TALKSPACE - ADVANCED EMOTION RECOGNITION AND PERSONALITY ANALYSIS PLATFORM

Submitted in partial fulfillment of the requirements

of the degree of

Bachelor of Engineering in Information Technology

By

ROHAN KADU (Roll No. 20101A0023)

KRISHNAKANT GANGURDE (Roll No. 20101A0016)

VAISHNAV RAJPUT (Roll No. 20101A0068)

Under the Guidance of

Prof. SAMUEL JACOB

Department of Information Technology Engineering

VIT Vidyalankar Institute of Technology
ACCREDITED A+ BY NAAC
Autonomous Institute affiliated University of Mumbai
Vidyalankar Institute of Technology
Wadala(E), Mumbai-400437

University of Mumbai

2023-24

CERTIFICATE OF APPROVAL

This is to certify that the project entitled

"TalkSpace - Advanced Emotion Recognition and Personality Analysis Platform"

is a bonafide work of

ROHAN KADU (Roll No. 20101A0023)

KRISHNAKANT GANGURDE (Roll No. 20101A0016)

VAISHNAV RAJPUT (Roll No. 20101A0068)

submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the

degree of

Undergraduate in **"INFORMATION TECHNOLOGY"**.

Prof. Samuel Jacob	Dr. Vipul Dalal	Dr. S. A. Patekar
Guide	Head of Department	Principal

Project Report Approval for B. E.

This project report entitled ***TalkSpace*** by

1. ***Rohan Kadu (20101A0023)***
2. ***Krishnakant Gangurde (20101A0016)***
3. ***Vaishanv Rajput (20101A0068)***

is approved for the degree of ***Bachelor of Engineering in Information Technology.***

Examiners

1.----------

2.----------

Date:

Place:

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

	Name of student	Roll No.	Signature
1.	Rohan Kadu	20101A0023	
2.	Krishnakant Gangurde	20101A0016	
3.	Vaishnav Rajput	20101A0068	

Date:

ACKNOWLEDGEMENT

We express our profound gratitude and sincere thanks to **PROF. SAMUEL JACOB** our respectable project guide, for her gigantic support and guidance. Without her counselling our project would not have seen the light of the day.

We extend our sincere thanks to **DR. VIPUL DALAL**, Head of the Department of Information Technology Engineering and **PROF. KANCHAN DHURI**, Project Coordinator, for offering valuable advice at every stage of this undertaking. We would like to thank all the staff members who willingly helped us. We are grateful to **VIDYALANKAR INSTITUTE OF TECHNOLOGY** for giving us this opportunity.

We would also like to thank our staff members and lab assistant for permitting us to use computer in the lab as when required. We thank our college for providing us with excellent facilities that helped us to complete and present this project.

1. Rohan Kadu
2. Krishnakant Gangurde
3. Vaishnav Rajput

ABSTRACT

TalkSpace stands as a groundbreaking platform at the forefront of emotional recognition technology, employing sophisticated Deep Learning and Machine Learning models to analyze emotions through multiple modalities such as facial expressions, audio cues, and textual inputs. This revolutionary system offers an unprecedented level of accuracy in identifying and understanding various emotional states in individuals, representing a significant advancement in our understanding of human emotions. By providing real-time emotion analysis and contextual understanding, TalkSpace has the potential to revolutionize communication and interaction across a wide range of domains, including mental health, customer service, and education. In the realm of mental health, therapists can utilize TalkSpace to gain invaluable insights into their clients' emotional states, enabling them to tailor treatment plans more effectively. In customer service, TalkSpace can enhance user experiences by detecting customer emotions and adapting responses accordingly, leading to more personalized and effective interactions. Moreover, in education, TalkSpace can assist educators in understanding students' emotional engagement with learning materials, allowing for the optimization of teaching methods to better meet individual needs. With its comprehensive suite of capabilities, TalkSpace aims to transform how emotions are perceived, understood, and addressed in diverse contexts, ultimately leading to improved human-computer interaction and emotional well-being .

Table of Contents

List Of Figures

List of Tables

1 Introduction

1.1 Introduction

TalkSpace stands at the forefront of innovation in emotion recognition and personality trait classification, driven by a fusion of cutting-edge technologies. Through the seamless integration of text mining, signal processing, and computer vision techniques, TalkSpace provides a pioneering solution for understanding and interpreting human emotions in a variety of contexts.

At its essence, TalkSpace represents a transformative approach to emotional analysis, leveraging advanced methodologies to decode the intricacies of human behavior. By employing text mining techniques, TalkSpace delves deep into textual inputs, extracting valuable insights into individuals' personality traits. Through the analysis of language patterns and textual cues, the system accurately classifies personality traits, offering profound understandings of behavioural tendencies and characteristics. This enables TalkSpace to provide personalized and tailored insights into individuals' emotional makeup.

Additionally, TalkSpace harnesses sophisticated signal processing algorithms to uncover emotional cues embedded within audio signals. Through precise analysis of voice intonations, speech patterns, and acoustic features, TalkSpace achieves nuanced emotion recognition, enabling the detection of subtle emotional nuances in spoken communication. This capability allows TalkSpace to interpret emotional states with a high degree of accuracy, even in complex or ambiguous contexts.

In parallel, TalkSpace utilizes the power of computer vision for emotion recognition by analysing facial expressions to decipher underlying emotions. Leveraging advanced image processing techniques, the system identifies key facial features and dynamics, allowing for real-time assessment of emotional states with remarkable precision. This capability enables TalkSpace to provide users with immediate feedback on their emotional expressions, facilitating deeper self-awareness and understanding.

By seamlessly integrating these state-of-the-art technologies, TalkSpace transcends traditional approaches to emotion recognition, offering a multifaceted and holistic understanding of human emotions. Whether used in mental health assessments, customer service interactions, or educational settings, TalkSpace empowers users to interpret and respond to emotional cues effectively, fostering improved communication and interaction across various domains. With its innovative approach and comprehensive capabilities, TalkSpace is poised to revolutionize the way we understand and interact with emotions in the digital age.

1.2 Problem Statement

Understanding and interpreting human emotions and personality traits remains a challenging task despite technological advancements. Traditional methods for emotion recognition and personality trait classification often rely on subjective assessments or manual analyses, leading to inaccuracies and inefficiencies in domains like mental health, customer service, and education. Moreover, existing solutions tend to focus on single modalities such as text analysis or facial recognition, disregarding the holistic nature of human communication which involves verbal, non-verbal, and contextual cues. This fragmented approach hampers accurate interpretation of complex emotions, limiting effective communication and interaction. There's a critical need for a unified framework leveraging multiple modalities like text mining, signal processing, and computer vision to enable accurate and holistic emotion recognition and personality trait classification, enhancing our understanding of human emotions and facilitating the development of more effective communication strategies and personalized services.

1.3 Scope

In terms of objectives, the scope encompasses assessing user engagement and strategies for user retention, the quality of mental health resources offered within the app, and the analysis of user feedback to identify areas for improvement. Clinical validation and competitive analysis are integral parts of this evaluation. The scope extends to examining accessibility measures, ensuring inclusivity for users with varying needs. Additionally, technological aspects, the scope of this project involves the creation of TalkSpace, an integrated platform for advanced emotion recognition and personality analysis. Utilizing text mining, signal processing, and computer vision techniques, TalkSpace aims to accurately interpret human emotions and personality traits in diverse contexts. It will address the limitations of existing methods by taking a holistic approach that considers verbal, non-verbal, and contextual cues in communication.

TalkSpace will consist of modules for text, audio, and image analysis. The text analysis module will extract insights from language patterns and textual cues to classify personality traits. Meanwhile, the audio analysis module will decode emotional cues within audio signals, including voice intonations and speech patterns. Simultaneously, the image analysis module will analyze facial expressions to discern underlying emotions. This platform will

find applications across mental health assessments, customer service interactions, and educational settings, providing users with accurate insights into emotional states and personality traits. With its comprehensive approach, TalkSpace aims to revolutionize our understanding and handling of human emotions, facilitating improved communication and interaction across diverse domains.

1.4 Motivation

The motivation behind TalkSpace lies in the necessity for more accurate and comprehensive methods of understanding human emotions and personality traits. Current approaches often rely on subjective assessments or single-modal analyses, resulting in limited accuracy and effectiveness across various domains like mental health, customer service, and education. By integrating advanced technologies such as text mining, signal processing, and computer vision, TalkSpace aims to overcome these limitations and provide a more nuanced understanding of human emotions. Its ability to analyze verbal, non-verbal, and contextual cues offers a holistic view of human communication, allowing for more accurate interpretation of emotions and personality traits. TalkSpace addresses the growing demand for personalized interventions by providing real-time insights into emotional states and behavioral tendencies. It empowers individuals to better understand themselves and others, fostering improved communication, enhanced mental well-being, and more effective interactions in both personal and professional settings.

2 Literature Review

2.1 Literature Survey

Early methods of emotion recognition primarily relied on handcrafted features and traditional machine learning techniques [1,2,4]. These methods often involved feature extraction from various modalities such as audio, text, and images, followed by the application of classifiers to recognize emotions. For example, Kratzwalda et al. (2018) explored text-based emotion recognition using deep learning approaches, emphasizing the importance of linguistic features and sentiment analysis [1]. Majumder et al. (2017) investigated deep learning-based document modeling techniques for personality detection from text, focusing on capturing semantic representations [2]. In the realm of audio, Basharirad and Moradhaseli (2017) conducted a literature review on speech emotion recognition methods, highlighting the significance of acoustic features and classification models [4]. These early approaches showed promise in controlled settings but faced challenges in handling complex real-world scenarios, where emotions are often expressed through multifaceted cues and in diverse contexts.

The limitations of traditional methods led to the emergence of deep learning techniques, which have revolutionized the field of emotion recognition [3,5,6,7]. Deep learning models, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated superior performance in capturing intricate patterns and variations in emotional expressions across different modalities [3]. For instance, Dhall et al. (2015) explored video and image-based emotion recognition challenges, highlighting the potential of deep learning in analyzing facial expressions and visual cues [5]. Pramerdorfer and Kampel (2017) investigated facial expression recognition using CNNs, showcasing the effectiveness of deep learning in extracting discriminative features from facial images [7]. Furthermore, Majumder et al. (2017) proposed deep learning-based approaches for personality detection from text, emphasizing the ability of deep models to learn complex representations from textual data [6].

The shift towards deep learning has enabled the development of more robust and accurate emotion recognition systems capable of handling diverse input modalities and challenging real-world scenarios. By automatically learning discriminative features from raw data, deep learning models have significantly improved emotion recognition accuracy and generalization [8,9,10]. For example, Giannakopoulos (2015) developed pyAudioAnalysis,

an open-source Python library for audio signal analysis, which facilitates research in various audio processing tasks, including emotion recognition [8]. Moreover, the Facial Emotion Recognition Challenge from Kaggle provides researchers with a benchmark dataset for evaluating and improving facial expression recognition models, fostering advancements in the field [10].

Furthermore, advancements in multimodal emotion recognition have gained attention, as emotions are often conveyed through multiple modalities simultaneously [11,12,13]. Integrating information from audio, visual, and textual sources allows for more comprehensive emotion understanding. For instance, Ekman et al. (1987) studied universals and cultural differences in facial expressions of emotion, highlighting the importance of visual cues in emotion recognition [13]. Similarly, End-to-End Multimodal Emotion Recognition using Deep Neural Networks proposes a framework for integrating information from multiple modalities to improve emotion recognition accuracy [12].

In addition to academic research, industry applications of emotion recognition are expanding rapidly, particularly in fields such as marketing, healthcare, and human-computer interaction [14,15,16]. For instance, Al-Hajjar and Syed (2015) applied sentiment and emotion analysis on brand tweets for digital marketing purposes, demonstrating the value of emotion recognition in understanding consumer behavior [15]. Eichstaedt et al. (2015) investigated the predictive power of psychological language on Twitter in determining public health outcomes, indicating the potential of emotion analysis in healthcare [16]. These real-world applications underscore the practical significance of emotion recognition technologies and their impact on various aspects of society.

3 System Design

3.1 Proposed System

The proposed system integrates multimodal emotion recognition techniques to provide a comprehensive understanding of human emotions. It utilizes text, audio, and video inputs to capture emotional cues from various channels.

For text analysis, natural language processing algorithms extract linguistic features and identify patterns indicative of emotions and personality traits. In audio analysis, signal processing techniques decode emotional content from voice intonations and acoustic features. Facial recognition algorithms analyze facial expressions in video inputs to detect underlying emotions.

The system employs machine learning models to classify emotions and personality traits based on the extracted features from each modality. Fusion techniques combine information from different modalities to improve the accuracy of emotion recognition.

The proposed system aims to provide real-time insights into emotional states across diverse contexts, including mental health assessments, customer service interactions, and educational settings.

