Facial Image Based Emotion Detection and Music Recommendation System

Group No. 14

Abhinay Gupta (2018209)
Akriti Agrawal (MT20021)
Deepankar Kansal (MT20007)
Saloni Gupta (MT20016)
Sudeep Vig (MT20097)

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Problem Statement

- We propose a system which takes human's facial images as input and make predictions of emotions on that and after which it suggests a list of songs which relates to the emotion detected of the input image.

- We try to apply different Information Retrieval techniques in suggesting or recommending the related music from a huge list of songs.

- To better understand the facial expressions we also try to apply a *Computer Vision state-of-the-art* to detect the emotions efficiently and easy to understand techniques.

Dataset Used

We will be using the two datasets for our project :

- The first dataset contains facial images available in Kaggle named as [FER-2013](#) .

- Another one is the songs dataset that we have created ourselves named as [PLAYLIST DATA](#) .
  - Enhanced dataset containing 100 playlists.

Literature Review

We read multiple research papers where different researchers proposed different algorithms and approaches to recommend music based on facial emotion.

Rafael Cabredo Emil Ian V. Ascalon. proposed *Lyrics-Based Music Mood Recognition System*. The authors have focused on mood recognition of OPM songs using lyrics. They have created the Word level features such as *TF-IDF and keyGraph keyword generation* algorithm, using different thresholds and parameters to determine how well these methods worked. The keyGraph feature extraction gives an *80%* average accuracy for music recommendation.

Literature Review

Shlok Gilda, Husain Zafar, Chintan Soni, and Kshitija Waghurdekar proposed Smart music player integrating *facial emotion recognition and music mood recommendation system*.

A music player contains three main modules: *Emotion Module, Music Classification Module, Recommendation Module*.

They worked on four different classes *happy, sad, angry and neutral,* but the dataset used, contains seven emotion classes which leave a future aspect to work on all emotional classes.

Emotion module achieved 72.4% accuracy by applying CNN's on the images. Overall they achieved 72.4% accuracy in Emotion module using OpenCV, 97.69% accuracy in correctly classifying the correct songs of emotions caught.

Literature Review

Sergio Oramas, Luis Espinosa-Anke, Aonghus Lawlor, Xavier Serra, and Horacio Saggion proposed a music information retrieval system from the ***user's reviews on the music and combining metadata and acoustic features related to that song***. They have used Multimodal Album Reviews Dataset (MARD) dataset, which contains Amazon customer reviews and also applied mappings of MusicBrainz (MB) metadata and acoustic features from AcousticBrainz (AB) to the albums. Training to test ratio made as ***80-20% with 5-fold cross validation*** and using Linear SVM with BoW+SEM features noted the highest accuracy of ***69.08%.***

