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.

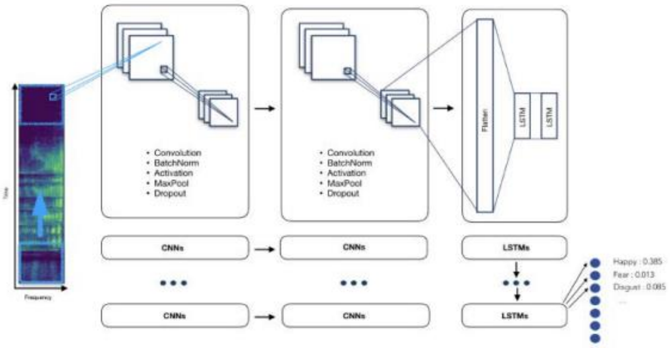


Fig. 1: Time distributed CNN

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.

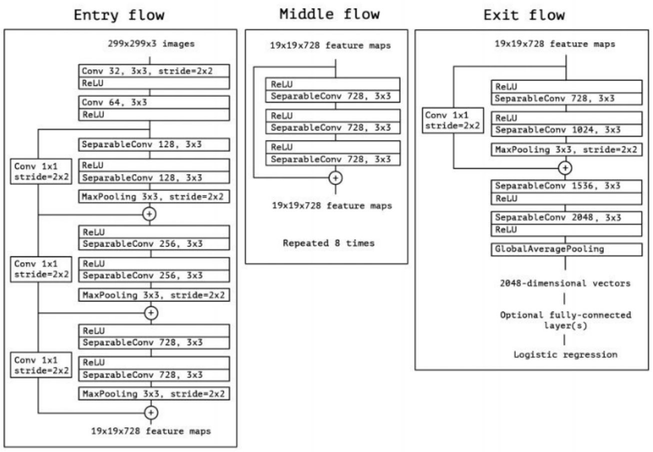


Fig. 2: XCeption Structure

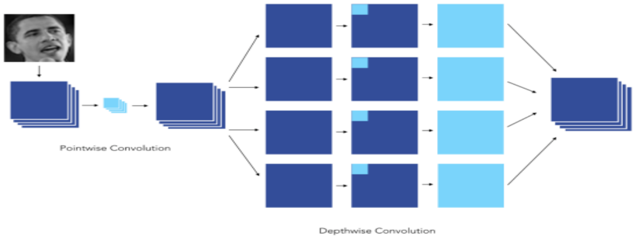


Fig. 3: XCeption

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:

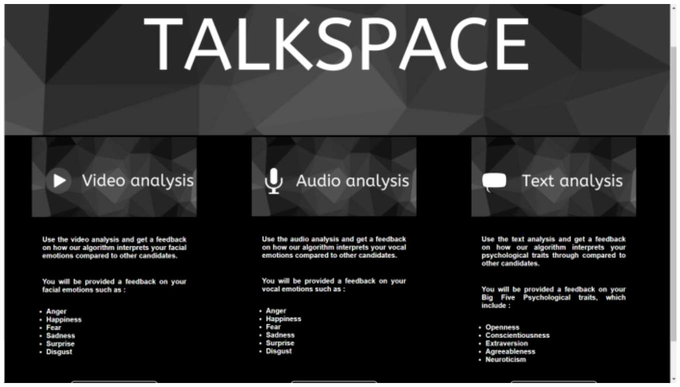
Utilizing a Time Distributed CNN, we analyze audio signals directly from log-mel-spectrograms. This model combines CNNs with LSTM networks to capture temporal patterns and dependencies for accurate emotion prediction.

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.



TALKSPACE

Use the video analysis and get a feedback on how our algorithm interprets your facial emotions compared to other candidates.

You will be provided a feedback on your facial emotions such as :

- Anger
- Happiness
- Fear
- Sadness
- Surprise
- Disgust

Use the audio analysis and get a feedback on how our algorithm interprets your vocal emotions compared to other candidates.

