CS 539 Project Process Notebook

Sumanth Pashuparthi

May 2, 2024

1 Overview

The emotion detection project is a multifaceted exploration into the realm of machine learning and human-computer interaction, with the goal of discerning users' emotional states from facial expressions. By delving into facial image data, this project seeks to develop a sophisticated model capable of accurately categorizing emotions such as happiness, sadness, anger, surprise, fear, and disgust. Drawing upon various machine learning techniques and leveraging a dataset specifically curated for emotion detection, the project endeavors to create a robust system capable of real-time analysis. Through meticulous preprocessing, feature extraction, model selection, training, and evaluation processes, the aim is to achieve high accuracy and generalizability across diverse demographics and environmental conditions. The ultimate aspiration is to contribute to the advancement of emotionally intelligent systems, enriching human-computer interaction experiences across a wide range of applications.

2 Summary

This project focuses on developing an emotion detection system using machine learning techniques to analyze facial expressions. The aim is to accurately classify emotions such as happiness, sadness, anger, surprise, fear, and disgust. Leveraging a curated dataset, the project employs various machine learning algorithms for preprocessing, feature extraction, model selection, training, and evaluation. The ultimate goal is to create a robust system capable of real-time emotion analysis with high accuracy and generalizability across diverse demographics and environmental conditions. The project contributes to advancing emotionally intelligent systems, enriching human-computer interaction experiences in various applications.

3 Project Description

3.1 Motivation

This project is motivated by the critical role of emotion detection in improving human-computer interaction and user experience. Understanding users' emotional states can enable systems to respond more effectively and empathetically. Leveraging machine learning techniques, this project aims to develop an emotion detection system capable of accurately interpreting facial expressions, thereby contributing to the advancement of emotionally intelligent systems.

3.2 Aims

1. Develop a robust machine learning model to accurately classify various emotions depicted in facial expressions, including happiness, neutral, sadness, anger, surprise, disgust, and fear.
2. Explore different feature extraction methods to capture nuanced emotional cues effectively and enhance the model's performance.
3. Evaluate the model's accuracy and generalizability using comprehensive training and testing datasets.
4. Investigate potential real-world applications of the developed emotion detection system in diverse domains such as healthcare, education, and human-computer interaction.

3.3 Dataset Description

The primary dataset used for this project is sourced from Kaggle and is titled "Emotion Detection - FER." It comprises 35,685 examples of 48x48 pixel grayscale images of faces, divided into training and testing datasets. The images are categorized based on the emotions depicted in the facial expressions, including happiness, neutral, sadness, anger, surprise, disgust, and fear. This dataset provides a substantial number of images for robust model training and evaluation.

- **Dataset Source:** [Kaggle - Emotion Detection - FER] (https://www.kaggle.com/datasets/ananthu017/emotion-detection-fer/data)
- **Contents:** 35,685 grayscale images of faces categorized based on emotions (happiness, neutral, sadness, anger, surprise, disgust, fear)
- **Size:** The dataset includes a significant number of images, facilitating comprehensive model training and evaluation.

This dataset serves as the foundation for our project, enabling the development of a powerful emotion detection system capable of real-time analysis and fostering advancements in human-computer interaction.

4 Methods

4.1 Data Preprocessing

- The dataset containing 48x48 pixel grayscale images of faces is loaded into memory.
- Preprocessing techniques such as normalization, resizing, and data augmentation are applied to enhance model performance and generalizability.

4.2 Feature Extraction

- Facial feature extraction is a crucial step in emotion detection. Techniques such as histogram of oriented gradients (HOG), local binary patterns (LBP), and deep learning-based feature extraction using pre-trained convolutional neural networks (CNNs) like VGG or ResNet are explored.

4.3 Model Selection

- Various machine learning and deep learning models are considered for emotion detection, including Support Vector Machines (SVM), Random Forests, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their combinations.
- Hyperparameter tuning is conducted to optimize model performance.

4.4 Model Training

- The selected models are trained on the preprocessed data using appropriate training algorithms and optimization techniques.
- Training is performed using both the training and validation datasets to prevent overfitting.

