Multimodal Emotion Recognition System

K Supriya
BL.EN.U4EAC22024
Amrita School of Engineering
Electronics and Computer Engineering

Shrinandh N H
BL.EN.U4EAC22037
Amrita School of Engineering
Electronics and Computer Engineering

Smrithi Warrier
BL.EN.U4EAC22050
Amrita School of Engineering
Electronics and Computer Engineering

***Abstract*—This project presents a comprehensive emotion recognition system integrating text, voice, and facial expression analysis to enhance human-computer interaction, healthcare monitoring, and user experience. Utilizing datasets from platforms like Kaggle, RAVDESS, and CREMA-D, we classify emotions such as Angry, Fear, Happy, Neutral, Sad, and Surprise. Facial expressions are analyzed using OpenCV and MediaPipe FaceMesh, while voice features are extracted using Zero Crossing Rate, MFCC, and Mel Spectrograms. Text sentiment analysis leverages TF-IDF and Word2Vec embeddings. Machine learning models, including Random Forest and Gradient Boosting, are trained on these features, with SMOTE addressing class imbalances. The system's real-time prediction capabilities are demonstrated through an interface combining all three modalities. Future work includes incorporating deep learning models and expanding datasets for improved accuracy and generalization.**

Keywords—Emotion Recognition, Machine Learning, Facial Expression Analysis, Voice Analysis, Text Sentiment Analysis, Multi-modal Integration.

I. Introduction (*Heading 1*)

Emotion recognition is a critical component in enhancing human-computer interaction, providing better healthcare monitoring, and improving user experiences across various domains. This project presents a multi-modal emotion recognition system that integrates text, voice, and facial expression analysis. The system classifies emotions into categories such as Angry, Fear, Happy, Neutral, Sad, and Surprise. By leveraging computer vision techniques for face detection and feature extraction, advanced audio analysis for speech emotion detection, and natural language processing for text sentiment analysis, the system aims to deliver a comprehensive solution for real-time emotion detection and classification. Emotion recognition in speech and text involves discerning emotional states from spoken words or written content. Speech analysis considers features like prosody and voice quality, whereas text analysis relies on word choice, sentence structure, and contextual understanding. It utilizes Emotion Datasets, incorporating preprocessing steps and advanced techniques such as part-of-speech tagging. The application of TF-IDF for feature extraction enhances model discernment. Various machine learning models (SVM, Decision Tree, Random Forest, Logistic Regression, K-nearest Neighbour (KNN) Classifier and Convolutional Neural Networks (CNNs)) undergo comprehensive assessment.

Emotion recognition through facial analysis leverages computer vision techniques to identify and classify human emotions based on facial expressions. This process involves capturing images or video streams of faces and using advanced algorithms to detect and interpret facial features. These features are then analysed to determine the underlying emotional state of the person, such as happiness, sadness, anger, or surprise. The extracted data is processed through machine learning models, which are trained to recognize various emotional patterns based on the position and movement of facial landmarks.

II. Related Works

[2] The study uses a neural network model with OpenCV for group facial emotion analysis, achieving a 90.5% accuracy rate. It combines CNN models and OpenCV for precise feature extraction and real-time applications in security and human-computer interaction.

[3] This study presents a sentiment analysis model for job interviews using supervised ML and ANN. It includes eye tracking to analyze participants' eye movements during questions. The model uses an MLP with two hidden layers, a sigmoid function, and input normalization, featuring a client-server model with a web-based interface.

[4] The study improves facial emotion recognition by fusing CNN and LBP, and using transfer learning on datasets like CASIA WebFace and SFEW, achieving a 15% accuracy improvement. It demonstrates the practical use of CNNs in real-time emotion detection.

[5] The paper "Emotion Detection of Contextual Text using Deep Learning" introduces the Aimens system, which uses a Bi-LSTM model with word2vec and doc2vec embeddings for emotion detection in conversations, achieving an F-score of 0.7185. The system enhances understanding of user emotions and human-computer interaction through hyper-parameter tuning and improved preprocessing.

[6] This study develops a real-time algorithm for lie detection based on facial expressions during conversations. It identifies six key facial features linked to lying, such as blinking rate and eyebrow movements, using computer vision and signal processing. The model employs the Viola-Jones algorithm for feature extraction and discusses related works and experimental results in facial lie detection.

III. Dataset Description

A. *Text Emotion Recognition*

The text sentiment dataset comprises labelled text samples representing various emotions, sources from social media, online forums, and curated emotion-labelled datasets.

