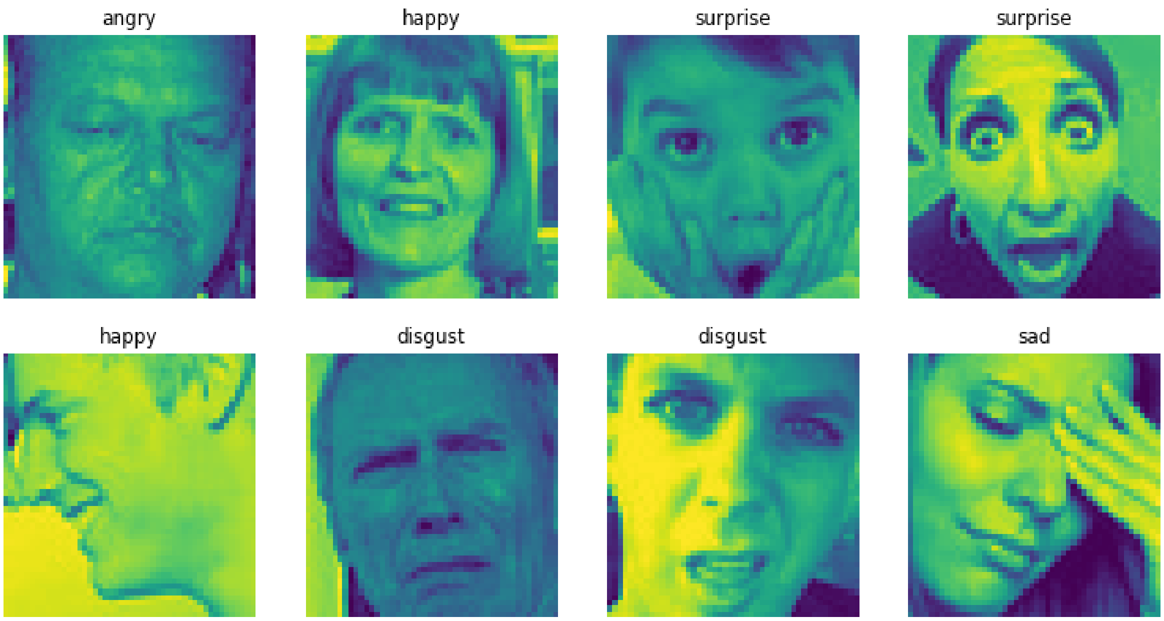


Fig 3.2 Dataset of FER-2013

3.2.1 .Pre-Processing

Preprocessing involves detecting and locating faces in an image, aligning them to a standard position, cropping to a consistent size, and normalizing the pixel values to reduce variability in the data. This helps improve the performance of the model by providing it with consistent, high-quality input data.

3.2.2 .Face Detection

Face detection in facial emotion recognition involves identifying the presence and location of faces in an image. This step is crucial for the subsequent processing steps, as the model needs to know where the face is in order to perform accurate analysis.

3.2.3. Feature Extraction

Feature extraction involves identifying and analyzing significant facial features such as points and textures, to understand the emotions being expressed. This information is used to generate an expression vector, which can then be processed by the model to classify the emotion being expressed.

3.2.4. Split the Dataset

Dataset splitting is the process of dividing a large dataset into smaller parts for training, validation, and testing purposes. The training set is used to train the model, the validation set is used for hyper-parameter tuning and model selection, and the test set is used for evaluating the final model performance. The split ratio for each set is usually determined based on the size of the dataset and the research requirements, but common ratios are 80/10/10 or 70/15/15 for training, validation, and testing, respectively.

3.2.5. Build CNN Model

Building a Convolutional Neural Network (CNN) model for facial emotion recognition involves defining the network architecture, hyper-parameters, and training the model on a dataset of facial images with corresponding emotions. This includes choosing the number and type of layers, activation functions, and optimizing the model using gradient descent.

CHAPTER-4

RESULTS AND DISCUSSION

The project of plant disease detection system was able to produce a model that is able to classify the emotions of humans. The model was built using Convolutional Neural Network (CNN).The model is defined and then it is trained by giving a huge dataset containing 28709 images for training and 7178 for testing. Thus, trained the model using the dataset and then when an input image is given, it classifies the image, predicts the category of emotion. Here, the sequential model was built and the layers of convolution, normalization, pooling, dropout, and activation were added in place. Then I initialized the Adam optimizer with the learning rate and attenuation parameters.