4.5 Model Evaluation

- The trained models are evaluated using the test dataset to assess their performance in accurately classifying emotions.
- Evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix are computed to measure model performance.

4.6 Cross-Validation

- Cross-validation techniques such as k-fold cross-validation are employed to validate model robustness and generalizability.

4.7 Ensemble Methods

- Ensemble learning methods such as bagging, boosting, and stacking are explored to further improve model performance by combining multiple base learners.

4.8 Interpretability Analysis

- Techniques for interpreting model decisions, such as feature importance analysis and visualization of learned representations, are applied to gain insights into how the models classify emotions based on facial expressions.

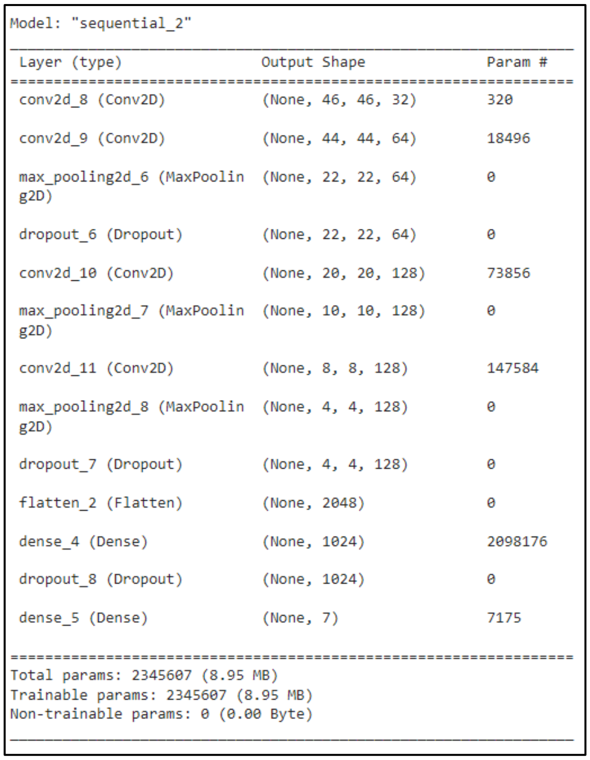
These methods collectively form the foundation of the data analysis process, enabling the development of a robust emotion detection system capable of accurately interpreting facial expressions.

5 Conclusion

In the pursuit of developing an emotion detection system, a variety of machine learning models were explored and evaluated. Each model brought forth unique insights and performance characteristics, contributing to a comprehensive understanding of the task at hand.

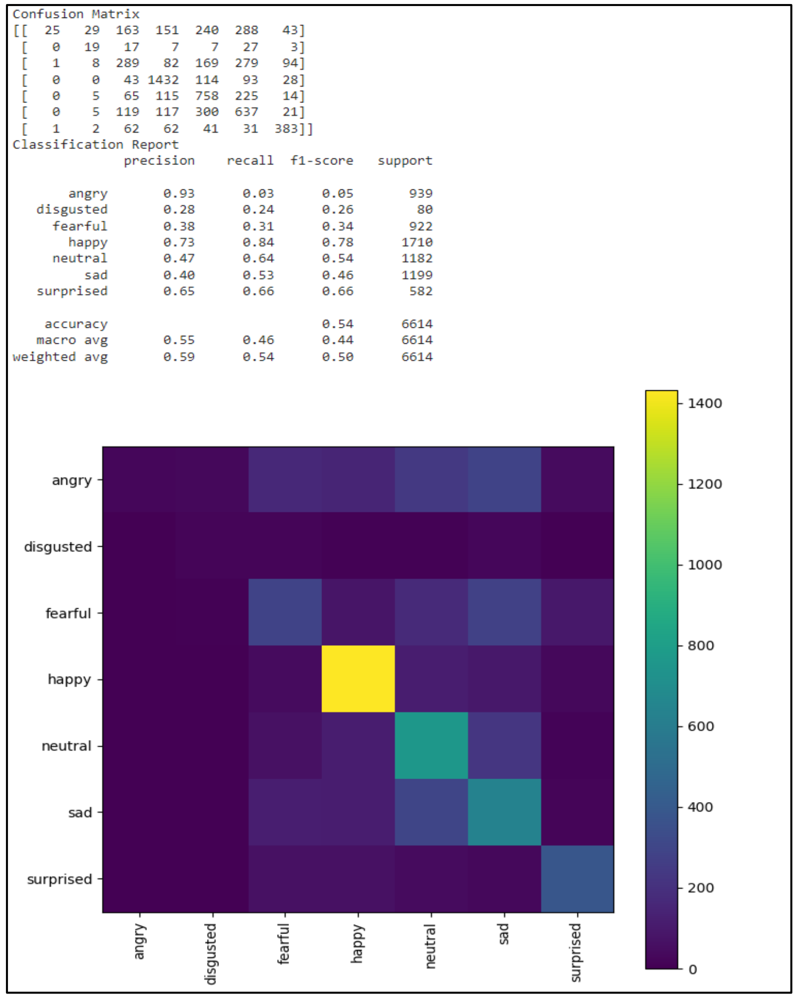
5.1 Key Findings of Emotion Detection using CNNs