3.2 Proposed System Algorithm

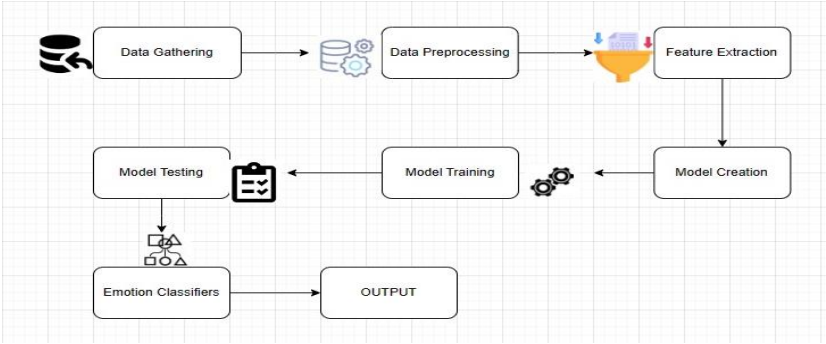


Fig 3.1 Proposed Model

As outlined above, the primary step in the proposed TalkSpace system is data gathering, followed by data preprocessing and feature extraction. Text inputs undergo convolutional processing followed by LSTM modeling for capturing sequential information. For audio, a time-distributed CNN architecture is applied to extract features from the acoustic signals. Meanwhile, video inputs are processed using the Xception model to analyze facial expressions and visual cues.

3.3 Methodology

In developing TalkSpace, the methodology encompasses a multi-step process integrating various techniques for each modality. Firstly, for text analysis, we studied SVM, LSTM, and Convolution + LSTM, with Convolution + LSTM showing superior results. Secondly, in audio analysis, we examined SVM and Time Distributed Convolution, with Time Distributed LSTM demonstrating better performance. Lastly, for video analysis, we studied VGG-16, ResNet, and Inception V3, but found Xception to yield the best results.

For text analysis, natural language processing algorithms are employed to extract linguistic features and identify patterns indicative of personality traits. These features undergo sentiment analysis to detect emotional cues. In audio analysis, signal processing algorithms decode emotional content from voice intonations and acoustic features, followed by emotion classification using Time Distributed LSTM. In image analysis, computer vision techniques are applied to analyze facial expressions. Facial landmark detection and feature extraction are followed by emotion recognition using Xception. The outputs from these modalities are fused to provide a comprehensive understanding of human emotions and personality traits.

We are going to explore several categorical targets depending on the input considered. Table 1 gives a summary of all the categorical targets we are evaluating depending on the data type.

Table 3.1 Data Types and Categorical targets

Data types	Categorical target
Textual	Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism along with Happy, Sad, Angry, Fearful, Surprise, Neutral and Disgust
Sound	Happy, Sad, Angry, Fearful, Surprise, Neutral and Disgust
Video	Happy, Sad, Angry, Fearful, Surprise, Neutral and Disgust

3.3.1 Text Analysis Using Conv + LSTM.

- **Data Collection:**

The dataset utilized in this study was collected by Pennebaker and King [1999], comprising 2,468 daily writing submissions from 34 psychology students (29 women and 5 men) aged between 18 and 67, with a mean age of 26.4.

- **Preprocessing:**
    - **Tokenization:** Involves dividing the document into individual words or tokens.
    - **Standardization:** Utilizes regular expressions to replace formulations like "can't" with "cannot" and "'ve" with "have".
    - **Deletion of Punctuation:** Removes punctuation marks from the tokens.
    - **Lowercasing:** Converts all tokens to lowercase.
    - **Removal of Stopwords:** Involves discarding predefined stopwords like 'a', 'an', etc.
    - **Part-of-Speech Tagging:** Assigns part-of-speech tags to the remaining tokens.
    - **Lemmatization:** Utilizes part-of-speech tags for more accurate lemmatization of tokens.
    - **Padding:** Sequences of tokens of each document are padded to a fixed length of 300 tokens. Tokens beyond this index are deleted, and zeros are added at the beginning of the vector if the input vector has less than 300 tokens. The padding dimension is determined based on the average number of words in each essay, resulting in a padding dimension of 300.
- **Embedding:**

Each token is replaced by its embedding vector using Google's pre-trained Word2Vec vectors in 300 dimensions, which incorporates the most information. The embedding is set to be trainable.

- **Classifier:**

The neural network architecture combines one-dimensional Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The CNN layer extracts features from the text data, followed by Long Short-Term Memory (LSTM) cells to capture the sequential nature of natural language. The final model comprises three consecutive blocks, each consisting of a CNN layer, max pooling layer, spatial dropout layer, and batch normalization layer. The number of convolution filters for each block is 128, 256, and 512, respectively, with a kernel size of 8, max pooling size of 2, and dropout rate of 0.3. Three stacked LSTM cells with 180 outputs each are followed by a fully connected layer of 128 nodes before the final classification layer.

3.3.2 Audio Processing Using Time Distributed CNN

- **Data Collection:**

The RAVDESS database was utilized, consisting of acted emotional speech from both male (672) and female (672) actors. Actors were prompted to express six different emotions (happy, sad, angry, disgust, fear, surprise) and neutral at two levels of emotional intensity.

- **Signal Preprocessing:**
    - **Pre-emphasis filter:** Balances the frequency spectrum by amplifying high frequencies to prevent issues in Fourier Transform computation.
    - **Framing:** Divides the signal into short-term windows to capture frequency contours over time. Common settings include a frame size of 25ms with a 15ms overlap.
    - **Hamming:** Each frame is multiplied by a Hamming window function to reduce spectral leakage and signal discontinuities.
    - **Discrete Fourier Transform (DFT):** Converts the signal from time domain to frequency domain for frequency content analysis.
- **Short-term audio features:**
    - **Time-domain features:** Energy, entropy of energy, zero crossing rate.
    - **Frequency-domain features:** Spectrogram, log-mel-spectrogram, spectral centroid, spectral spread, spectral entropy, spectral flux.

- **Model:**

The Time Distributed Convolutional Neural Network integrates hierarchical CNNs with an LSTM-based recurrent neural network to identify sequential patterns in speech signals. It applies a rolling window mechanism, processing each segment with a CNN featuring Local Feature Learning Blocks (LFLBs). The output is fed into an LSTM network with two cells to capture long-term dependencies, followed by a fully connected layer for emotion prediction.

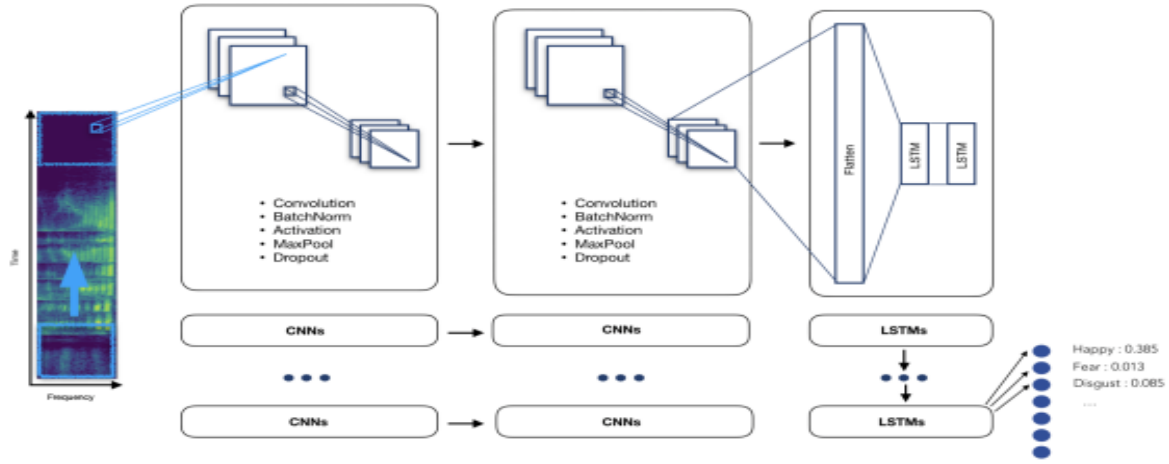


Fig 3.2 Time distributed CNN

- **Evaluation:**

The dataset was split into 80% for training, 15% for validation, and 5% for testing. Early stopping was employed to prevent overfitting, using Stochastic Gradient Descent with decay and momentum as the optimizer and a batch size of 64. Graphs display categorical cross-entropy loss and accuracy for training and validation sets.

- **Improvement:**

Our model achieved satisfactory results with prediction rates of approximately 65% for 7-way emotions and 75% for 6-way emotions (excluding surprise). To enhance performance further, we plan to explore more sophisticated classifiers such as Hidden Markov Models (HMM) and Convolutional Neural Networks (CNN) in the next phase.

3.3.3 Video Processing Using XCeption model

- **Data Collection:**

A dataset from FER2013 is collected, comprising images labeled with emotions such as Happy, Sad, Angry, Fearful, Surprise, Neutral, and Disgust. The train set contains 28,709 images, while the test set contains 3,589 images. Each image consists of grayscale pixels (48x48), representing facial expressions.

- **Data Preprocessing:**
    - **Grayscale Conversion:** Images are converted to grayscale using functions like cv2.cvtColor() to simplify processing.
    - **Face Detection and Zoom:** Face detection algorithms like cv2.CascadeClassifier.detectMultiScale() are used to identify and zoom in on faces.
    - **Multiple Face Management:** Techniques are employed to handle multiple faces, such as iterating through detected faces or selecting the primary face.
    - **Pixel Density Reduction:** cv2.resize() is used to reduce pixel density to match the training set's resolution.
    - **Image Transformation:** Preprocessed images are transformed using functions like cv2.resize() or cv2.normalize() to align with the model's input format.
    - **Emotion Prediction:** The preprocessed image is inputted into the XCeption Model for emotion prediction.
- **Model Architecture:**

The data passes through the entry flow, middle flow (repeated eight times), and exit flow. Convolution and SeparableConvolution layers are followed by batch normalization.

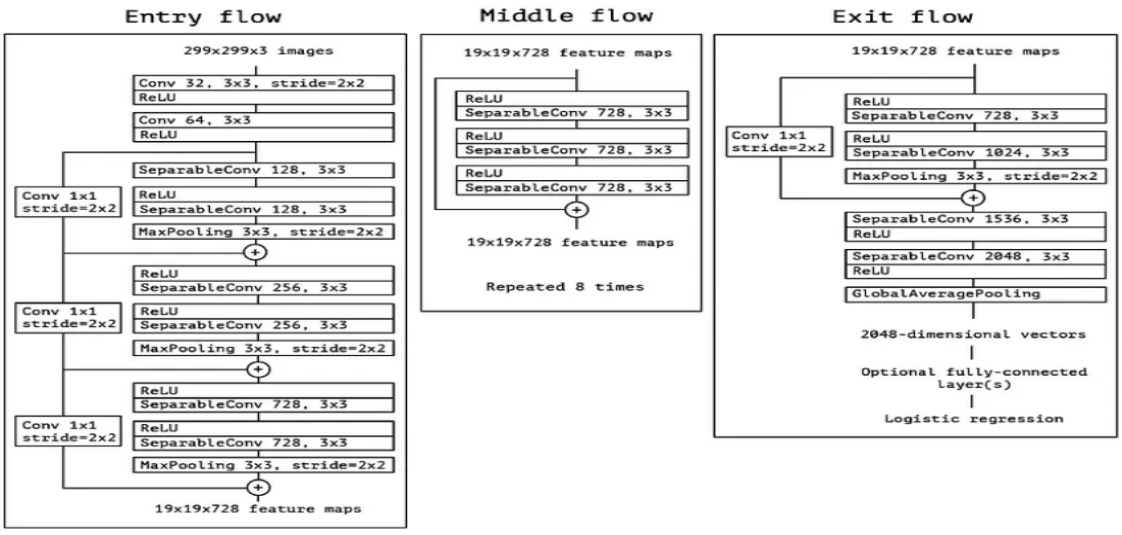


Fig 3.3 XCeption Structure

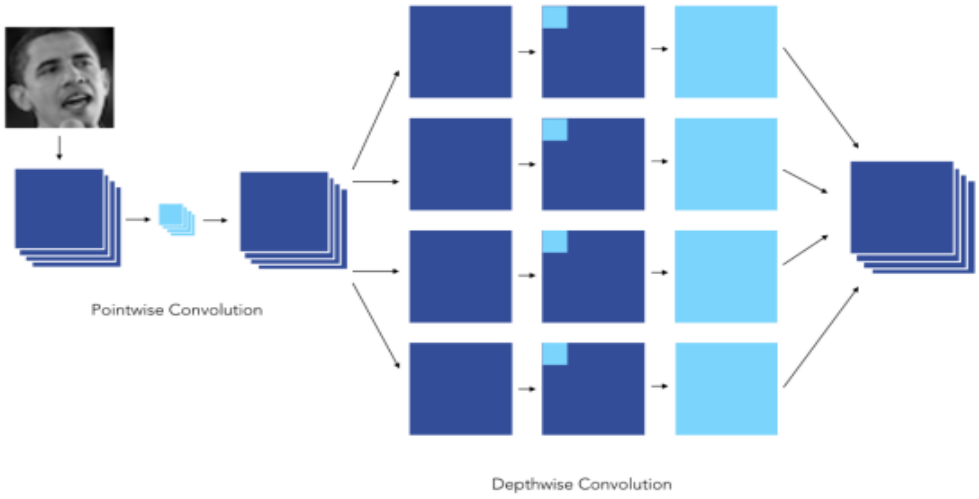


Fig 3.4 Xception

- **Training and Validation:**

The compiled model is trained on the training dataset using batches of images. During training, the model adjusts its parameters to minimize the loss function. It is evaluated on a separate validation dataset after each epoch to monitor performance and prevent overfitting. Validation accuracy and loss metrics are computed to assess generalization ability.