- Images are loaded and labels are identified:

```
for folder in os.listdir(TRAIN_DIR):
    files = gb.glob(pathname= str(TRAIN_DIR+ '/'+ folder + '/*.jpg'))
    print(f'For training data, found {len(files)} in folder {folder}')
```

```
For training data, found 4830 in folder sad
For training data, found 3995 in folder angry
For training data, found 3171 in folder surprise
For training data, found 4965 in folder neutral
For training data, found 7215 in folder happy
For training data, found 4097 in folder fear
For training data, found 436 in folder disgust
```

```
for folder in os.listdir(TEST_DIR):
    files = gb.glob(pathname= str(TEST_DIR+ '/'+ folder + '/*.jpg'))
    print(f'For testing data, found {len(files)} in folder {folder}')
```

```
For testing data, found 831 in folder surprise
For testing data, found 958 in folder angry
For testing data, found 1774 in folder happy
For testing data, found 1024 in folder fear
For testing data, found 1247 in folder sad
For testing data, found 1233 in folder neutral
For testing data, found 111 in folder disgust
```

Fig 4.1.Loaded images form dataset.

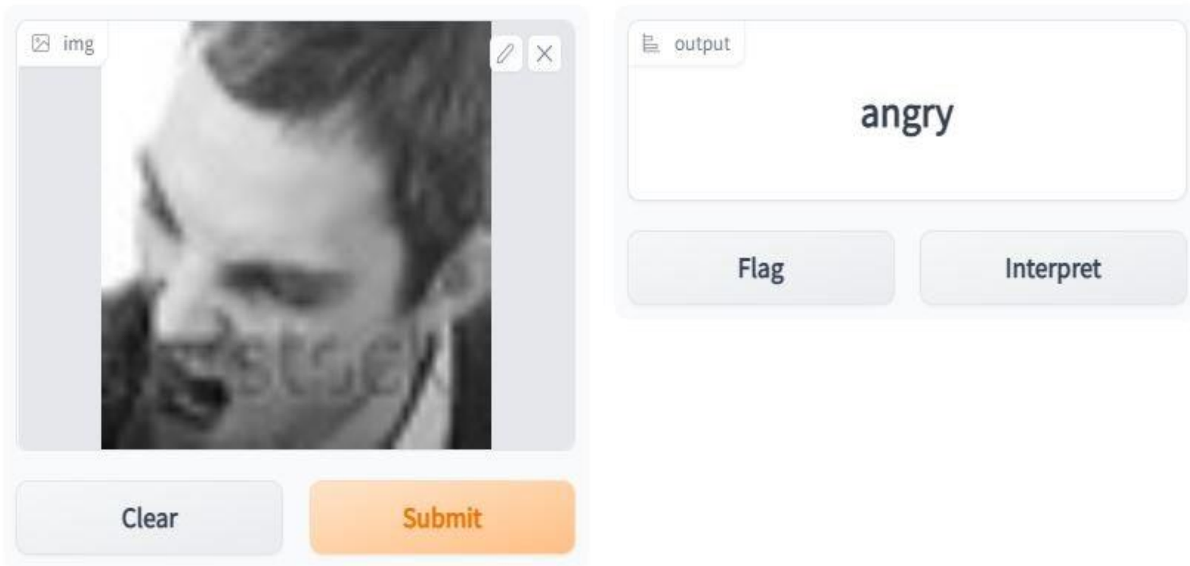


Fig 4.2.Interface using Gradio

The system's accuracy might be calculated based on the results. Based on the available dataset, the system's accuracy is around 65 percent. Certain factors that affect the accuracy of the system are the number of epochs and the number of images used for training. As both parameters increase, so does the accuracy.

CHAPTER-5

CONCLUSION & FUTURE SCOPE

The project is simple and effective, and it is used to evaluate face expression. In order to decide whether the emotion state of a person, various procedures must be followed, such as preprocessing, feature extraction, classifier training, and classification. When you preprocess an image, you reduce the size of all the photos to the same size. Following that, the features of a preprocessed image are extracted. The algorithm used is a Convolutional Neural Network (CNN).