Emotions dataset for NLP

- Number of files: 3
    - train.txt: [Training Dataset]
    - test.txt: [Testing Dataset]
    - val.txt: [Validation Dataset]
- Total Size: 206.76 kB(training dataset), 1.66MB(testing dataset), 204.24kB(validation dataset)
- Data Format: Text files containing textual data for sentiment analysis
- List of documents with emotion flag, dataset is split into train, test & validation for building the machine learning model

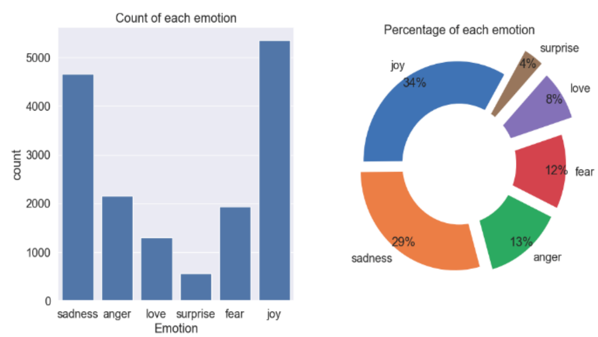

Fig. 1. The count of emotion samples for each category in the training dataset

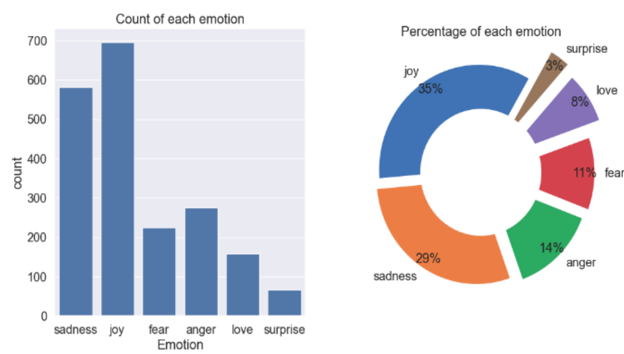

Fig. 2. The count of emotion samples for each category in the testing dataset

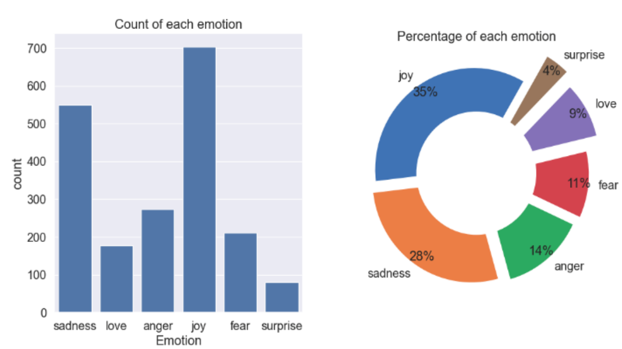

Fig. 3. The count of emotion samples for each category in the validation dataset

B. Voice Emotion Recognition

Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) - [Dataset Link]

- This portion of the RAVDESS contains 1440 files: 60 trials per actor x 24 actors = 1440
- RAVDESS contains 24 professional actors (12 female, 12 male)
- Emotions include calm, happy, sad, angry, fearful, surprise, and disgust.
- The filename consists of a 7-part numerical identifier (e.g., 03-01-06-01-02-01-12.wav). These identifiers define the stimulus characteristics:

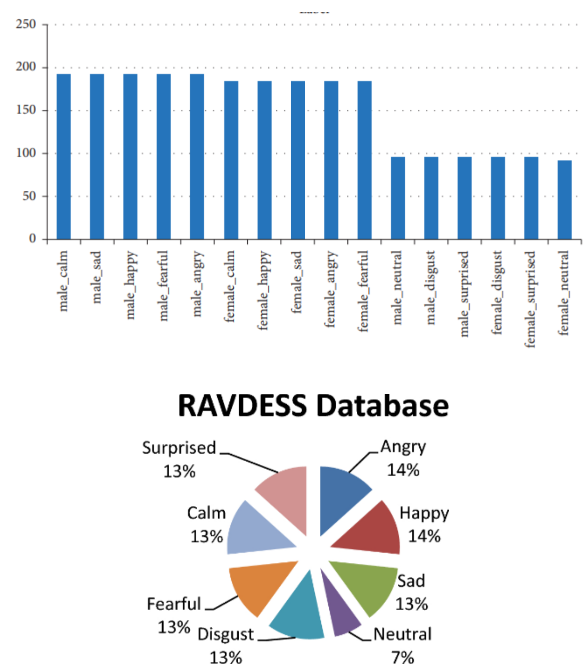

Fig. 4. The count of emotion samples for each category in the dataset

C. Facial Emotion Recognition

FER-2013 - [Dataset Link]