- The CNN architecture was custom-designed to effectively capture nuanced emotional cues from gray scale facial images. By incorporating multiple convolutional layers, max-pooling layers, and dense layers, the model learns hierarchical representations of facial features crucial for emotion classification.
- Evaluation of the trained CNN revealed promising results in terms of accuracy and robustness. Performance metrics such as accuracy were monitored closely during training, with validation datasets used to gauge the model's generalization capabilities.
- The CNN demonstrated proficiency in discerning subtle emotional nuances from facial expressions, underscoring its potential for real-world applications.
- Although training the model was a complex and time taking task,model developed using this architecture achieved an accuracy of 60
- Optimizers like Adam, SGD with momentum, RMSprop, and Adagrad are pivotal in training CNN models. Adam's adaptive learning rates and momentum offered faster convergence.
- As gray scale images are used for modelling,I have analysed various image properties such as dynamic range,blurriness,aspect ratio to understand the image quality and experimented with various haarcascades for face detection,$frontalface\_alt$ haarcascades(a publicly availabe opencv module) provided a promising results boosting the model accuracy.



```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_8 (Conv2D)           (None, 46, 46, 32)        320

 conv2d_9 (Conv2D)           (None, 44, 44, 64)        18496

 max_pooling2d_6 (MaxPoolin  (None, 22, 22, 64)        0
 g2D)

 dropout_6 (Dropout)         (None, 22, 22, 64)        0

 conv2d_10 (Conv2D)          (None, 20, 20, 128)       73856

 max_pooling2d_7 (MaxPoolin  (None, 10, 10, 128)       0
 g2D)

 conv2d_11 (Conv2D)          (None, 8, 8, 128)         147584

 max_pooling2d_8 (MaxPoolin  (None, 4, 4, 128)         0
 g2D)

 dropout_7 (Dropout)         (None, 4, 4, 128)         0

 flatten_2 (Flatten)         (None, 2048)              0

 dense_4 (Dense)             (None, 1024)              2098176

 dropout_8 (Dropout)         (None, 1024)              0

 dense_5 (Dense)             (None, 7)                 7175

=================================================================
Total params: 2345607 (8.95 MB)
Trainable params: 2345607 (8.95 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Fig1: Summary of the Model



```
Confusion Matrix
[[  25   29  163  151  240  288   43]
 [   0   19   17    7    7   27    3]
 [   1    8  289   82  169  279   94]
 [   0    0   43 1432  114   93   28]
 [   0    5   65  115  758  225   14]
 [   0    5  119  117  300  637   21]
 [   1    2   62   62   41   31  383]]
Classification Report
              precision    recall  f1-score   support

       angry       0.93      0.03      0.05       939
   disgusted       0.28      0.24      0.26        80
     fearful       0.38      0.31      0.34       922
       happy       0.73      0.84      0.78      1710
     neutral       0.47      0.64      0.54      1182
         sad       0.40      0.53      0.46      1199
   surprised       0.65      0.66      0.66       582