- **Hyperparameter Tuning:**

Hyperparameters like learning rate, batch size, and dropout rate are tuned to optimize performance. Techniques like learning rate scheduling or early stopping may be employed to improve training efficiency and prevent overfitting.

- **Deployment:**

The trained Xception model is deployed for inference on new unseen images, where it can classify emotions based on learned representations.

3.4 Analysis

3.4.1 Process Model

The model used for the project is Rapid Application Development (RAD) model.

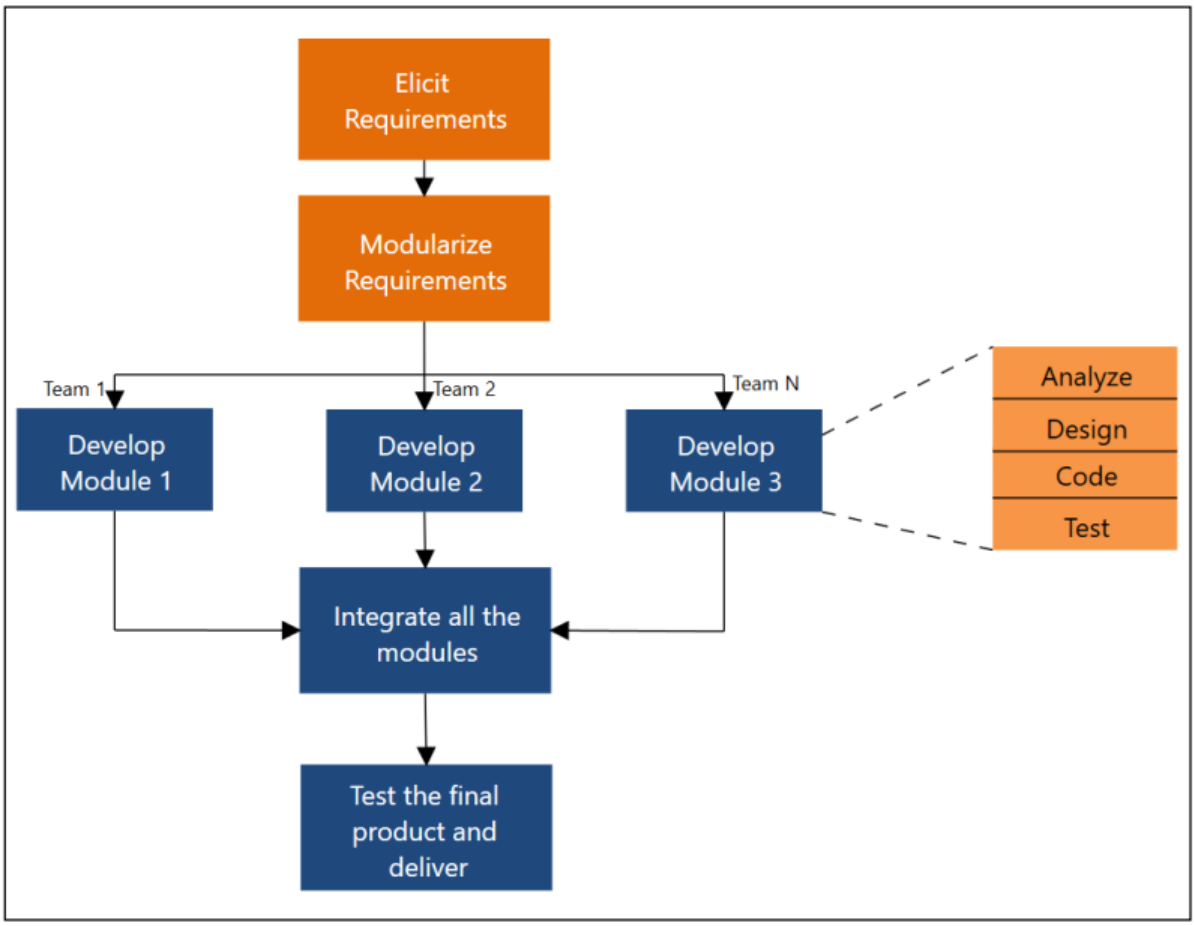


Fig 3.5 Process Model

This model can be used to implement a software project if the project is split down into discrete modules. These modules can then be joined to produce the final product. Another

notable feature of this model is the short time window for delivery (time-box), which is approximately 60-90 days.

3.4.2 Feasibility Analysis

• Data Availability

Various datasets are available for text, audio, and video inputs. Text data can be obtained from Pennebaker and King [1999]. Audio data can be sourced from The RAVDESS database, which contains acted emotional speech from both male and female actors. Video data can be collected from the FER2013 dataset, which includes images labeled with emotions such as Happy, Sad, Angry, Fearful, Surprise, Neutral, and Disgust.

• Hardware and Software

Adequate hardware resources, including GPUs, and software tools for deep learning are available for processing text, audio, and video data. This includes tools for text mining, signal processing, and computer vision tasks.For deployment, the TalkSpace platform will utilize Flask as the web framework, along with TensorFlow for machine learning tasks. The platform will be developed using Python for backend logic, and JavaScript, HTML, and CSS for frontend user interface.

• Model Complexity

Implementation of CNNs, LSTMs, and Xception model is feasible, given the available computational resources. These models have been successfully used in various applications, including emotion recognition tasks.

• Model Performance

Existing CNNs, LSTMs, and Xception models have shown promise in emotion recognition tasks. CNNs and LSTMs are effective in capturing patterns from sequential data such as text and audio, while the Xception model excels in analyzing visual data from images and videos. Their performance in previous studies indicates their technical feasibility for this project.

3.4.3 UML Diagram

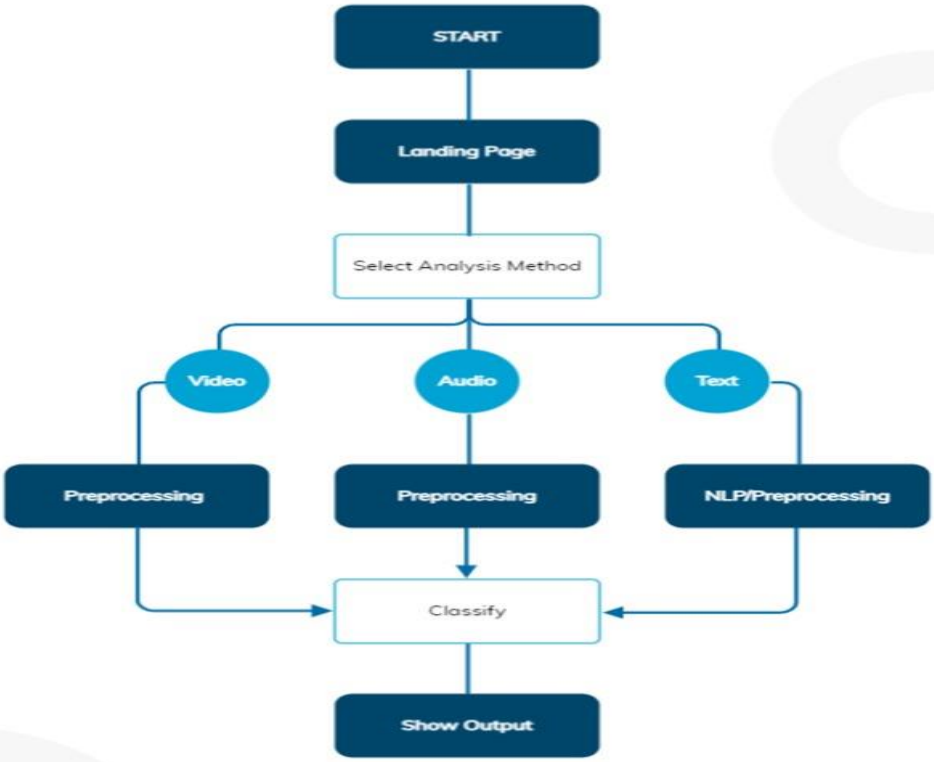


Fig 3.6 UML Diagram

4 System Implementation

4.1 Making of Prototype

4.1.1 Dataset

To build a functional prototype for emotion recognition, acquiring a suitable dataset is crucial. The dataset provides the foundation for training and evaluating the performance of the emotion recognition model. Several sources of data will be explored to ensure diversity and adequacy in emotion representation. One potential dataset is the RAVDESS database, which contains acted emotional speech and song recordings from male and female actors, covering various emotions at different intensity levels. Additionally, other publicly available emotion databases and labeled datasets will be considered to augment the dataset. These datasets will undergo preprocessing to ensure consistency and quality before being used for training and testing the model.

4.1.2 Model Overview

The prototype's model architecture is designed to effectively recognize and classify emotions from different modalities, including text, audio, and video. The model consists of multiple components tailored to process each modality appropriately.

For text processing, natural language processing (NLP) techniques will be employed. This involves tokenization, lemmatization, and sentiment analysis to extract relevant features from textual inputs. Machine learning algorithms such as Support Vector Machines (SVM), Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNN) will be utilized to classify emotions based on text.

In audio processing, the model will analyze audio signals to capture emotional cues. Techniques such as Fourier Transform and Mel-Frequency Cepstral Coefficients (MFCCs) will be used for feature extraction. Deep learning models like Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) will be trained to classify emotions from audio inputs.

For video processing, the model will analyze facial expressions to infer emotions. Computer vision techniques, including facial landmark detection and feature extraction, will be employed to capture facial expressions. Deep learning architectures such as Convolutional Neural Networks (CNNs), particularly pre-trained models like Xception, will be used for emotion recognition from video inputs.

The model components will be integrated using a fusion approach at the decision-making level to combine the outputs from text, audio, and video modalities. This fusion mechanism aims to provide a more comprehensive and accurate understanding of emotions by leveraging information from multiple sources.

4.1.3 Flask Framework

The Flask framework will serve as the backbone for developing the prototype's web interface. Flask is a lightweight and flexible web framework for Python, well-suited for building web applications with minimal overhead. It provides features like URL routing, template rendering, and request handling, making it ideal for developing interactive web applications.

In the context of the emotion recognition prototype, Flask will facilitate the integration of the trained model into a user-friendly web interface. The prototype's frontend will be developed using HTML, CSS, and JavaScript to create a visually appealing and interactive user interface. Flask will handle the backend logic, including processing user requests, invoking the emotion recognition model, and returning the predicted emotions to the user interface.

The web interface will allow users to input text, upload audio files, or provide video streams for emotion recognition. The backend server powered by Flask will process these inputs, apply the trained model for emotion recognition, and display the predicted emotions back to the user. Additionally, Flask will handle error handling, ensuring smooth functioning of the prototype under various scenarios.

Overall, the Flask framework will enable the development of a functional and user-friendly prototype for emotion recognition, allowing users to interact with the model seamlessly through a web interface.

4.2 Making of GUI

TalkSpace project presents a holistic solution for emotion recognition and personality trait classification. Leveraging advanced deep learning and machine learning techniques, including signal processing for emotion recognition, computer vision for facial emotion analysis, and text mining for personality trait classification, TalkSpace achieves accurate and real-time assessments of individuals' emotional states and personality traits. By integrating these components, TalkSpace facilitates enhanced understanding of human behavior, fosters personalized interactions, and opens avenues for applications in mental health, customer

service, and education. Overall, TalkSpace revolutionizes the landscape of emotion analysis and personality assessment, paving the way for more insightful and effective human-computer interactions.