- RAVDESS contains 24 professional actors (12 female, 12 male)
- Emotions include calm, happy, sad, angry, fearful, surprise, and disgust. The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less cantered and occupies about the same amount of space in each image.
- The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).
- The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

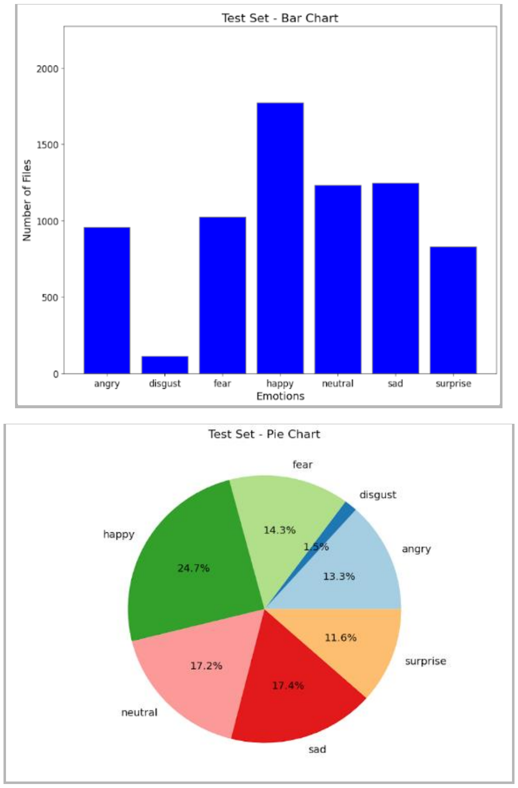

Fig. 5. The count of emotion samples for each category in the testing dataset

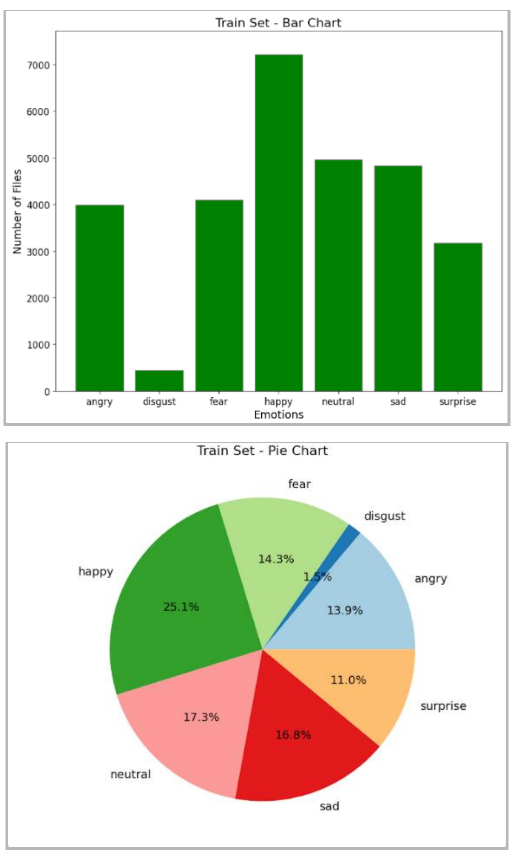

Fig. 6. The count of emotion samples for each category in the training dataset

IV. Model Selection

These models serve as fundamental components in the project's architecture, facilitating the analysis and classification of data across diverse input modalities. Regardless of whether the input is derived from speech, text, or other sources, these models are applied consistently to discern and interpret patterns within the data.

A. *Random Forest*

Random Forest is like a group of decision trees each specialized in recognizing different emotions. When a new text comes in, each tree votes on which emotion it thinks the text expresses. The forest then aggregates the votes to predict the primary emotion based on the majority opinion of all the trees.

B. *Logistic Regression*

Logistic Regression is like drawing a dividing line through a cloud of points, where each point represents a piece of text and the line separates them into different emotion categories. Logistic regression calculates the probability that a piece of text belongs to each emotion category.

C. *Decision Tree*

Decision Tree is like playing a game of multiple-choice questions about emotions. It asks a series of questions based on the features of the text, gradually narrowing down the possibilities until it reaches a conclusion about the primary emotion expressed in the text.

D. *Support Vector Machine (SVM)*

Support Vector Machine is like finding the best hyperplane to separate different emotions in a high-dimensional space of text features. It identifies the optimal boundary that maximizes the margin between different emotion categories, enabling accurate classification of text into the appropriate emotion classes.

E. *K-nearest Neighbor (KNN) Classifier*

K-nearest Neighbor is like consulting with the closest neighbors to determine the primary emotion expressed in a piece of text. It looks at the emotions of similar texts (neighbors) based on their feature representations and assigns the primary emotion based on the majority vote of its k nearest neighbors.