Methodology:
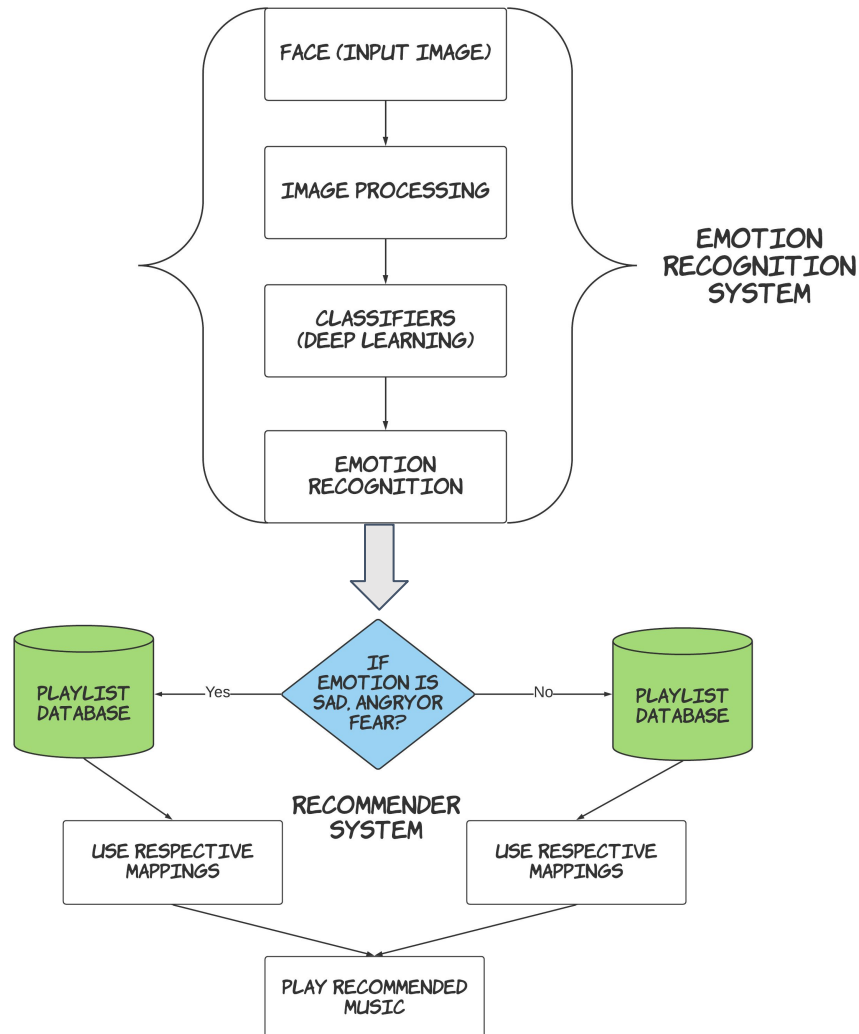
Results (On 7 Classes)

We achieved an overall classification accuracy of ***65.58%*** (on the testing data) on seven emotion classes using ***Convolutional Neural Network (CNN)***.

Other models results:

- VGG-16, VGG-19(accuracy of 57.63% on the testing data and accuracy of 97.53% on training data).
- InceptionV3(accuracy of 33.54% on the testing data and accuracy of 36.12% on training data).
- ResNet50(accuracy of 42.61% on the testing data and accuracy of 94.33% on training data).
- A machine learning model SVM (Support Vector Machine). But neither of them is giving good accuracy, or some models like VGG-16, VGG-19 and ResNet50 overfits the data a lot.
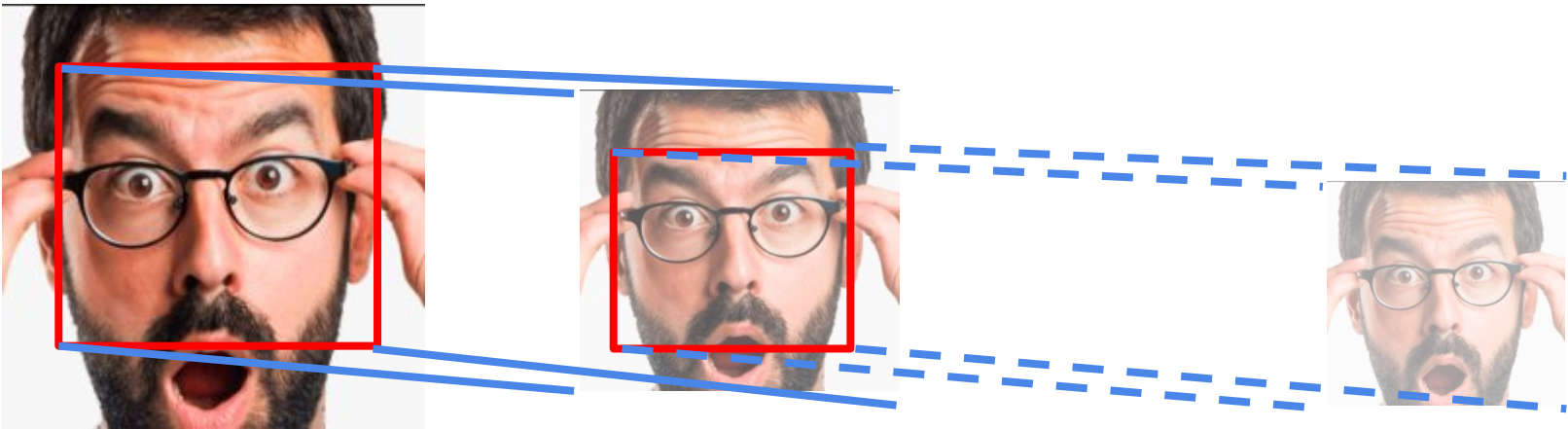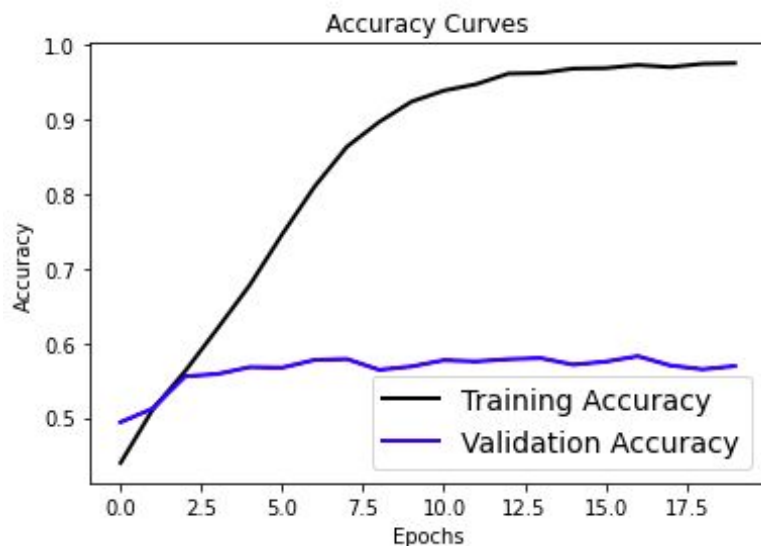
Best Model's Architecture
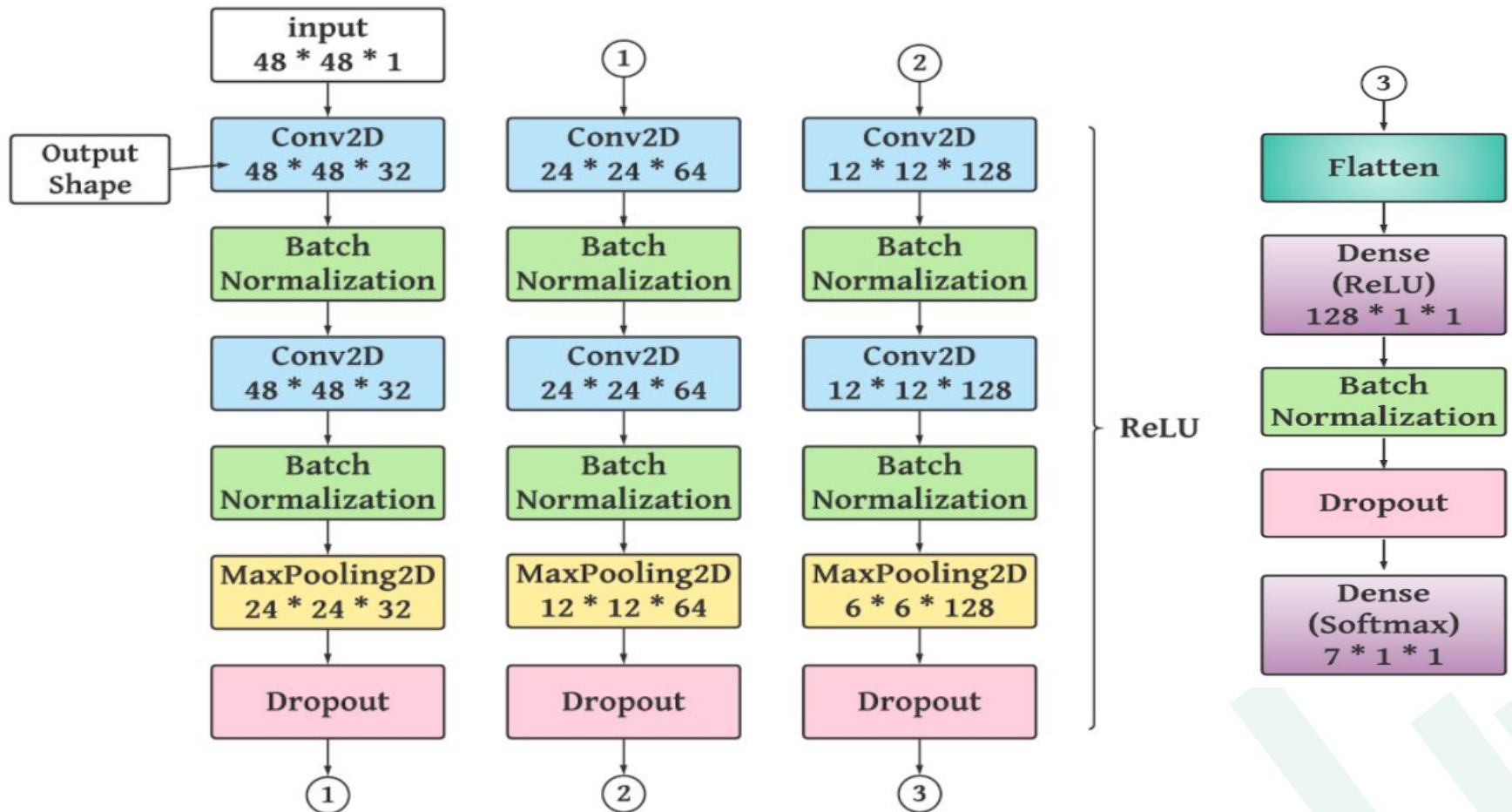
CNN Model Working