    accuracy                           0.54      6614
   macro avg       0.55      0.46      0.44      6614
weighted avg       0.59      0.54      0.50      6614
```


Fig2: Confusion Matrix and Classification Report

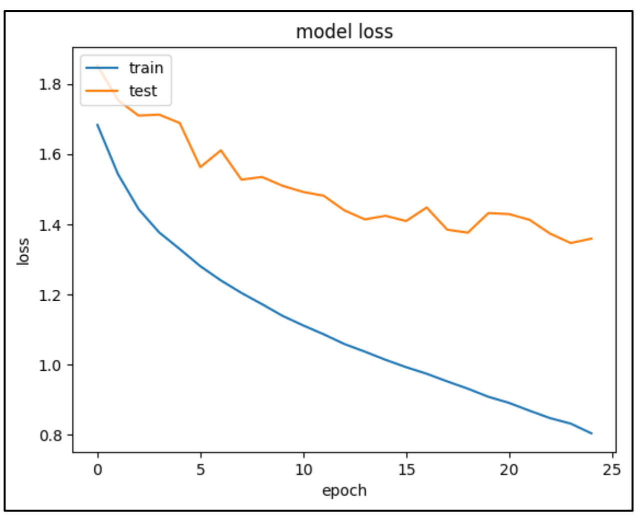


Fig3: Model Loss

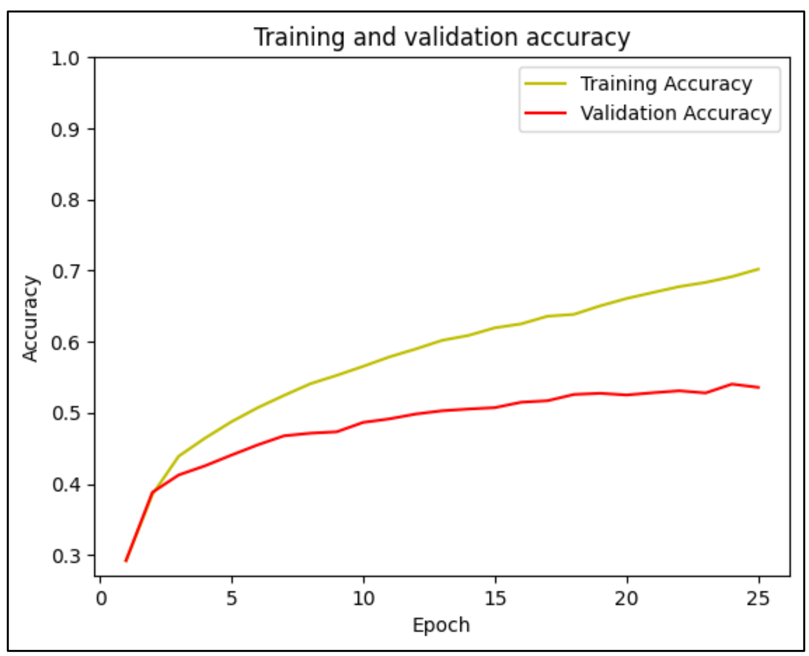
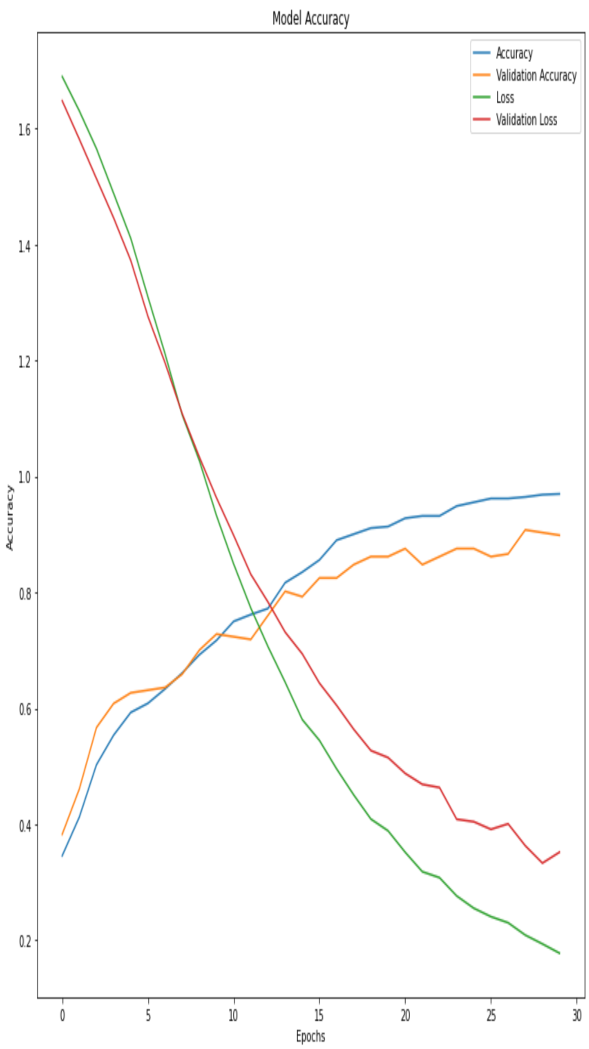


Fig4: Training and Validation accuracy

5.2 Key Findings of Emotion Detection using VGG16

- VGG-16, a deep convolutional neural network architecture, has been employed for emotion recognition by developing 16 layers, including convolutional and pooling layers, enabling robust feature extraction crucial for discerning emotional cues.
- Availability of pre-trained weights facilitates efficient training and deployment, contributing to the model's widespread adoption.
- The hierarchical structure of VGG-16 allows it to capture intricate patterns in facial expressions, body language, and contextual information.
- The model consistently achieved accuracy rates ranging from 75 to 80,demonstrating its efficacy in practical applications.



(a) fig5:Accuracy and Loss plots

input_1 (InputLayer)	(None, 224, 224, 3)	0
conv1_1 (Conv2D)	(None, 224, 224, 64)	1792
conv1_2 (Conv2D)	(None, 224, 224, 64)	36928
pool1 (MaxPooling2D)	(None, 112, 112, 64)	0
conv2_1 (Conv2D)	(None, 112, 112, 128)	73856