F. *Convolutional Neural Networks (CNNs)*

They are deep learning models tailored for structured grid data like images and audio. With layers of filters, they learn hierarchical features, making them adept at tasks such as image and speech recognition.

V. Methodology

A. *Sentiment Analysis using text*

The goal is to identify emotions in texts, encompassing complex states like happiness, sadness, anger, surprise, fear, etc. The task involves categorizing texts into predefined emotion classes.

1) Data Collection

The data is split into parts for training, testing and validating making sure it's all good quality. The dataset comprises two primary columns: one for textual data and the other for denoting the corresponding emotion associated with each text entry. This structure facilitates the organization and analysis of the data, enabling the development and evaluation of machine learning models tailored to emotion recognition tasks. Each row in the dataset pairs a piece of text, such as a sentence or phrase, with an emotion label, providing a clear and structured framework for training and validation processes.

2) Data Preprocessing

Data preprocessing is vital in the analysis and machine learning pipeline, aiming to enhance data quality, address issues like missing values or outliers, and ensure data is in a format suitable for algorithms.Some of the data preprocessing steps conducted include:

a) Convert to lowercase - Ensures consistency and eliminates the impact of letter case on subsequent processing.
b) Remove Special characters - Eliminates non-informative elements, focusing on the actual content and enhancing text cleanliness.
c) Tokenize the text- Breaks down the text into individual words or tokens, facilitating granular analysis and under- standing of the text.
d) Apply Part-Of-Speech tagging - Labels each token with its grammatical category (e.g., noun, verb), which is useful for understanding syntactic structure and capturing word relationships.
e) Remove stop words - Eliminates common words (e.g.," the"," and") that don't carry significant meaning, enhancing focus on more meaningful words for emotion classification.
f) Perform lemmatization - Lemmatization considers con- text, reducing dimensionality and capturing core word meanings.
g) Reassemble the text - After preprocessing tokens, re-assembles text to maintain structure, ensuring modifications during preprocessing do not disrupt the text's overall coherence.

3) Feature Extraction

TF-IDF (Term Frequency-Inverse Document Frequency) vectorization is essential for converting processed text into numerical features, aligning with machine learning model requirements. TF-IDF combines "Term Frequency" (TF) to measure term occurrence and "Inverse Document Frequency" (IDF) to assess a term's global importance, resulting in a weighted numerical representation. This reduction technique improves model performance, reduces computational ineffi-ciencies, and optimizes textual data for machine learning applications by capturing term importance both locally and globally.

4) Model Training

In emotion model selection, various machine learning and ensemble models, including SVM, Naive Bayes, Decision Tree, Random Forest, XGBoost, LightGBM, AdaBoost, Gradient Boosting, Bagging, and Ensemble Classifier, are considered for a thorough exploration of algorithmic approaches in emotion recognition. The goal is to identify the most suitable model or ensemble through rigorous evaluation, aligning with the complexities of emotion classification and establishing a robust foundation.

a) Random Forest
b) Logistic Regression
c) Decision Tree
d) Support Vector Machine
e) K-neighbor Classifier

5) Evaluation

Following training on the complete dataset, evaluating the model's performance on the test set provides crucial insights into its generalization ability and effectiveness across different data subsets. Metrics such as accuracy and detailed classification reports, including precision, recall, F1-score, support and confusion matrix offer nuanced assessments of the model's predictive capabilities for each emotion category. By comparing metrics across different models, we can gauge their performance in recognizing various emotions. The model exhibiting the highest overall accuracy and balanced precision, recall, and F1-scores signifies superior performance and suitability for real-world emotion recognition tasks. This evaluation process guides the selection of the best-performing model, ensuring optimal outcomes and informed decision-making in deploying emotion recognition systems.

6) Predictions/Performance

After training, the model is put to the test by using it to guess emotions in fresh text. This step is crucial because it helps us understand how well the model performs in real-life situations. We carefully examine its strengths and weaknesses, making changes to improve its accuracy. By going through this process several times, we end up with a reliable model that can accurately identify emotions like happiness, sadness, anger, or fear in text.

B. *Sentiment Analysis using speech*

1) Data Collection

The RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) dataset. This dataset provides a comprehensive collection of audio samples that are essential for sentiment analysis. The dataset includes audio recordings of actors performing emotional speech, with each recording labeled according to the emotion expressed as well as music files. The labeled emotions include a wide range of categories that are critical for training the models effectively.