Facial emotion recognition (FER) is a field of study in computer vision that focuses on identifying human emotions from facial expressions. Over the past few years, Convolutional Neural Networks (CNNs) have become a popular approach for FER, achieving high accuracy in many benchmark datasets. Despite these advances, there are still some limitations and challenges to be addressed in order to make FER more practical and reliable.

One of the main limitations of current FER systems is their robustness to variations in facial expressions and illumination. In real-world scenarios, facial expressions can vary greatly due to individual differences, cultural background, and environmental factors. In addition, the quality of images can be degraded by changes in lighting conditions. These factors can result in significant performance degradation for current FER system. Another challenge for FER is the limited ability to recognize complex and multi-dimensional emotions. Emotions are not always easily expressed through facial expressions, and multiple emotions can co-occur in a single expression. In order to accurately capture the full range of human emotions, it may be necessary to integrate additional modalities such as audio and text.

Despite these challenges, there are many exciting directions for future research in FER. Here are a few potential areas of focus:

- Robust and generalizable models: One of the main goals of future FER research will be to develop models that are robust to variations in facial expressions and illumination, and can generalize well to new scenarios and subjects. This will require the integration of new techniques such as data augmentation, transfer learning, and domain adaptation.
- Multi-modal emotion recognition: As mentioned, emotions are complex and multi-dimensional, and facial expressions are not always a reliable indicator. Integrating additional modalities such as audio, text, and physiological signals may allow for more comprehensive and accurate emotion recognition.
- Real-world applications: FER has the potential to be applied in many real-world scenarios such as human-computer interaction, video analysis, and mental health diagnosis. In order to be practical and useful in these applications, FER systems need to be reliable, fast, and user-friendly.
- Novel CNN architectures: In recent years, there has been a significant push towards developing new and innovative CNN architectures for various computer vision tasks, including FER. There is a great deal of room for improvement in this area, and new architectures may help overcome current limitations and achieve higher accuracy.

Overall, the field of facial emotion recognition using CNNs is a rapidly growing and exciting area of research with many exciting possibilities for future development.

REFERENCES

[1] Mehendale, Ninad. "Facial emotion recognition using convolutional neural networks (FERC)." SN Applied Sciences 2.3 (2020): 1-8.

[2] Pranav, E., et al. "Facial emotion recognition using deep convolutional neural network." 2020 6th International conference on advanced computing and communication Systems (ICACCS). IEEE, 2020.

[3] Mohammadpour, Mostafa, et al. "Facial emotion recognition using deep convolutional networks." 2017 IEEE 4th international conference on knowledge-based engineering and innovation (KBEI). IEEE, 2017

[4] Y. A. Bachtiar., 2019. Convolutional Neural Network and Maxpooling Architecture on Zynq SoC FPGA.

[5] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). Comput. Vis. Image Underst. 110, 346–359. doi: 10.1016/j.cviu.2007.09.014.

[6] Rzayeva, Zeynab, and Emin Alasgarov. "Facial emotion recognition using convolutional neural networks." 2019 IEEE 13th international conference on application of information and communication technologies (AICT). IEEE, 2019.

[7] Akhand, M. A. H., et al. "Facial emotion recognition using transfer learning in the deep CNN." Electronics 10.9 (2021): 1036.

[8] Verma, Abhishek, Piyush Singh, and John Sahaya Rani Alex. "Modified convolutional neural network architecture analysis for facial emotion recognition." *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, 2019.

[9] Monica, Shrinitha, and R. Roseline Mary. "Face and Emotion Recognition from Real-Time Facial Expressions Using Deep Learning Algorithms." *Congress on Intelligent Systems*. Springer, Singapore, 2022.

[10] Duncan, Dan, Gautam Shine, and Chris English. "Facial emotion recognition in real time." *Computer Science* (2016): 1-7.

[11] Mellouk, Wafa, and Wahida Handouzi. "Facial emotion recognition using deep learning: review and insights." *Procedia Computer Science* 175 (2020): 689-694.