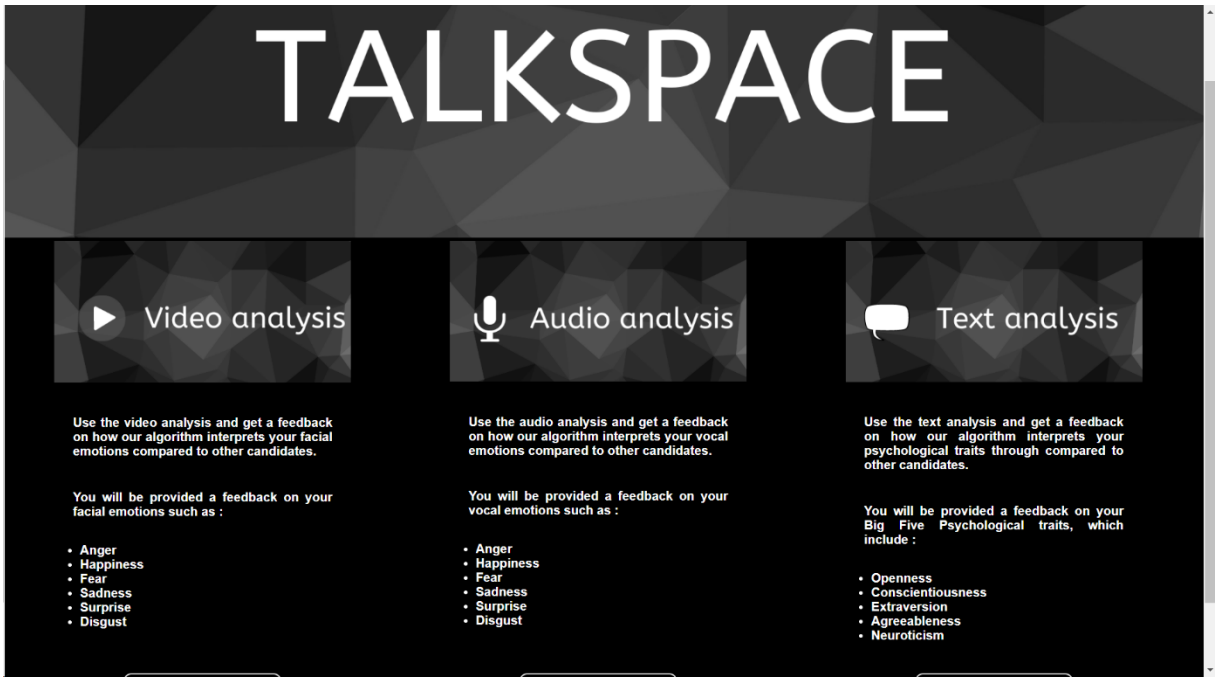


Fig 4.1 Homepage

As seen in Fig 4.2.1, A page is dedicated to each communication channel (audio, video, text) and allows the user to be evaluated. A typical interview question is asked on each page, for instance:" Tell us about the last time you showed leadership". The audio/video extract (recorded via computer microphones/webcam) or text block can be retrieved once saved and processed by our algorithms (in the case of the text channel the user can also upload a .pdf document that will be parsed by our tool)

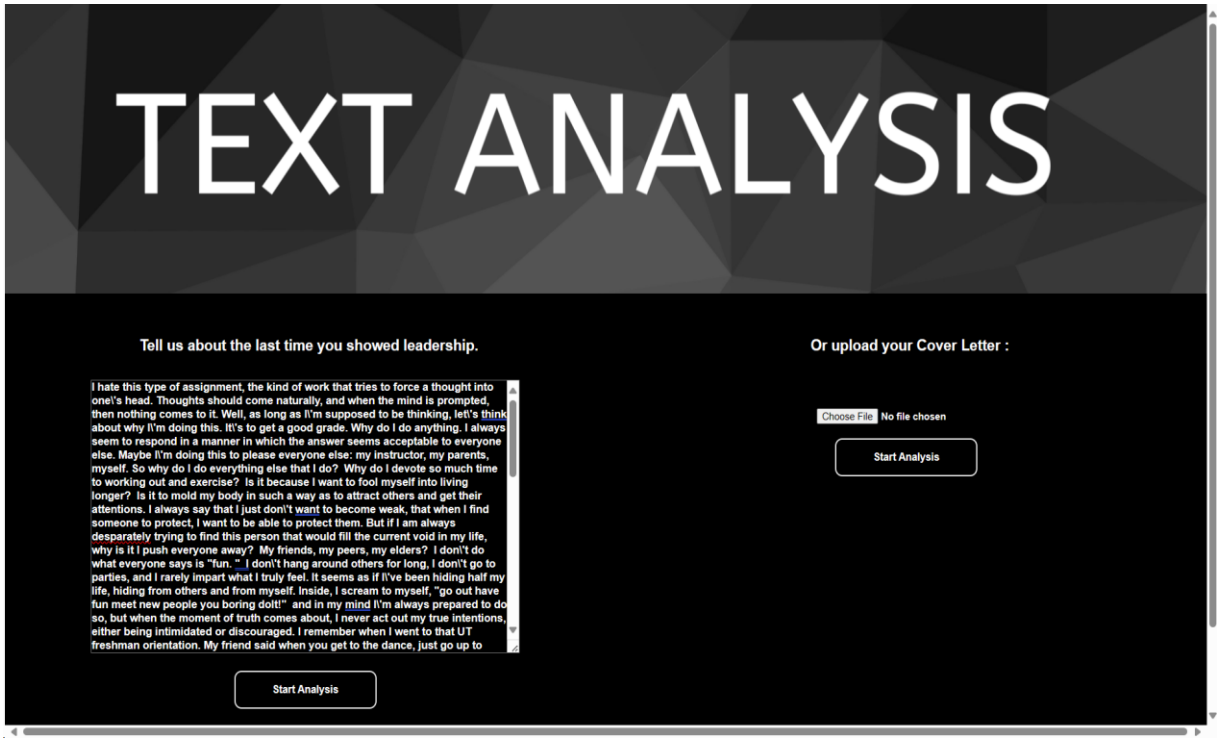


Fig 4.2 Text Analysis

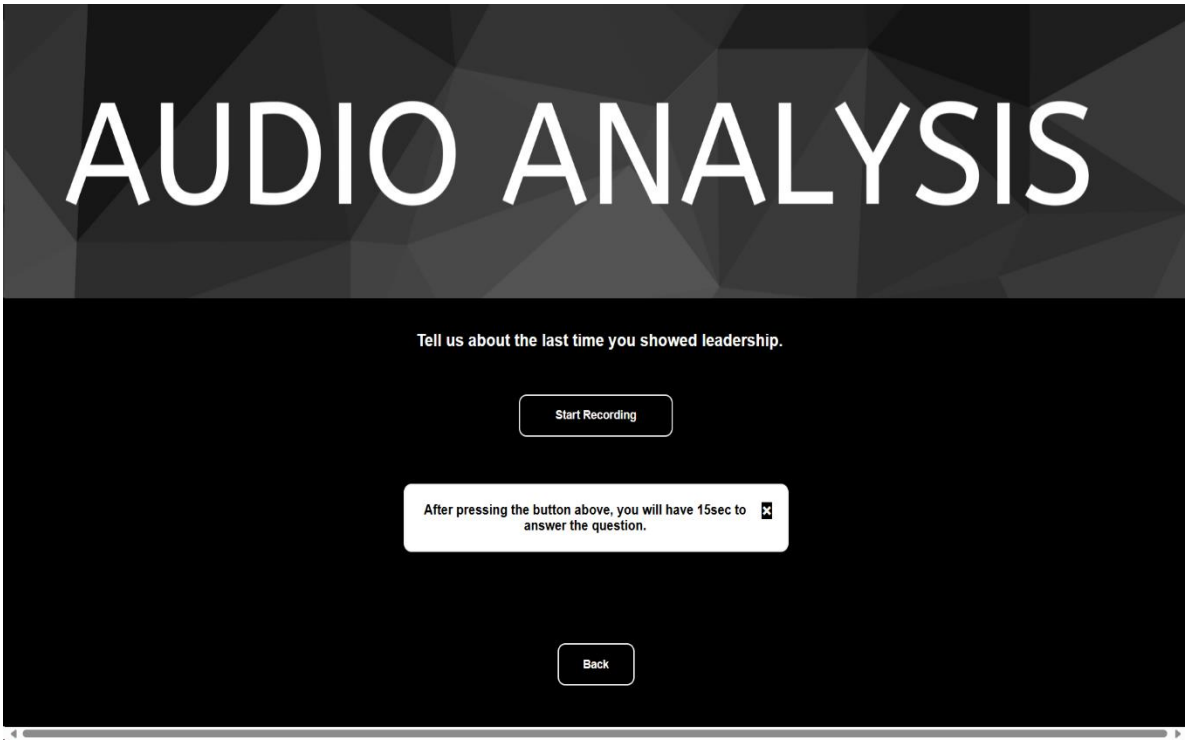


Fig 4.3 Audio Analysis

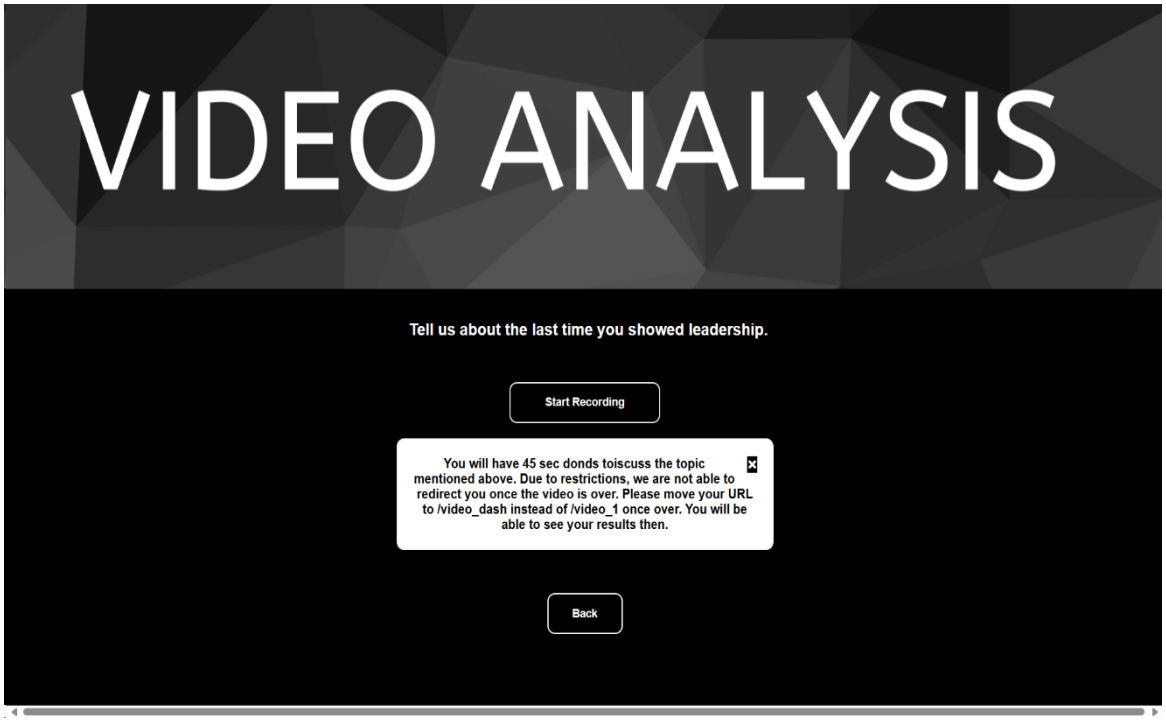


Fig 4.4 Video Analysis

The text and video/audio summaries are slightly different: for the text interview summary, not only we chose to display the percentage score of identified personality traits for both the user and the other candidates, but also the most frequently used word in the answer. For the video and audio interview summaries, we displayed the perceived emotions scores of the user and the other candidates. Following are the summary pages for both the text and video interviews.