2) Data Preprocessing

Data preprocessing is a crucial step to ensure that the audio samples are in the optimal format for analysis. The preprocessing steps include: MFCC feature (Mel-frequency cepstral coefficients): MFCCs are extracted from the audio files. MFCC is a feature widely used in speech recognition tasks, representing the short-term power spectrum of sound and capturing the essential characteristics of the audio signals.Audio Trim Technique: To ensure consistency, audio files are trimmed to a uniform length. This step is important to standardize the input data, as varying lengths of audio samples can introduce noise and affect the performance of the models.One-Hot Encoding: The categorical emotion labels are converted into a numerical format using one-hot encoding. One-hot encoding transforms the categorical labels into binary vectors, which is essential for machine learning algorithms to process the data effectively.

3) Feature Extraction

Feature extraction involves normalizing the extracted features to improve model performance. The steps include:

a) Standard Scaler:
Standard scaling is applied to the extracted MFCC features. This technique standardizes the features by removing the mean and scaling to unit variance, ensuring that the features are on a similar scale and improving the efficiency of the learning algorithms.

b) Z-Score Normalization:
Each feature is standardized using its z-score, where the mean and standard deviation are calculated.

c) Fitting and Transforming:
The Standard Scaler from scikit-learn is used to fit the scaler on the training data and transform both the training and test data to ensure consistent scaling. This involves the following steps: Fitting: The scaler is fitted on the training dataset to compute the mean and standard deviation for each feature. Transforming: The computed mean and standard deviation are used to transform the training and test datasets, standardizing the features based on the training data statistics.

4) Model Training

a) CNN Model
b) Random Forest (RF)
c) K-Nearest Neighbors (KNN)
d) Decision Tree (DT)
e) Support Vector Machine (SVM)

5) Evaluation

The performance of the audio emotion recognition models was evaluated using a variety of metrics to ensure a comprehensive assessment. These metrics help in understanding how well the models perform in classifying emotions from audio signals.

a) Accuracy: For audio signals, accuracy measures the ratio of correctly predicted emotion instances to the total instances in the dataset.
b) Precision: High precision indicates that the model has a low false positive rate, meaning it rarely misclassified other emotions as the target emotion.
c) Recall: Recall, or sensitivity, measures the model's ability to identify all positive instances of a particular emotion in audio signals.
d) F1-Score: For audio signals, it is particularly useful when the dataset is imbalanced, as it ensures that both precision and recall are considered in the model's performance evaluation.
e) Confusion Matrix: The confusion matrix is a comprehensive tool used to evaluate the performance of the classification model on audio signals. This helps in understanding specific areas of improvement, such as which emotions are being confused with others.

By utilizing these metrics, we can assess the effectiveness of the emotion recognition models in classifying emotions from audio signals, ensuring a robust and reliable performance analysis.

6) Predictions/Performance

The predictions and performance of the models in this project are determined by their ability to classify emotions from audio signals using features like Mel-frequency cepstral coefficients (MFCCs). Convolutional Neural Networks (CNNs) excel in capturing intricate patterns in MFCC features, making them highly effective for audio data. Random Forest (RF) models combine multiple decision trees to enhance prediction accuracy, while k-Nearest Neighbors (KNN) relies on the similarity between audio samples for classification. Decision Trees (DT) offer an intuitive approach, and Support Vector Machines (SVM) excel in high-dimensional spaces. Each model brings unique strengths to emotion recognition in audio data.

C. *Sentiment Analysis using Facial Recognition*

The goal of this project is to develop an emotion recognition system that accurately classifies human emotions using facial landmarks detected from images. By leveraging data augmentation, feature extraction, and machine learning models, the system aims to improve the robustness and accuracy of emotion classification.

1) Data Collection

The dataset consists of images categorized by emotion labels. Each image is manually labeled with one of the six emotions. The images are stored in a structured directory with each emotion having its folder containing respective images.

Haar Cascades are employed to detect faces and eyes within the images. The system uses pre-trained models provided by OpenCV for face (haarcascade_frontalface_default.xml) and eye detection (haarcascade_eye.xml). Images with at least two detected eyes are selected for further processing, ensuring the accuracy of face localization.

2) Data Preprocessing

Data preprocessing is a crucial step in the machine learning process. It involves improving the quality of the data, dealing with problems like missing values or outliers, and ensuring the data is in the right format for the algorithms. Here are some of the data preprocessing steps we carried out:

a) Cropping and Resizing-Images are cropped around the detected faces to focus on the relevant facial region. Each cropped image is resized to a standard dimension (48x48 pixels) to maintain consistency across the dataset.
b) To enhance the dataset's diversity and robustness, various augmentation library. These include horizontal flipping, random cropping, Gaussian blur, contrast adjustment, additive Gaussian noise, brightness modification, and affine transformations. Each image undergoes a series of random augmentations to create multiple variants, significantly increasing the training dataset's size.