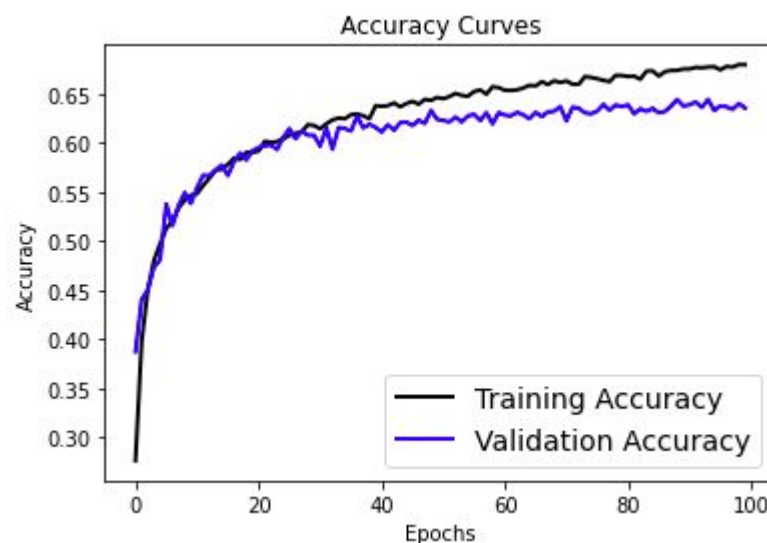Proposed CNN model reduces image size after each pooling layer.