conv2_2 (Conv2D)	(None, 112, 112, 128)	147584
pool2 (MaxPooling2D)	(None, 56, 56, 128)	0
conv3_1 (Conv2D)	(None, 56, 56, 256)	295168
conv3_2 (Conv2D)	(None, 56, 56, 256)	590080
conv3_3 (Conv2D)	(None, 56, 56, 256)	590080
pool3 (MaxPooling2D)	(None, 28, 28, 256)	0
conv4_1 (Conv2D)	(None, 28, 28, 512)	1180160
conv4_2 (Conv2D)	(None, 28, 28, 512)	2359808
conv4_3 (Conv2D)	(None, 28, 28, 512)	2359808
pool4 (MaxPooling2D)	(None, 14, 14, 512)	0
conv5_1 (Conv2D)	(None, 14, 14, 512)	2359808
conv5_2 (Conv2D)	(None, 14, 14, 512)	2359808
conv5_3 (Conv2D)	(None, 14, 14, 512)	2359808
pool5 (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc6 (Dense)	(None, 4096)	102764544
fc6/relu (Activation)	(None, 4096)	0
fc7 (Dense)	(None, 4096)	16781312
fc7/relu (Activation)	(None, 4096)	0
fc8 (Dense)	(None, 2622)	10742334
fc8/softmax (Activation)	(None, 2622)	0

Total params: 145,002,878
Trainable params: 145,002,878

(b) fig6:VGG16 model summary

5.3 Key Findings of Emotion Detection using SVM

The SVM model achieved a perfect accuracy score of 100 in emotion detection. The classification report shows that the model achieved a precision, recall, and F1-score of 1.00 for all emotions, indicating that it correctly classified all instances of each emotion. The confusion matrix also confirms that there were no misclassifications.