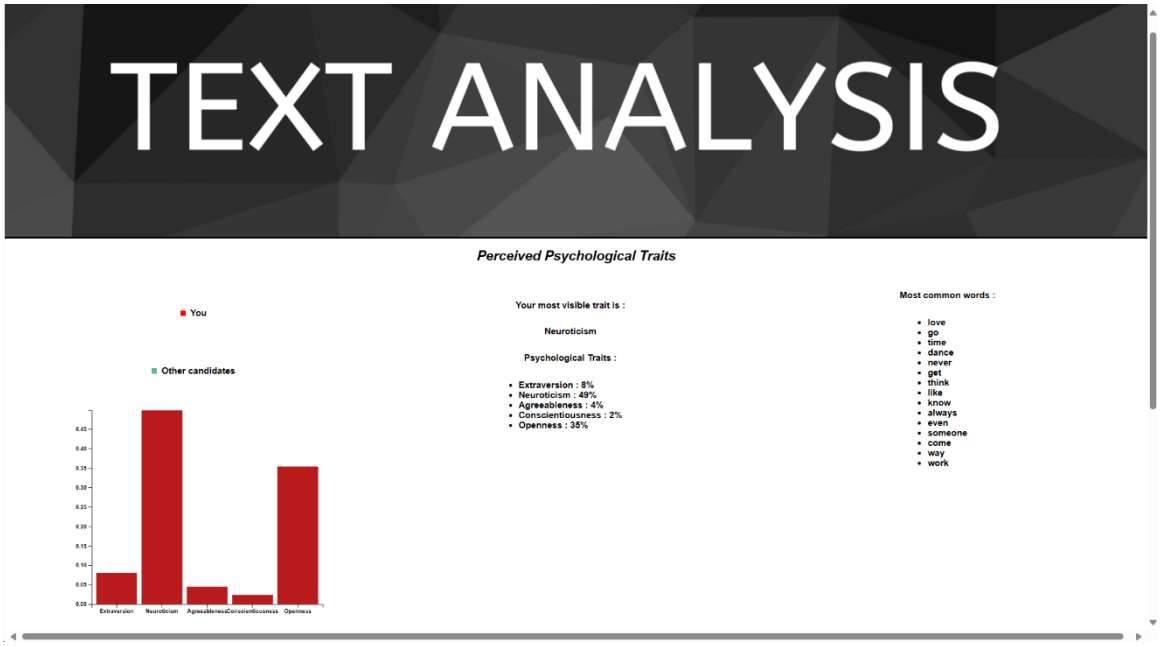


Fig 4.5 Text Analysis Result

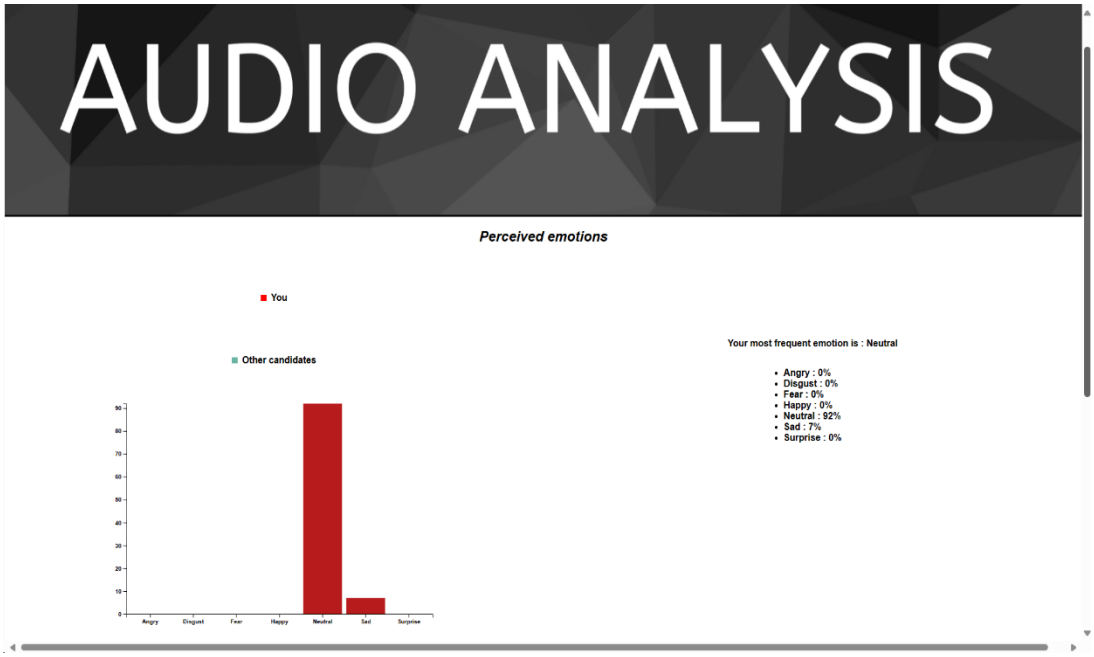


Fig 4.6 Audio Analysis Result

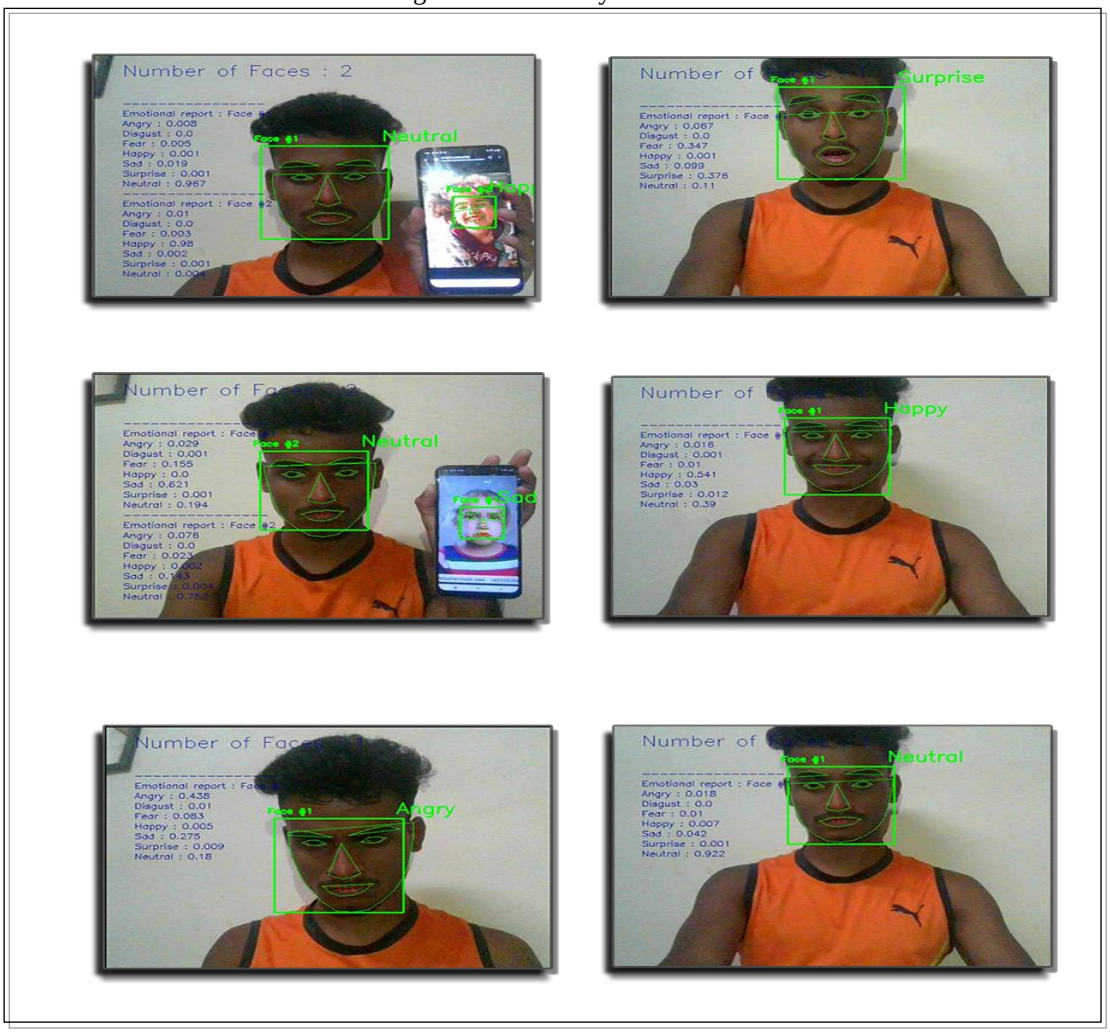


Fig 4.7 Emotion Detected

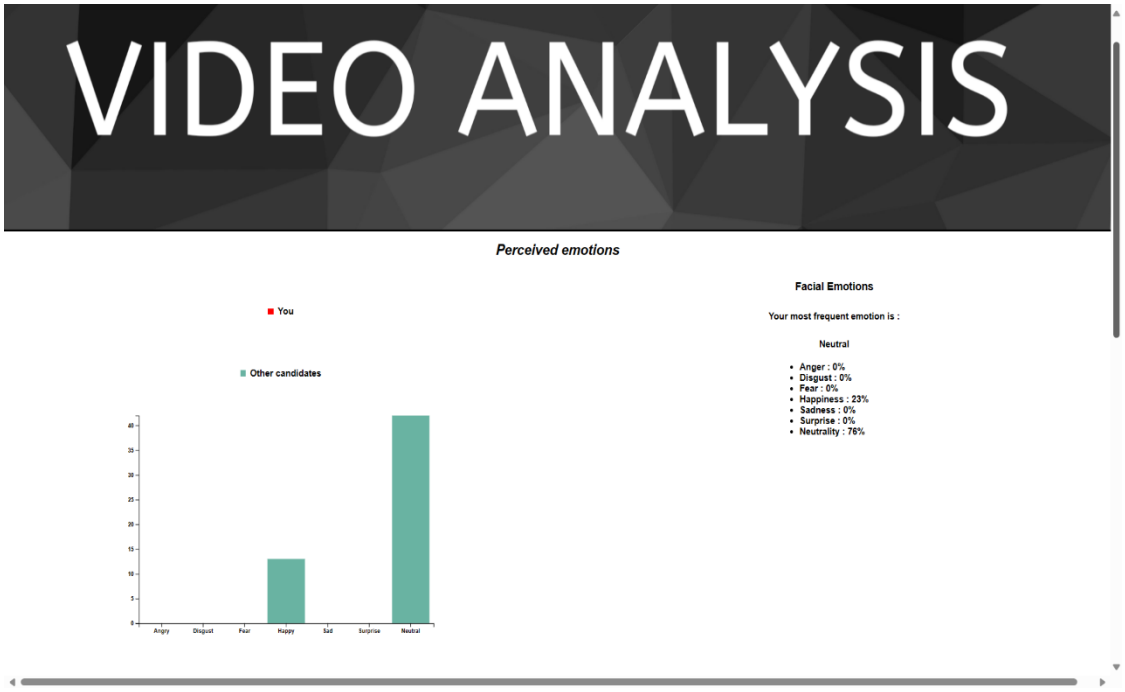


Fig 4.8 Video Analysis Result

5 Results and Discussions

5.1 Results and Discussions

5.1.1 Using Conv + LSTM For Text Analysis

Accuracy: 96.34%

```
Accuracy of our model on test data :  96.34121060371399 %
```

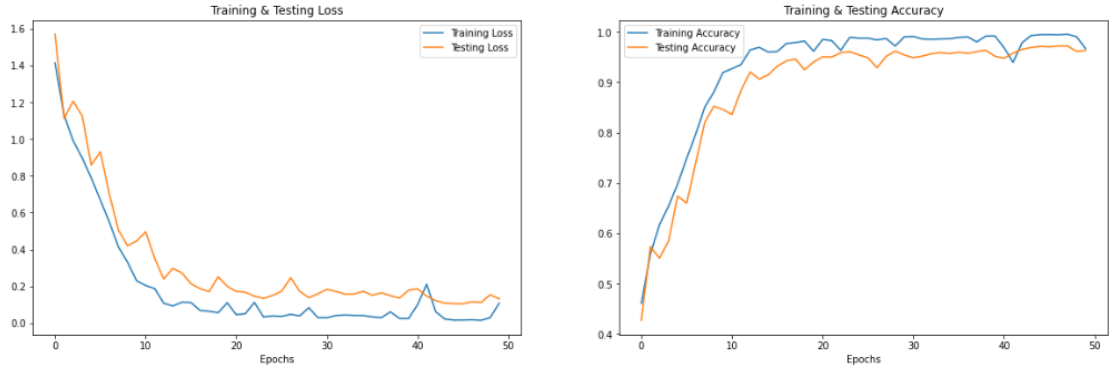


Fig 5.1 Loss and Accuracy curve for Text Analysis

The general health questionnaire, comprising 12 questions, serves as the initial round of

5.1.2 Using Time Distributed LSTM For Audio Analysis

Accuracy: 96%

	precision	recall	f1-score	support
angry	0.96	0.97	0.97	1484
disgust	0.97	0.95	0.96	1558
fear	0.96	0.97	0.96	1505
happy	0.96	0.95	0.96	1619
neutral	0.97	0.98	0.97	1558
sad	0.96	0.97	0.96	1478
surprise	0.98	0.97	0.97	528
accuracy			0.96	9730
macro avg	0.96	0.96	0.96	9730
weighted avg	0.96	0.96	0.96	9730

Fig 5.2 Performance matrix for Audio Analysis

5.1.3 Using Xception Model for Video Analysis

Accuracy: 97.25%

```
305/305 [==============================] - 8s 24ms/step - loss: 0.1117 - accuracy: 0.9725
accuracy: 97.25%
```

Fig 5.3 Loss and Accuracy curve for Video Analysis

6 Conclusion

6.1 Conclusion

In conclusion, the proposed emotion recognition system offers a multifaceted approach to understanding and interpreting human emotions across various modalities. By integrating text analysis, audio processing, and computer vision techniques, the system can capture subtle nuances in emotional expressions, providing a holistic understanding of users' feelings. The utilization of advanced machine learning algorithms, such as SVM, LSTM, CNNs, and pre-trained models like Xception, ensures accurate emotion classification based on textual, auditory, and visual inputs. Additionally, the fusion mechanism combines outputs from different modules to enhance the accuracy and reliability of emotion recognition. Through this comprehensive approach, the system enables effective communication, personalized services, and tailored interventions in domains such as mental health, customer service, and education.