After final pool layer, model flatten() the results and dense layer architecture can be imposed. For compiling 'adam' optimizer with categorical cross-entropy and accuracy metrics is used.
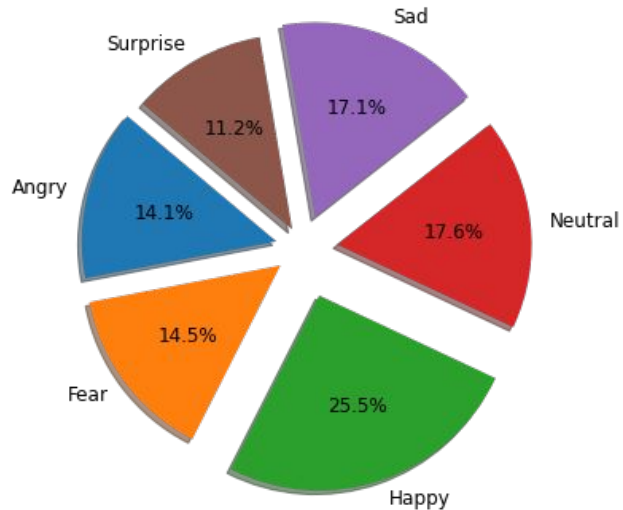
Accuracy plots & Results

VGG-19

CNN Variant

	VGG16	VGG19	InceptionV3	ResNet50	Xception	CNN 1	CNN 2	CNN3	CNN 4	CNN 5
7 Classes	63.29%	57.63%	33.54%	42.61%	38.90%	65.58%	64.07%	61.67%	64.30%	63.06%

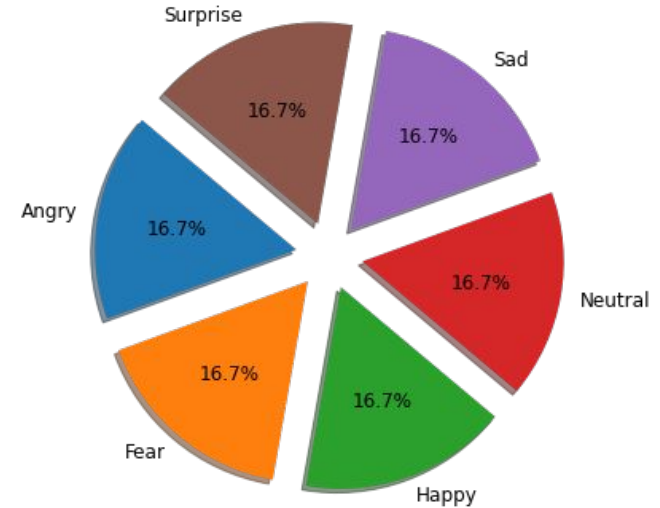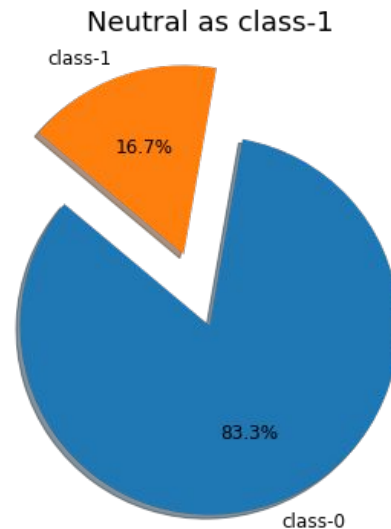| Baselines	Proposed CNN variants

Under Sampling of the Data

Training Data Before Undersample

Training Data After Undersample

Neutral as class-1

Single emotion class proportions for Ensemble Technique.

On 4 Classes and Ensemble Technique

We achieved an overall classification accuracy of ***73.58%*** (on the testing data) on four emotion classes using ***Convolutional Neural Network (CNN)***.

Ensemble Technique

- ***Five class classification***; Happy, Angry, Sad, Surprise, Neutral
- 5 different data labels for each emotion classification. Example: for happy model, happy images labels set to 1 rest are 0.
- After ensembling, got accuracy of ***75.3%*** and ***0.753*** F1-score.