```
Train Result:
================================================
Accuracy Score: 100.00%
_______________________________________________
CLASSIFICATION REPORT:
              precision    recall  f1-score   support

       anger       1.00      1.00      1.00        88
    contempt       1.00      1.00      1.00        36
     disgust       1.00      1.00      1.00       136
        fear       1.00      1.00      1.00        49
       happy       1.00      1.00      1.00       147
     sadness       1.00      1.00      1.00        58
    surprise       1.00      1.00      1.00       172

    accuracy                           1.00       686
   macro avg       1.00      1.00      1.00       686
weighted avg       1.00      1.00      1.00       686

_______________________________________________
Confusion Matrix:
 [[ 88   0   0   0   0   0   0]
 [  0  36   0   0   0   0   0]
 [  0   0 136   0   0   0   0]
 [  0   0   0  49   0   0   0]
 [  0   0   0   0 147   0   0]
 [  0   0   0   0   0  58   0]
 [  0   0   0   0   0   0 172]]
```

(a) fig7:Training Results

```
Test Result:
================================================
Accuracy Score: 98.98%
_______________________________________________
CLASSIFICATION REPORT:
              precision    recall  f1-score   support

       anger       1.00      1.00      1.00        47
    contempt       1.00      1.00      1.00        18
     disgust       1.00      1.00      1.00        41
        fear       1.00      0.88      0.94        26
       happy       1.00      1.00      1.00        60
     sadness       0.90      1.00      0.95        26
    surprise       1.00      1.00      1.00        77

    accuracy                           0.99       295
   macro avg       0.99      0.98      0.98       295
weighted avg       0.99      0.99      0.99       295

_______________________________________________
Confusion Matrix:
 [[47  0  0  0  0  0  0]
 [ 0 18  0  0  0  0  0]
 [ 0  0 41  0  0  0  0]
 [ 0  0  0 23  0  3  0]
 [ 0  0  0  0 60  0  0]
 [ 0  0  0  0  0 26  0]
 [ 0  0  0  0  0  0 77]]
```

(b) fig8:Test Results

5.4 Key Findings of Emotion Detection using Random Forest

In emotion detection using Random Forest, the model achieves a training accuracy of approximately 50.58 and a validation accuracy of about 49.02. These accuracies indicate that the model's performance is only slightly better than random guessing, suggesting the need for further refinement or exploration of different features and hyperparameters. Despite its limited accuracy, this model still holds potential for identifying emotional cues in textual data, albeit with room for improvement. Additional techniques such as feature engineering, ensemble methods, or fine-tuning hyperparameters could be explored to enhance the model's performance and make it more reliable for real-world applications.

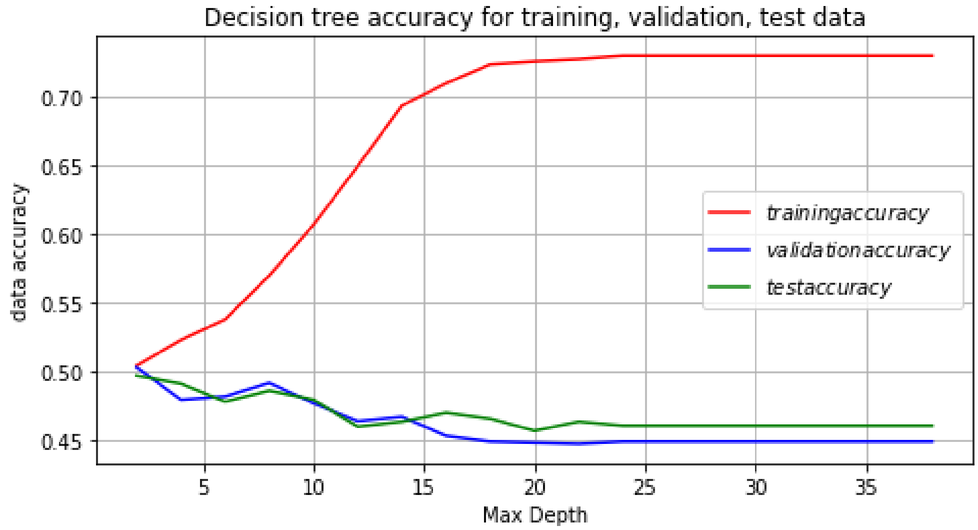


Figure 3: Accuracy plots of Random forest model

- Pre-trained models such DeepFace and CVLIB also shows to be promising results to an extent.