Furthermore, the development of this emotion recognition system underscores the potential of technology in addressing complex human behaviors and interactions. As society becomes increasingly reliant on digital platforms for communication and engagement, the ability to accurately interpret and respond to human emotions becomes paramount. The proposed system not only demonstrates the feasibility of leveraging machine learning and deep learning techniques for emotion recognition but also highlights the importance of interdisciplinary approaches in solving real-world challenges. Moving forward, continued research and innovation in this field hold the promise of further improving the accuracy, efficiency, and accessibility of emotion recognition systems, contributing to enhanced human-computer interaction and overall well-being.

7 Future Scope

7.1 Future Scope

Looking ahead, there are several avenues for future development and enhancement of the emotion recognition system. One potential area of focus is the refinement of existing machine learning models and algorithms to improve accuracy and efficiency. This could involve exploring new architectures, optimizing hyperparameters, and integrating more sophisticated feature extraction techniques. Additionally, incorporating real-time processing capabilities would enable the system to provide instantaneous feedback, enhancing its utility in dynamic environments such as live chat support or virtual classrooms.

Furthermore, expanding the scope of emotion recognition to include more diverse cultural and linguistic contexts presents an exciting opportunity. Adapting the system to recognize and interpret emotions across different languages, dialects, and cultural norms would enhance its applicability and accessibility on a global scale. Additionally, integrating multimodal fusion techniques to combine information from multiple modalities, such as text, audio, and video, could further improve the accuracy and richness of emotion recognition results.

Moreover, exploring applications beyond traditional domains, such as mental health and customer service, could unlock new opportunities for the system. For instance, integrating emotion recognition into virtual reality (VR) environments could enhance immersive experiences by enabling virtual characters to respond empathetically to users' emotions. Additionally, leveraging emotion recognition in educational settings could support personalized learning experiences tailored to students' emotional states and preferences.

Overall, the future scope of the emotion recognition system lies in continual innovation, refinement, and diversification to meet evolving user needs and technological advancements. By embracing interdisciplinary approaches and leveraging emerging technologies, we can unlock the full potential of emotion recognition systems in enhancing human-computer interaction and facilitating positive emotional experiences.

8 References

8.1 References

[1]. **B.Kratzwalda, S.Ilie´, M.Kraus, S.Feuerriegel, H.Prendinger. Deep learning for affective computing: text-based emotion recognition in decision support, Sep. 2018.** This paper explores deep learning techniques for text-based emotion recognition in decision support systems, presenting approaches and methodologies for affective computing.

[2]. **N.Majumder, S.Poria, A.Gelbukh, E.Cambria. Deep Learning-Based Document Modeling for Personality Detection from Text, 2107.** This study investigates deep learning-based approaches for personality detection from textual documents, focusing on document modeling techniques.

[3]. **The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS).** This database provides emotional speech and song samples for research in emotion recognition from audio signals.

[4]. **B.Basharirad, and M.Moradhaseli. Speech emotion recognition methods: A literature review. AIP Conference Proceedings 2017.** This literature review discusses various methods and approaches for speech emotion recognition, providing insights into the state of the art in the field.

[5]. **L.Chen, M.Mao, Y.Xue and L.L.Cheng. Speech emotion recognition: Features and classification models. Digit. Signal Process, vol 22 Dec. 2012.** This paper examines speech emotion recognition, focusing on feature extraction methods and classification models.

[6]. **T.Vogt, E.Andre´ and J.Wagner. Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realisation. Affect and Emotion in Human-Computer Interaction, 2008.** This review paper discusses automatic recognition of emotions from speech and provides recommendations for practical implementation.

[7]. **T.Vogt and E.Andre´. Improving Automatic Emotion Recognition from Speech via Gender Differentiation. Language Resources and Evaluation Conference, 2006.** This paper proposes methods to improve automatic emotion recognition from speech by considering gender differentiation.

[8]. **T.Giannakopoulos. pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis. Dec. 2015.** The pyAudioAnalysis library provides tools for audio signal analysis, facilitating research in various audio processing tasks.

[9]. **The Facial Emotion Recognition Challenge from Kaggle.** This Kaggle challenge provides a dataset for facial expression recognition, allowing researchers to develop and evaluate facial emotion recognition models.

[10]. **C.Pramerdorfer, and M.Kampel. Facial Expression Recognition using Convolutional Neural Networks: State of the Art. Computer Vision Lab, TU Wien.** This paper presents the state of the art in facial expression recognition using convolutional neural networks.

[11]. **OpenCV open source library for image feature extraction.** OpenCV is a popular open-source library for computer vision tasks, including image feature extraction and processing.

[12]. **End-to-End Multimodal Emotion Recognition using Deep Neural Networks.** This paper introduces a method for multimodal emotion recognition using deep neural networks.

[13]. **Agrawal, A., & An, A. (2012). Unsupervised emotion detection from text using semantic and syntactic relations.** This study explores unsupervised methods for emotion detection from text by leveraging semantic and syntactic relations.

[14]. **Al-Hajjar, D., & Syed, A. Z. (2015). Applying sentiment and emotion analysis on brand tweets for digital marketing.** This paper discusses the application of sentiment and emotion analysis on brand tweets for digital marketing purposes.

[15]. **Alm, C. O., Roth, D., & Sproat, R. (2005). Emotions from text.** This paper explores methods for extracting emotions from text using linguistic analysis techniques.

[16]. **I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," Neural Networks, vol. 64, pp. 59–63, 2015.** This paper discusses challenges in representation learning, focusing on three machine learning contests.

[17]. **E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 37, no. 6, pp. 1113–1133, 2015.** This survey paper discusses automatic analysis of facial affect, covering registration, representation, and recognition techniques.

[18]. **M. V. B. Martinez, "Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition," in Advances in Face Detection and Facial Image Analysis.** This book chapter explores advances, challenges, and opportunities in automatic facial expression recognition.

[19]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep

[20]. **A. Dhall, O. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon. Video and Image Based Emotion Recognition Challenges in the Wild: EmotiW 2015.** This paper

discusses challenges and advancements in video and image-based emotion recognition, particularly focusing on the EmotiW 2015 challenge held at the ACM International Conference on Multimodal Interaction (ICMI).

[21]. **Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., Jha, S., Agrawal, M., Dziurzynski, L. A., Sap, M., Weeg, C., Larson, E. E., Ungar, L. H., & Seligman, M. E. P. Psychological language on Twitter predicts county-level heart disease mortality.** This study investigates the predictive power of psychological language used on Twitter in determining county-level heart disease mortality rates, offering insights into the relationship between language and public health.

[22]. **Danisman, T., & Alpkocak, A. Feeler: Emotion classification of text using vector space model.** This paper presents Feeler, a system for emotion classification of text using vector space models, exploring methods for analyzing and classifying emotions expressed in text.

[23]. **Chitturi, R., Raghunathan, R., & Mahajan, V. Form versus function: How the intensities of specific emotions evoked in functional versus hedonic trade-offs mediate product preferences.** This research examines the impact of emotions, specifically those evoked in functional versus hedonic trade-offs, on product preferences, shedding light on consumer behavior and decision-making.

[24]. **Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., & Ricci-Bitti, P. E. Universals and cultural differences in the judgments of facial expressions of emotion.** This seminal work by Ekman et al. investigates universals and cultural differences in the judgments of facial expressions of emotion, contributing significantly to our understanding of cross-cultural psychology and emotion recognition.

9 Appendix

9.1 Published Paper

IJSREM

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 08 ISSUE: 04 | APRIL - 2024 SJIF RATING: 8.448 ISSN: 2582-3930

TalkSpace - Advanced Emotion Recognition and Personality Analysis Platform

Rohan Kadu
BE in Information Technology
Vidyalankar Institute of Technology
rohan.kadu@vit.edu.in

Krishnakant Gangurde
BE in Information Technology
Vidyalankar Institute of Technology
krishnakant.gangurde@vit.edu.in

Vaishnav Rajput
BE in Information Technology
Vidyalankar Institute of Technology
vaishnav.rajput@vit.edu.in

Prof. Samuel Jacob
Assistant Professor
Department of Information Technology
Vidyalankar Institute of Technology
samuel.jacob@vit.edu.in

***Abstract*— TalkSpace is a revolutionary platform designed to analyze emotions through facial expressions, audio cues, and textual inputs using advanced Deep Learning and Machine Learning models. This project represents a significant advancement in emotional recognition technology, enabling precise identification and understanding of various emotional states in individuals. By leveraging state-of-the-art algorithms, TalkSpace offers an unparalleled level of accuracy in emotion detection, facilitating improved communication and interaction in various domains such as mental health, customer service, and education. With its comprehensive suite of capabilities, including real-time emotion analysis and contextual understanding, TalkSpace aims to revolutionize how emotions are perceived and addressed in diverse contexts, ultimately enhancing human-computer interaction and emotional well-being.**

I. INTRODUCTION

TalkSpace embodies a pioneering endeavor in the realm of emotion recognition and personality trait classification, leveraging an amalgamation of cutting-edge technologies. Through the convergence of text mining, signal processing, and computer vision techniques, TalkSpace offers a comprehensive solution for understanding and interpreting human emotions in diverse contexts.

At its core, TalkSpace utilizes text mining methodologies to delve into textual inputs, extracting valuable insights into individuals' personality traits. By analyzing language patterns and textual cues, the system accurately classifies personality traits, providing profound understandings of behavioral tendencies and characteristics.

Moreover, TalkSpace employs sophisticated signal processing algorithms to decode emotional cues embedded within audio signals. Through precise analysis of voice intonations, speech patterns, and acoustic features, TalkSpace achieves nuanced emotion recognition, enabling the detection of subtle emotional nuances in spoken communication.

In parallel, TalkSpace harnesses the power of computer vision for emotion recognition, analyzing facial expressions to decipher underlying emotions. Leveraging advanced image processing techniques, the system identifies key facial features and dynamics, allowing for real-time assessment of emotional states with remarkable accuracy.

By seamlessly integrating these technologies, TalkSpace transcends traditional approaches to emotion recognition, offering a multifaceted and holistic understanding of human emotions. Whether in mental health assessments, customer service interactions, or educational settings, TalkSpace empowers users to effectively interpret and respond to emotional cues, fostering improved communication and interaction across various domains.

II. PROBLEM STATEMENT

Despite the advancements in technology, understanding and interpreting human emotions and personality traits remain challenging tasks. Traditional methods for emotion recognition and personality trait classification often rely on subjective assessments or manual analyses, leading to inaccuracies and inefficiencies in various domains such as mental health, customer service, and education.

Moreover, existing solutions often focus on single modalities such as text analysis or facial recognition, overlooking the

holistic nature of human communication, which involves a combination of verbal, non-verbal, and contextual cues.

This fragmented approach hampers the ability to accurately capture and interpret the complexity of human emotions, limiting the effectiveness of communication and interaction in diverse contexts. Furthermore, the lack of integration between different modalities hinders the development of comprehensive solutions that can provide nuanced insights into human behavior.

Therefore, there is a critical need for a unified framework that leverages multiple modalities, including text mining, signal processing, and computer vision, to enable accurate and holistic emotion recognition and personality trait classification. Such a framework would not only enhance our understanding of human emotions but also facilitate the development of more effective communication strategies, personalized services, and tailored interventions across various domains.

III. RELATED WORK

This References [1] [23] [25] enhances emotion recognition in textual data using modifications like bidirectional processing and dropout regularization. Computational experiments show significant performance improvements, with pre-trained bidirectional LSTMs outperforming traditional models. The proposed sent2affect strategy, utilizing transfer learning from sentiment analysis tasks, further boosts performance.

References [10] [11] [12] [13] provides a thorough examination of CNN-based Facial Emotion Recognition (FER), highlighting performance differences and bottlenecks. We showcased notable enhancements achieved through modern CNNs and ensemble techniques. Looking ahead, our focus will be on addressing remaining challenges, with emphasis on data augmentation strategies tailored for FER and addressing biases in existing datasets like FER2013. Additionally, we aim to explore the development of a more comprehensive and publicly available FER dataset to advance research in this field.

References [4] [5] [8] [9] presents a concise review of audio-based emotion recognition systems, assessing their performance in terms of classifiers, features, recognition rates, and datasets. Notably, well-designed classifiers have demonstrated high accuracy across various emotional states. Current research emphasizes exploring different features, including Mel-frequency cepstral coefficients (MFCCs), to enhance recognition rates. Additionally, time-distributed CNNs show promise in improving performance. However, challenges with existing datasets hinder accurate evaluation of emotion recognition in audio recordings.

IV. PROPOSED ALGORITHM

Multimodal Emotion Recognition is a relatively new discipline that aims to include text inputs, as well as sound and video. This field has been rising with the development of social networks that gave researchers access to a vast amount of data. Recent studies have been exploring potential metrics to measure the coherence between emotions from the different channels. We are going to explore several categorical targets depending on the input considered. Table 1 gives a summary of all the categorical targets we are evaluating depending on the data type.