You will be provided a feedback on your vocal emotions such as :

- Anger
- Happiness
- Fear
- Sadness
- Surprise
- Disgust

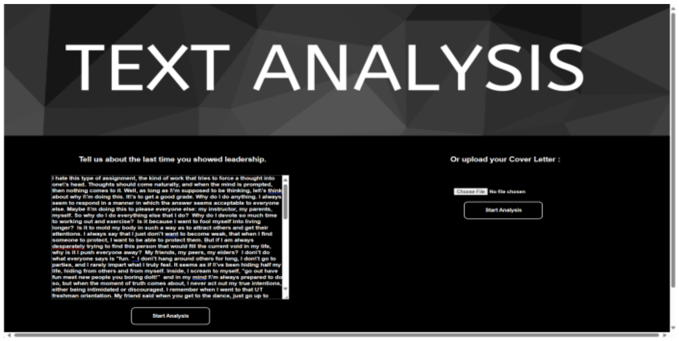
Use the text analysis and get a feedback on how our algorithm interprets your psychological traits through compared to other candidates.

You will be provided a feedback on your Big Five Psychological traits, which include :

- Openness
- Conscientiousness
- Extraversion
- Agreeableness
- Neuroticism

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)



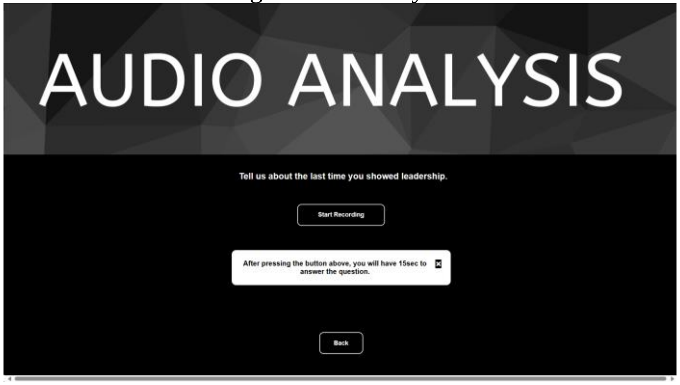
TEXT ANALYSIS

Tell us about the last time you showed leadership.

I hate this type of assignment, the kind of work that tries to force a thought into one\'s head. Thoughts should come naturally, and when the mind is prompted, then nothing comes to it. Well, as long as I\'m supposed to be thinking, let\'s think about why I\'m doing this. It\'s to get a good grade. Why do I do anything. I always seem to respond in a manner in which the answer seems acceptable to everyone else. Maybe I\'m doing this to please everyone else: my instructor, my parents, myself. So why do I do everything else that I do? Why do I devote so much time to working out and exercise? Is it because I want to fool myself into living longer? Is it to mold my body in such a way as to attract others and get their attentions. I always say that I just don\'t want to become weak, that when I find someone to protect, I want to be able to protect them. But if I am always desparately trying to find this person that would fill the current void in my life, why is it I push everyone away? My friends, my peers, my elders? I don\'t do what everyone says is "fun. " I don\'t hang around others for long, I don\'t go to parties, and I rarely impart what I truly feel. It seems as if I\'ve been hiding half my life, hiding from others and from myself. Inside, I scream to myself, "go out have fun meet new people you boring dolt!" and in my mind I\'m always prepared to do so, but when the moment of truth comes about, I never act out my true intentions, either being intimidated or discouraged. I remember when I went to that UT freshman orientation. My friend said when you get to the dance, just go up to

Or upload your Cover Letter :

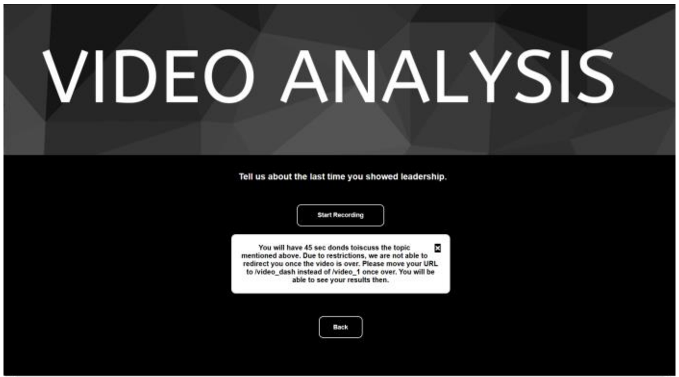
Fig. 5: Text Analysis



AUDIO ANALYSIS

Tell us about the last time you showed leadership.