3) Feature Extraction

The Mediapipe Face Mesh model is used to extract 468 facial landmarks from each image. These landmarks provide a detailed representation of the face's geometric structure. For each detected face, the 3D coordinates (x, y, z) of the 468 landmarks are flattened into a single feature vector. This vector includes all landmark points concatenated together, providing a comprehensive representation of facial geometry.

4) Model Training

The dataset is divided into training and testing sets. Each set includes feature vectors and corresponding emotion labels. Synthetic Minority Over-sampling Technique (SMOTE) is applied to the training set to address class imbalances. SMOTE generates synthetic examples for underrepresented classes, ensuring a balanced distribution of emotions.

Several machine learning models are evaluated:

a) Random Forest
b) Logistic Regression
c) Decision Tree
d) Support Vector Machine
e) K-neighbor Classifier
f) CNN
g) Gradient Boosting

Model parameters are optimized to improve performance. This includes adjusting the number of estimators in ensemble models, the regularization parameters in SVM and Logistic Regression, and the depth of Decision Trees.

5) Evaluation

Evaluation is the process of assessing the performance of trained machine learning models on a separate test dataset that was not used during training. This step ensures that the model generalizes well to new, unseen data.Models are evaluated using standard metrics: accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of the models' classification performance.The Random Forest and Gradient Boosting models demonstrate superior performance, with the highest F1-scores. These models are selected for further use based on their robustness and accuracy.A confusion matrix is generated for the best model to analyze the classification performance across different emotions. This helps identify any specific emotions that are challenging to classify.

6) Predictions/Performance

For real-time predictions, an ensemble approach combining Random Forest and Gradient Boosting is used. The predictions from both models are averaged to determine the final emotion label.A webcam-based system captures live video frames. Each frame is processed to detect facial landmarks, which are then passed through the trained models for emotion prediction.The predicted emotion is displayed on the video feed, providing real-time feedback to the user. This system runs continuously, updating the prediction as new frames are captured.The real-time system is tested for its responsiveness and accuracy. It successfully identifies emotions in various lighting conditions and facial expressions, demonstrating its practical applicability.

VI. RESULT

A. *Sentiment Analysis using text*

A variety of classifiers, including SVM, Logistic Regression, Decision Tree, Random Forest and K-nearest Neighbour(KNN) underwent training for emotion recognition, with summarized test results in Table 1. Random Forest achieved the highest accuracy (0.889). Logistic Regression also performed well, attaining an accuracy of 0.8685. KNN exhibited lower ac- curacy (0.792), indicating potential limitations.

Sl.no	Model	Accuracy
1	Random Forest	0.8890
2	Logistic Regression	0.8685
3	Decision Tree	0.8585
4	Support Vector Machine (SVM)	0.8055
5	K-nearest Neighbour (KNN) Classifier	0.7925

Table 1. The Accuracy of each model

	F1 score		F1 score
sadness	0.761221	sadness	0.858223
anger	0.762125	anger	0.823810
love	0.830137	love	0.893048
surprise	0.626761	surprise	0.703704
fear	0.835334	fear	0.915683
joy	0.520000	joy	0.626263

Table 2. F1 Scores of each emotion in Logistic Regression & K- Nearest Neighbor models

	F1 score
sadness	0.867857
anger	0.810811
love	0.889213
surprise	0.756098
fear	0.897678
joy	0.601504

	F1 score
sadness	0.734607
anger	0.737913
love	0.831611
surprise	0.544643
fear	0.891156
joy	0.527473

Table 3. F1 Scores of each emotion in Decision Tree & Support Vector Machine

	F1 score
sadness	0.899452
anger	0.864629
love	0.910238
surprise	0.744966
fear	0.935652
joy	0.611570

Table 4. F1 Scores of each emotion in Random Forest

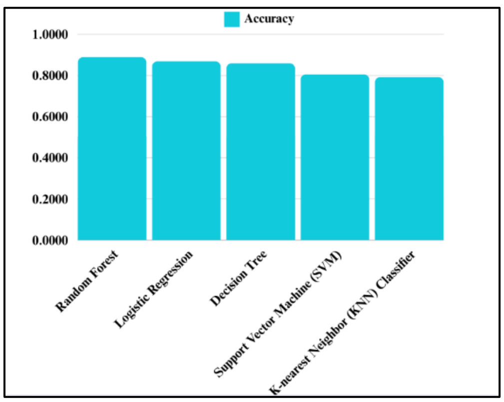

Fig. 7. The Accuracy of each model in a histogram

B. *Sentiment Analysis using speech*

The Convolutional Neural Network (CNN) stands out with the highest accuracy of 69%, showcasing its aptitude in capturing intricate patterns within audio data for emotion recognition. While other models like Random Forest (RF), k-Nearest Neighbors (KNN), Decision Trees (DT), and Support Vector Machines (SVM) offer varying accuracies, CNN's superior performance underscores its suitability for this task.