Data types	Categorical target
Textual	Openness, Conscientiousness, Extraversion, Agreeableness , Neuroticism
Sound	Happy, Sad, Angry, Fearful, Surprise, Neutral and Disgust
Video	Happy, Sad, Angry, Fearful, Surprise, Neutral and Disgust

Table 1: Categorical target depending on the input data type.

Algorithm 1: Text mining for personality trait using NN (CONV + LSTM)

Input: Any emotional text

Output: Openness , Conscientiousness , Extraversion , Agreeableness , Neuroticism .

1. Data Collection:

We are using data that was gathered in a study by Pen nebaker and King [1999]. It consists of a total of 2,468 daily writing submissions from 34 psychology students (29 women and 5 men whose ages ranged from 18 to 67 with a mean of 26.4).

2. Preprocessing:

-Tokenization: Involves dividing the document into individual words or tokens.
-Standardization: Includes using regular expressions to replace formulations such as "can't" with "cannot" and "'ve" with "have".
-Deletion of Punctuation: Removes punctuation marks from the tokens.
-Lowercasing: Converts all tokens to lowercase.
-Removal of Stopwords: Involves removing predefined stopwords like 'a', 'an', etc.
-Part-of-Speech Tagging: Assigns part-of-speech tags to the remaining tokens.

-Lemmatization: Utilizes part-of-speech tags for more accurate lemmatization of tokens.
-Padding: the sequences of tokens of each document to constrain the shape of the input vectors. The input size has been fixed to 300 : all tokens beyond this index are deleted. If the input vector has less than 300 tokens, zeros are added at the beginning of the vector in order to normalize the shape. The dimension of the padded sequence has been determine using the characteristics of our training data. The average number of words in each essay was 652 before any preprocessing. After the standardization of formulations, and the removal of punctuation characters and stopwords, the average number of words dropped to with a standard deviation of . In order to make sure we incorporate in our classification the right number of words without discarding too much information, we set the padding dimension to 300, which is roughly equal to the average length plus two times the standard deviation.

3. Embedding:

Each token is replaced by its embedding vector using Google's pre-trained Word2Vec vectors in 300 dimensions (which is the largest dimension available and therefore incorporates the most information), and this embedding is set to be trainable (our training corpus is to small to train our own embedding).

4. Classifier:

The neural network architecture combines one-dimensional convolutional neural networks (CNNs) and recurrent neural networks (RNNs). The one-dimensional convolutional layer extracts features from the text data, followed by Long Short-Term Memory (LSTM) cells to leverage the sequential nature of natural language. Unlike regular neural networks, LSTMs progressively accumulate and capture information through sequences, selectively remembering patterns for long durations.
The final model comprises three consecutive blocks, each consisting of a one-dimensional convolution layer, max pooling layer, spatial dropout layer, and batch normalization layer. The number of convolution filters for each block is 128, 256, and 512, respectively, with a kernel size of 8, max pooling size of 2, and dropout rate of 0.3.
Following the three blocks, three LSTM cells with 180 outputs each are stacked. Finally, a fully connected layer of 128 nodes is added before the last classification layer.

Algorithm 2: Signal processing for emotion recognition using Time Distributed CNN

Input: Audio file representing a emotion

Output: Happy , Sad , Angry , Fearful ,Surprise , Neutral , Disgust.

1. Data Collection:

The RAVDESS database was used, It contains acted emotions speech of male (672) and female (672) actors (gender balanced) that were asked to pretend six different emotions (happy, sad, angry, disgust, fear, surprise and neutral) at two levels of emotional intensity.

2. Signal Preprocessing:

-Pre-emphasis filter: Amplifies high frequencies to balance the frequency spectrum and prevent numerical issues in Fourier Transform computation.
-Framing: Divides the signal into short-term windows to capture frequency contours over time. Typically, window sizes range from 20ms to 50ms with 40% to 50% overlap between consecutive windows. Common settings include a frame size of 25ms with a 15ms overlap.
-Hamming: Each frame is multiplied by a Hamming window function to reduce spectral leakage and signal discontinuities, improving clarity. The Hamming function ensures that the beginning and end of frames match up smoothly.
-Discrete Fourier Transform (DFT): Converts the signal from the time domain to the frequency domain, allowing analysis of frequency content. This transformation facilitates the representation and analysis of audio features, which are often defined in the frequency domain.

3. Short-term audio features:

Time-domain features:
-Energy: Sum of squares of signal values normalized by frame length.
-Entropy of energy: Measures abrupt changes in energy amplitude of an audio signal.
-Zero Crossing Rate: Rate of sign changes of an audio signal.

Frequency-domain features:
-Spectrogram: Represents time evolution of frequency content.
-Log-mel-spectrogram: Utilizes mel-frequency scale for human auditory system resemblance.
-Spectral centroid: Center of gravity of the sound spectrum.
-Spectral spread: Second central moment of the sound spectrum.
-Spectral entropy: Measures normalized spectral energies.
-Spectral flux: Measures spectral changes between successive frames.

4. Model:

The Time distributed convolutional neural network integrates hierarchical CNNs with an LSTM-based recurrent neural network to identify sequential patterns in speech signals. Operating directly on log-mel-spectrograms, it applies a rolling window mechanism, processing each segment with a convolutional neural network featuring four Local Feature Learning Blocks (LFLBs). The output is then fed into a recurrent neural network with two LSTM cells to capture long-term dependencies, culminating in a fully connected layer with softmax activation for emotion prediction.

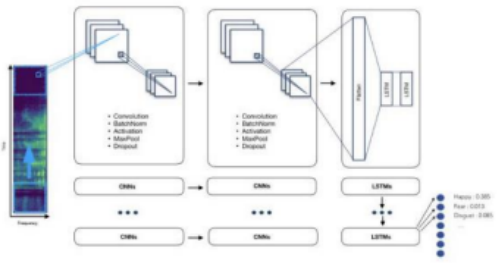


Fig. 1: Time distributed CNN

4. Evaluation:

We report the results of our deep learning model, divided the dataset into 80% for training, 15% for validation, and 5% for testing. Early stopping was employed to prevent overfitting, using Stochastic Gradient Descent with decay and momentum as the optimizer and a batch size of 64. Graphs display categorical cross-entropy loss and accuracy for training and validation sets.

5. Improvement:

Our model achieved satisfactory results with a prediction recognition rate of approximately 65% for 7-way emotions and 75% for 6-way emotions (surprised removed). To enhance performance further, we plan to explore more sophisticated classifiers such as Hidden Markov Models (HMM) and Convolutional Neural Networks (CNN) in the next phase.

Algorithm 3: Computer vision for emotion recognition using XCeption model

Input: WebCam Video representing a emotion

Output: Happy , Sad , Angry , Fearful ,Surprise , Neutral , Disgust.

1. Data Collection:

Collect a dataset of FER2013 data set containing number of class by emotions such as Happy , Sad , Angry , Fearful ,Surprise , Neutral , Disgust.
The train set has 28709 images, the test set has 3589 images. For each image, the data set contains the grayscale color of 2304 pixels (48x48), as well as the emotion associated.

2. Data Preprocessing:

-Grayscale Conversion: Frames are converted to grayscale using functions like cv2.cvtColor() to simplify processing and reduce input complexity.
-Face Detection and Zoom: Face detection algorithms, possibly using functions like cv2.CascadeClassifier. detectMultiScale(), are employed to identify and zoom in on faces within each frame.
-Multiple Face Management: Techniques such as iterating through detected faces or selecting the primary face can be implemented to handle multiple faces.
-Pixel Density Reduction: Functions like cv2.resize() may be used to reduce pixel density and match the training set's resolution.
-Image Transformation: The preprocessed image is transformed using functions like cv2.resize() or cv2.normalize() to align with the model's input format.
-Emotion Prediction: After preprocessing, the image is inputted into the model for emotion prediction, utilizing functions relevant to the XCeption Model

3. Model Architecture:

The data first goes through the entry flow, then through the middle flow which is repeated eight times, and finally through the exit flow. Note that all Convolution and SeparableConvolution layers are followed by batch normalization.

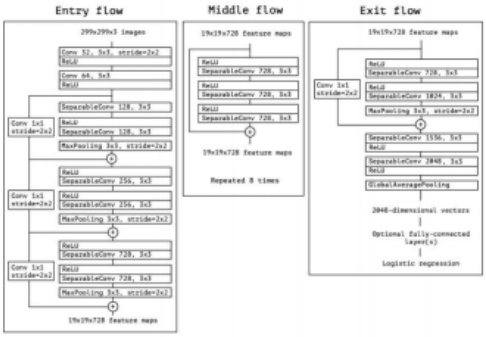


Fig. 2: XCeption Structure

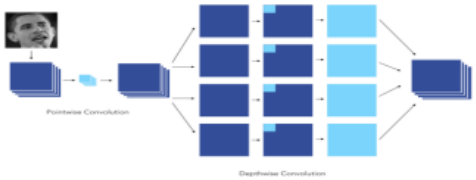


Fig. 3: XCeption

4. Training and Validation:

The compiled model is trained on the training dataset using batches of images. During training, the model iteratively adjusts its parameters to minimize the loss function, optimizing its ability to classify images accurately.
To monitor the model's performance and prevent overfitting, it is evaluated on a separate validation dataset after each epoch. Validation accuracy and loss metrics are computed to assess the model's generalization ability.

5. Hyperparameters -Tuning :

Hyperparameters such as learning rate, batch size, and dropout rate may be tuned to optimize the model's performance further. Techniques like learning rate scheduling or early stopping may also be employed to improve training efficiency and prevent overfitting.

6. Deployment:

Finally, the trained Xception model can be deployed for inference on new unseen images, where it can classify objects or perform other relevant tasks based on its learned representations.

V. IMPLEMENTATION

A. Text Analysis Using NN And LSTM:

Text analysis involves deriving insights from textual data. Neural networks, especially LSTM, are effective for this task. Preprocessing involves tokenization and encoding text into numerical representations. A neural network architecture, typically comprising LSTM layers, is trained on labeled data. Once trained, the model can be used for tasks like sentiment analysis and text classification. This approach offers a versatile solution for deriving insights from text data.

B. Video Analysis Using XCeption:

Xception, a powerful convolutional neural network, is utilized for video analysis. Each frame of the video is treated as an image and passed through the Xception model. The pre-trained Xception model extracts high-level features for tasks such as action recognition or object detection. This approach enables accurate and robust video analysis across various applications.

C. Audio Analysis Using Time Distributed CNN:

Utilizing a Time Distributed CNN, we analyze audio signals directly from log-mel-spectrograms. This model combines CNNs with LSTM networks to capture temporal patterns and dependencies for accurate emotion prediction.

D. Web Application Development Using Flask:

To provide users with seamless access to the TalkSpace system, a web application will be developed using the Flask framework. This web application will feature a user-friendly interface comprising views and templates for uploading input file, displaying results. Compatibility and responsiveness across different devices and browsers will be ensured to accommodate a diverse user base and enhance overall user experience.

VI. RESULTS

TalkSpace project presents a holistic solution for emotion recognition and personality trait classification. Leveraging advanced deep learning and machine learning techniques, including signal processing for emotion recognition, computer vision for facial emotion analysis, and text mining for personality trait classification, TalkSpace achieves accurate and real-time assessments of individuals' emotional states and personality traits. By integrating these components, TalkSpace facilitates enhanced understanding of human behavior, fosters personalized interactions, and opens avenues for applications in mental health, customer service, and education. Overall, TalkSpace revolutionizes the landscape of emotion analysis

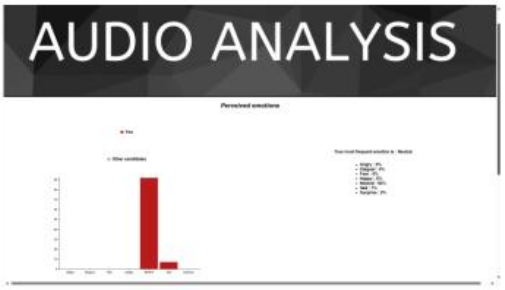
and personality assessment, paving the way for more insightful and effective human-computer interactions.