	Happy	Angry	Sad	Neutral	Surprise
Accuracy	92.36%	85.06%	82.15%	85.14%	91.87%

Results on Random Data

Predicted Emotion: **happy**
Intensity Of Emotion: **99.710%**
We suggest you this playlist:
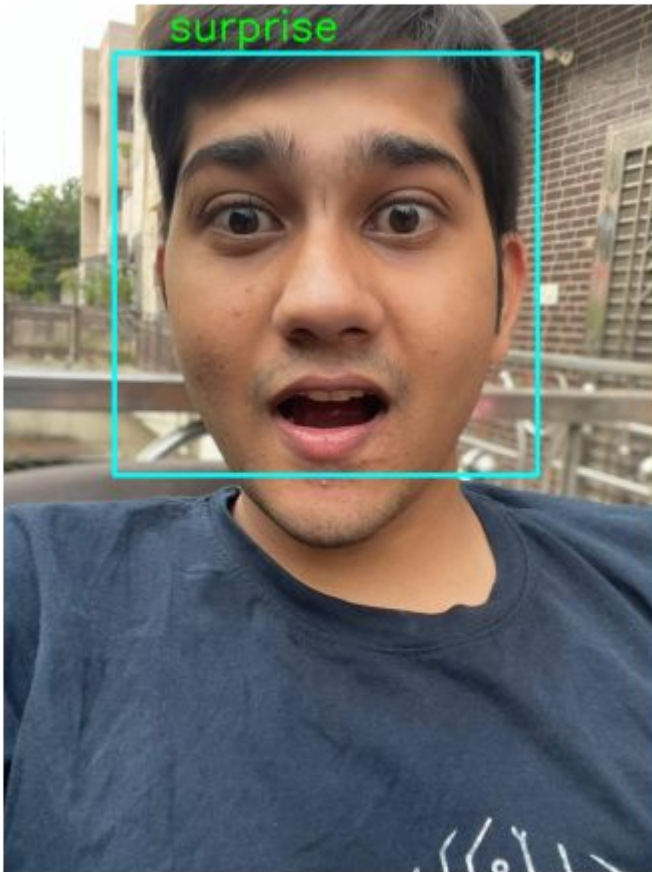https://youtube.com/playlist?list=RDCLAK5uy_nlOMew8qv8HGXb9HbshuU1OgH3aL_JMKA
Genre: EDM
Artist: Multiple
Language: Punjabi

Emotion Recognition and Music Recommendation in action

Results on Random Data

Predicted Emotion: **surprise**
Intensity Of Emotion: **49.497%**
We suggest you this playlist:
https://www.youtube.com/playlist?list=PLa2rOLrAc0awWA1Y3xN5uxqm31ui-49_4
Genre: EDM
Artist: Alan Walker
Language: English

Emotion Recognition and Music Recommendation in action

Midsem Future Aspect

✓ Checking results for images outside the dataset using **OpenCV** modules to detect face from images.

✓ Try to enhance the accuracy of predicting emotions by some machine learning techniques, for example; **oversampling**, etc.

✓ More deep learning models for image detection.

✓ Reduce number of emotion classes for better training.

✓ Enhance music recommendation playlist.

✓ Use intensity of emotion to recommend music.

Limitations

- Emotion Recognition
  - Lower accuracy on unseen data.
  - Baselines worked well if input image size is somewhat large, i.e. around (224 * 224 * 3). But actual image size is (48 * 48 * 1).
  - Try more datasets of facial images.
- Music Recommendation
  - Evaluation metric is not accurate.
  - Hard to compare results with existing system, as all are using their own personalized dataset.
  - Further increase dataset samples, currently it contains 100 playlists.

Future Aspect

- Enhance accuracy on unseen data by training model for specific emotion classes and then tune them in a way, so that model works fine for unseen data.

- Music recommender system by taking feedback from the user in a more constructive way.

Questions?

Contributions

Deepankar Kansal: Literature surveys, CNN-1 variant and Ensemble technique, Under-sampling, report writing, demo file maintenance including OpenCV.

Sudeep Vig: Literature surveys, CNN-2, 3, 4, methodology, maintaining code file, report writing, Playlist dataset creation.

Saloni Gupta: Literature surveys, VGG-16, 19, Xception, Data augmentation, report writing, Playlist dataset creation.

Akriti Agrawal: Literature surveys, InceptionV3, CNN-5, methodology, results, report writing, Playlist dataset creation.

Abhinay Gupta: Literature surveys, ResNet50, evaluation metrics, report writing, Playlist dataset creation.

Thank you