Classification Report:

	precision	recall	f1-score	support
0	0.735294	0.657895	0.694444	38.000000
1	0.720930	0.584906	0.645833	53.000000
2	0.767442	0.600000	0.673469	55.000000
3	0.661290	0.788462	0.719298	52.000000
4	0.740741	0.476190	0.579710	42.000000
5	0.441176	0.600000	0.508475	25.000000
6	0.478261	0.687500	0.564103	48.000000
7	0.687500	0.702128	0.694737	47.000000
accuracy	0.641667	0.641667	0.641667	0.641667
macro avg	0.654079	0.637135	0.635009	360.000000
weighted avg	0.667101	0.641667	0.644032	360.000000

Fig. 8. Classification Report of CNN

Sl.no	Model	Accuracy
1	Random Forest	0.62
2	CNN	0.69
3	Decision Tree	0.44
4	Support Vector Machine (SVM)	0.46
5	K-nearest Neighbour (KNN) Classifier	0.38

Table 5. The Accuracy of each model

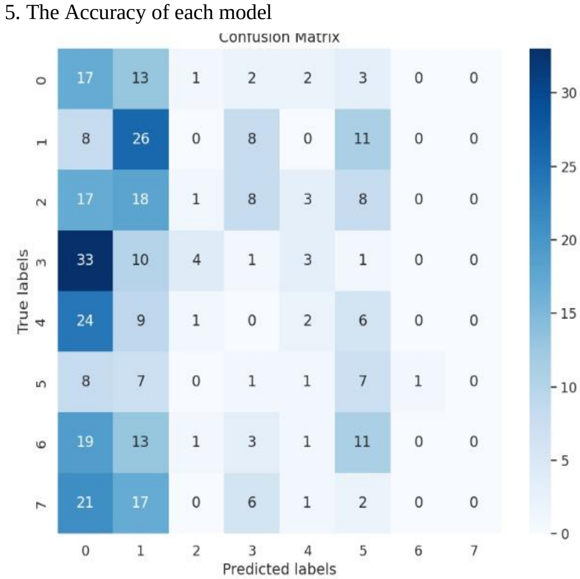

Fig. 9. Confusion matrix of each emotion in a CNN Model

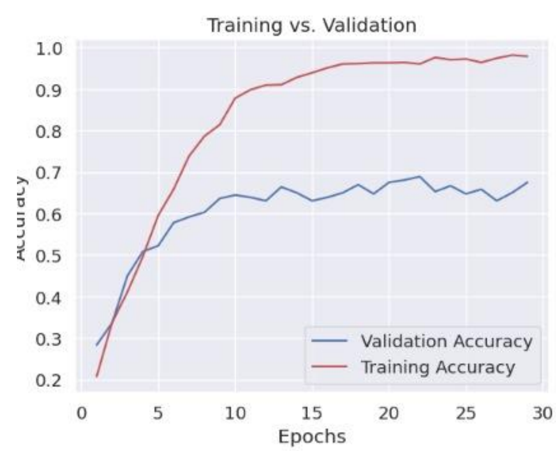

Fig. 10. Epochs for Training and Validation Data

C. *Sentiment Analysis using facial recognition*

The results of our emotion recognition system demonstrate a high level of accuracy(55%) and F1-score(0.55), particularly for the Random Forest model. This model significantly outperformed others, indicating its robustness in classifying emotions from facial landmarks. In contrast, the Convolutional Neural Network (CNN) model yielded the lowest accuracy (20.73%) and F1-score(0.20), suggesting that it was less effective in this specific task. The superior performance of the Random Forest model highlights its suitability for this application, making it a reliable choice for emotion recognition tasks.

Sr. No.	Model	Accuracy
1	Random Forest	0.5500
2	Logistic Regression	0.4132
3	Decision Tree	0.3725
4	Support Vector Machine	0.4311
5	K-nearest Neighbour	0.4673
6	CNN	0.4451
7	Gradient Boosting	0.2073

Table 6. The Accuracy of each model

Sr. No.	Model	F1-Score
1	Random Forest	0.55
2	Logistic Regression	0.40
3	Decision Tree	0.39
4	Support Vector Machine	0.44
5	K-nearest Neighbour	0.47
6	CNN	0.07
7	Gradient Boosting	0.20