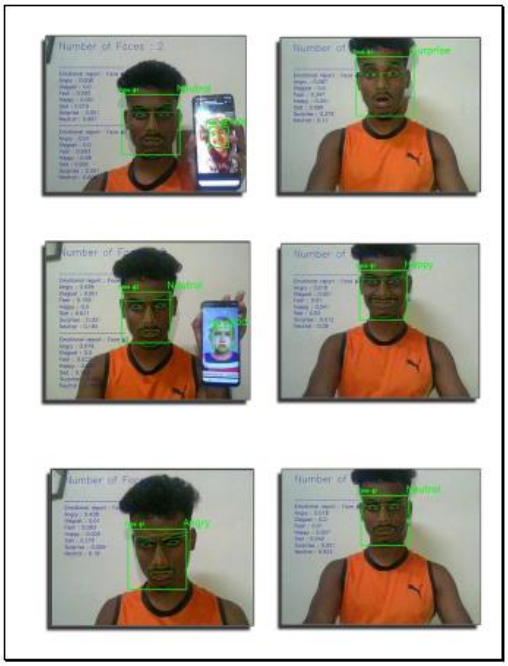
Fig. 4: TalkSpace Homepage

As seen in Fig 1, A page is dedicated to each communication channel (audio, video, text) and allows the user to be evaluated. A typical interview question is asked on each page, for instance:" Tell us about the last time you showed leadership". The audio/video extract (recorded via computer microphones/webcam) or text block can be retrieved once saved and processed by our algorithms (in the case of the text channel the user can also upload a .pdf document that will be parsed by our tool)

Fig. 5: Text Analysis

The text and video/audio summaries are slightly different : for the text interview summary, not only we chose to display the percentage score of identified personality traits for both the user and the other candidates, but also the most frequently used word in the answer. For the video and audio interview summaries, we displayed the perceived emotions scores of the user and the other candidates. Following are the summary pages for both the text and video interviews.





VII. CONCLUSION

TalkSpace stands at the forefront of emotion recognition and personality trait classification. By integrating cutting-edge technologies such as deep learning and signal processing, it offers a transformative approach to understanding human emotions across text, audio, and video modalities. This innovative platform has the potential to revolutionize various domains, from mental health diagnostics to customer sentiment analysis, by providing invaluable insights into human behaviour and emotional states.

With its versatility and accuracy, TalkSpace paves the way for more personalized interactions, tailored interventions, and informed decision-making processes. As we continue to refine and expand its capabilities, TalkSpace holds the promise of fostering greater empathy, understanding, and connection in our increasingly digital world. Through its seamless integration of technology and human experiences, TalkSpace is poised to make a profound impact on how we perceive, analyze, and respond to emotions in today's fast-paced and interconnected society.

ACKNOWLEDGMENT

I extend my sincere thanks to Professor Samuel Jacob for his invaluable guidance and expertise, which have been instrumental in shaping the development of TalkSpace - An Innovative Emotion Recognition and Personality Trait Classification Platform. Professor Jacob's insights and mentorship have greatly contributed to the success of this project.

Furthermore, I would like to express my gratitude to Vidyalankar Institute of Technology for providing the necessary infrastructure and resources that made the development of TalkSpace possible. The support and facilities provided by Vidyalankar Institute of Technology have been essential in facilitating the implementation and execution of this project.

REFERENCES

[1] B.Kratzwalda, S.Ilie´, M.Kraus, S.Feuerriegel, H.Prendinger. Deep learning for affective computing: text-based emotion recognition in decision support, Sep. 2018. https://arxiv.org/pdf/1803.06397.pdf

[2] N.Majumder, S.Poria, A.Gelbukh, E.Cambria. Deep Learning-Based Document Modeling for Personality Detection from Text, 2107. http://sentic.net/deep-learning-based-personality-detection.pdf

[3] The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS), https://zenodo.org/record/1188976/?f=3.XAcEs5NKhQK

[4] B.Basharirad, and M.Moradhaseli. Speech emotion recognition methods: A literature review. AIP Conference Proceedings 2017. https://aip.scitation.org/doi/pdf/10.1063/1.5005438

[5] L.Chen, M.Mao, Y.Xue and L.L.Cheng. Speech emotion recognition: Features and classification models. Digit. Signal Process, vol 22 Dec. 2012.

[6] T.Vogt, E.Andre´ and J.Wagner. Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realisation. Affect and Emotion in Human-Computer Interaction, 2008.

[7] T.Vogt and E.Andre´. Improving Automatic Emotion Recognition from Speech via Gender Differentiation. Language Resources and Evaluation Conference, 2006.

[8] T.Giannakopoulos. pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis. Dec. 2015 https://doi.org/10.1371/journal.pone.

[9] T.Giannakopoulos. pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis. Dec. 2015 https://doi.org/10.1371/journal.pone.0144610.

[10] The Facial Emotion Recognition Challenge from Kaggle, https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/da

[11] . C.Pramerdorfer, and M.Kampel. Facial Expression Recognition using Convolutional Neural Networks: State of the Art. Computer Vision Lab, TU Wien. https://arxiv.org/pdf/1612.02903.pdf

[12] .OpenCV open source library for image feature extraction, https://opencv.org/2013

[13] End-to-End Multimodal Emotion Recognition using Deep Neural Networks, https://arxiv.org/pdf/1704.08619.pdf

[14] Agrawal, A., & An, A. (2012). Unsupervised emotion detection from text using semantic and syntactic relations. In International Conference on Web Intelligence and Intelligent Agent Technology.

[15] Al-Hajjar, D., & Syed, A. Z. (2015). Applying sentiment and emotion analysis on brand tweets for digital marketing. In Applied Electrical Engineering and Computing Technologies. IEEE.

[16] Alm, C. O., Roth, D., & Sproat, R. (2005). Emotions from text. In Human Language Technology and Empirical Methods in Natural Language Processing (pp. 579–586).

[17] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," Neural Networks, vol. 64, pp. 59–63, 2015..

[18] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 37, no. 6, pp. 1113–1133, 2015.

[19] M. V. B. Martinez, "Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition," in Advances in Face Detection and Facial Image Analysis. Springer, 2016, pp. 63–100.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1097–1105.

[21] K. He, X. Zhang, haoqing Ren, and J. Sun, "Deep Residual Learning for Image Recognition," CoRR, vol. 1512, 2015.

[22] A. Dhall, O. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon, "Video and Image Based Emotion Recognition Challenges in the Wild: EmotiW 2015," in ACM International Conference on Multimodal Interaction (ICMI), 2015, pp. 423–426.

[23] Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., Jha, S., Agrawal, M., Dziurzynski, L. A., Sap, M., Weeg, C., Larson, E. E., Ungar, L. H., & Seligman, M. E. P. (2015). Psychological language on Twitter predicts county-level heart disease mortality. Psychological Science, 26 , 159–169.

[24] Danisman, T., & Alpkocak, A. (2008). Feeler: Emotion classification of text using vector space model. In Communication, Interaction and Social Intelligence. volume 1.

[25] Chitturi, R., Raghunathan, R., & Mahajan, V. (2007). Form versus function: How the intensities of specific emotions evoked in functional versus hedonic trade-offs mediate product preferences. Journal of Marketing Research, 44 , 702–714.

[26] Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., & Ricci-Bitti, P. E. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. Journal of Personality and Social Psychology, 53 , 712–717.

9.2 Certificates

Research Paper Certificate:

ISSN: 2582-3930
Impact Factor: 8.448

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT
An Open Access Scholarly Journal || Index in major Databases & Metadata

CERTIFICATE OF PUBLICATION

International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to

Rohan Kadu

in recognization to the publication of paper titled

TalkSpace - Advanced Emotion Recognition and Personality Analysis Platform

published in IJSREM Journal on Volume 08 Issue 04 April, 2024

www.ijsrem.com
Editor-in-Chief
IJSREM Journal
ijsremjournal@gmail.com

ISSN: 2582-3930
Impact Factor: 8.448

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT
An Open Access Scholarly Journal || Index in major Databases & Metadata

CERTIFICATE OF PUBLICATION

International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to

Krishnakant Gangurde

in recognization to the publication of paper titled

TalkSpace - Advanced Emotion Recognition and Personality Analysis Platform

published in IJSREM Journal on Volume 08 Issue 04 April, 2024

www.ijsrem.com
Editor-in-Chief
IJSREM Journal
ijsremjournal@gmail.com


IJSREM
e-Journal
ISSN: 2582-3930
Impact Factor: 8.448
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT
An Open Access Scholarly Journal || Index in major Databases & Metadata
CERTIFICATE OF PUBLICATION
International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to
Vaishnav Rajput
in recognization to the publication of paper titled
TalkSpace - Advanced Emotion Recognition and Personality Analysis Platform
published in IJSREM Journal on Volume 08 Issue 04 April, 2024
www.ijsrem.com
Editor-in-Chief
IJSREM Journal
ijsremjournal@gmail.com


IJSREM
e-Journal
ISSN: 2582-3930
Impact Factor: 8.448
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT
An Open Access Scholarly Journal || Index in major Databases & Metadata
CERTIFICATE OF PUBLICATION
International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to
Prof. Samuel Jacob
in recognization to the publication of paper titled
TalkSpace - Advanced Emotion Recognition and Personality Analysis Platform
published in IJSREM Journal on Volume 08 Issue 04 April, 2024
www.ijsrem.com
Editor-in-Chief
IJSREM Journal
ijsremjournal@gmail.com

Project Competition:

ATHARVA COLLEGE OF ENGINEERING
(APPROVED BY AICTE, RECOGNIZED BY GOVERNMENT OF MAHARASHTRA
& AFFILIATED TO UNIVERSITY OF MUMBAI - ESTD. 1999 - 2000)
ISO 21001: 2018 ISO 14001 : 2015 ISO 9001 : 2015 ACE/PROJECTATHON 2.0/CMPN/FR/—/23-24
NAAC ACCREDITED A+

DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE OF PARTICIPATION

This is to certify that *Rohan Kadu* from *Vidyalankar Institute of Technology* has participated in the PROJECTATHON 2.0 NATIONAL LEVEL PROJECT COMPETITION 2024, held on 19th April 2024 organized by Computer Engineering Department, Atharva College of Engineering,Mumbai

Dr. Suvarna Pansambal
HOD,CMPN

Dr. Ramesh Kulkarni
Principal,ACE

ATHARVA COLLEGE OF ENGINEERING
(APPROVED BY AICTE, RECOGNIZED BY GOVERNMENT OF MAHARASHTRA
& AFFILIATED TO UNIVERSITY OF MUMBAI - ESTD. 1999 - 2000)
ISO 21001: 2018 ISO 14001 : 2015 ISO 9001 : 2015 ACE/PROJECTATHON 2.0/CMPN/FR/—/23-24
NAAC ACCREDITED A+

DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE OF PARTICIPATION

This is to certify that *Krishnakant Gangurde* from *Vidyalankar Institute of Technology* has participated in the PROJECTATHON 2.0 NATIONAL LEVEL PROJECT COMPETITION 2024, held on 19th April 2024 organized by Computer Engineering Department, Atharva College of Engineering,Mumbai

Dr. Suvarna Pansambal
HOD,CMPN

Dr. Ramesh Kulkarni
Principal,ACE

ATHARVA COLLEGE OF ENGINEERING
(APPROVED BY AICTE, RECOGNIZED BY GOVERNMENT OF MAHARASHTRA
& AFFILIATED TO UNIVERSITY OF MUMBAI - ESTD. 1999 - 2000)
ISO 21001: 2018 ISO 14001 : 2015 ISO 9001 : 2015 ACE/PROJECTATHON 2.0/CMPN/FR/—/23-24
NAAC ACCREDITED A+

DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE OF PARTICIPATION

This is to certify that *Vaishnav Rajput* from *Vidyalankar Institute of Technology* has participated in the PROJECTATHON 2.0 NATIONAL LEVEL PROJECT COMPETITION 2024, held on 19th April 2024 organized by Computer Engineering Department, Atharva College of Engineering,Mumbai

Dr. Suvarna Pansambal
HOD,CMPN

Dr. Ramesh Kulkarni
Principal,ACE

9.3 Github Link

https://github.com/devil2003cyber/TalkSpace

9.4 Plagiarism Report:

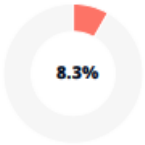
Plagiarism and AI Content Detection Report

BlackBook.pdf

Scan details

Scan time:	Total Pages:	Total Words:
April 28th, 2024 at 19:53 UTC	27	6705

Plagiarism Detection



Types of plagiarism		Words
Identical	1.6%	104
Minor Changes	0%	0
Paraphrased	0%	0
Omitted Words	81.4%	5455

Plagiarism Results: (7)

SOLUTION: Final report - Studypool — 8.3%

https://www.studypool.com/documents/24351429/final-report

Post a Question Provide details on what you need help with along with a budget and time limit. Questions are posted...

MTech_Thesis.doc — 8.3%

https://www.nitmz.ac.in/computer%20science%20and%20engineering/mtech_thesis.doc

[image: image1.jpg] (The title is in Times New Roman Font with 18 to 22-point size, Bold, one-and-a-half line spacing) A Thesis Submitted...

specimen.doc — 8.3%

https://oldweb.nita.ac.in/nitamain/academics/specimen.doc

[image: image1.png]Specimen A (The title is in Times New Roman Font with 18 to 22-point size, Bold, one-and-a-half line spacing) [image...

Declaration.tex — 8.2%

https://www.ee.iitb.ac.in/course/~shalini/reportmtech/declaration.tex

\chapter*{Declaration of Academic Ethics} \thispagestyle{empty} \doublespace % {\LARGE \bf \center Declaration of Academic Ethics} % ...

Certified by Copyleaks

About this report
help.copyleaks.com

copyleaks.com