Table 7. The F1-Score of each model

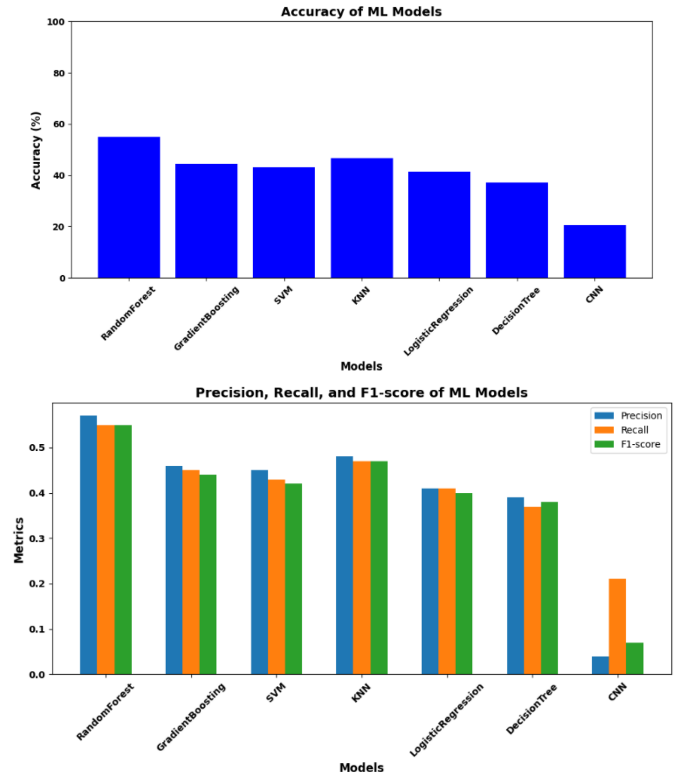

Fig. 11. The evaluation metrics of each model in a histogram

VII. Conclusion and Future Scope

The multi-modal emotion recognition system represents a significant advancement in understanding human emotions, with promising implications for personalized recommendation systems across diverse media platforms. By integrating facial expressions, voice analysis, and text sentiment, this system can discern users' emotional states more accurately, offering tailored recommendations for music, movies, and books based on their emotional responses.

Looking ahead, incorporating advanced deep learning models presents an avenue for further enhancing accuracy. Expanding datasets will improve model generalization, enabling the system to recognize a wider range of emotional nuances. Optimizing real-time processing will enhance user experiences by delivering timely and relevant recommendations. Extending the system's capabilities to develop recommendation systems for various media types will provide users with personalized experiences, enriching their interactions with digital content.

The future scope of this system lies in leveraging cutting-edge technology to refine emotion recognition, expand dataset diversity, optimize processing speed, and extend personalized recommendations across different media platforms, thereby revolutionizing how users engage with digital content.

VIII. ACKNOWLEDGMENT

We thank our faculty, Dr. Manju Venugopalan , for her guidance, encouragement, and co-operation throughout the completion of this project

IX. REFERENCES

[1] Zeeshan Farhan Hanif, Sohelkhan Samirkhan Pathan, Talha Sameer Sayyed, Minhaj Khan, “Emotional Analysis using OpenCV”, International Journal of Novel Research and Development 2013

[2] Kaviya P, Arumuga Prakash T, “Group Facial Emotion Analysis System Using Convolutional Neural Network”, Fourth International Conference on Trends in Electronics and Informatics (ICOEI 2020)

[3] Julio Martínez Zárate, Sandra Mateus Santiago, “Sentiment Analysis Through Machine Learning for the Support on Decision-Making in Job Interviews”, C. Stephanidis (Ed.): HCII 2019, LNCS 11786, pp. 202–213, 2019

[4] Dan Duncan, Gautam Shine, Chris English, “Facial Emotion Recognition in Real Time”

[5] K. Elissa, “Title of paper if known,” unpublished.

[6] Umar Rashid; Muhammad Waseem Iqbal; Emotion Detection of Contextual Text using Deep learning.International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)

[7] Minh-Triet Tran-Le; Anh-Tu Doan Lie Detection by Facial Expressions in Real Time 2021 International Conference on Decision Aid Sciences and Application (DASA)

[8] Shreshth Saxena, Lauren K. Fink, Elke B. Lange, “ Deep learning models for webcam eye tracking in online experiment”, 3 July 2023

[9] Acheampong, F.A., Nunoo-Mensah, H. and Chen, W., 2021, December. Recognizing emotions from texts using an ensemble of transformer- based language models. In 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWA TIP) (pp. 161-164). IEEE. Journal of Artificial Intelligence and Systems, 2(1), pp.53-79.

[10] Saxena, A., Khanna, A. and Gupta, D., 2020. Emotion recognition and detection methods: A comprehensive survey

.