After pressing the button above, you will have 15sec to answer the question.



VIDEO ANALYSIS

Tell us about the last time you showed leadership.

You will have 45 sec donds toiscuss the topic mentioned above. Due to restrictions, we are not able to redirect you once the video is over. Please move your URL to /video_dash instead of /video_1 once over. You will be able to see your results then.

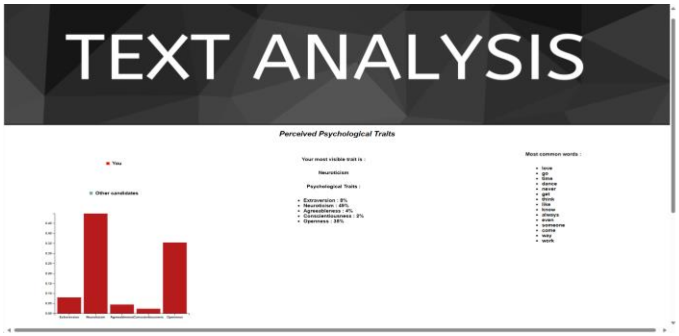
TEXT ANALYSIS

Tell us about the last time you showed leadership.

I hate this type of assignment, the kind of work that tries to force a thought into one\'s head. Thoughts should come naturally, and when the mind is prompted, then nothing comes to it. Well, as long as I\'m supposed to be thinking, let\'s think about why I\'m doing this. It\'s to get a good grade. Why do I do anything. I always seem to respond in a manner in which the answer seems acceptable to everyone else. Maybe I\'m doing this to please everyone else: my instructor, my parents, myself. So why do I do everything else that I do? Why do I devote so much time to working out and exercise? Is it because I want to fool myself into living longer? Is it to mold my body in such a way as to attract others and get their attentions. I always say that I just don\'t want to become weak, that when I find someone to protect, I want to be able to protect them. But if I am always desparately trying to find this person that would fill the current void in my life, why is it I push everyone away? My friends, my peers, my elders? I don\'t do what everyone says is "fun. " I don\'t hang around others for long, I don\'t go to parties, and I rarely impart what I truly feel. It seems as if I\'ve been hiding half my life, hiding from others and from myself. Inside, I scream to myself, "go out have fun meet new people you boring dolt!" and in my mind I\'m always prepared to do so, but when the moment of truth comes about, I never act out my true intentions, either being intimidated or discouraged. I remember when I went to that UT freshman orientation. My friend said when you get to the dance, just go up to

Or upload your Cover Letter :

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.



TEXT ANALYSIS

Perceived Psychological Traits

Your most visible trait is :

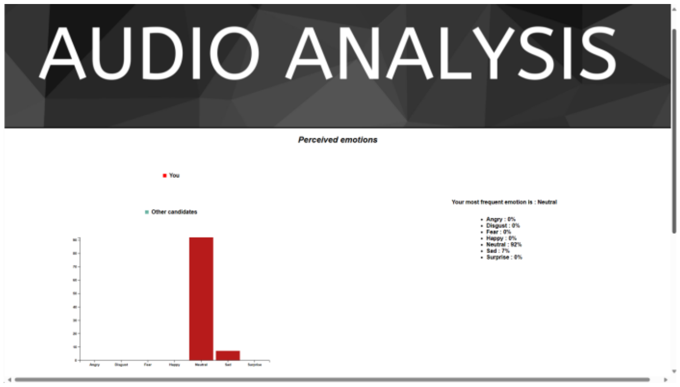
Neuroticism

Psychological Traits :

- Extraversion : 8%
- Neuroticism : 48%
- Agreeableness : 4%
- Conscientiousness : 2%
- Openness : 38%

Most common words :

- love
- go
- time
- dance
- never
- get
- think
- like
- know
- always
- even
- someone
- come
- way
- work

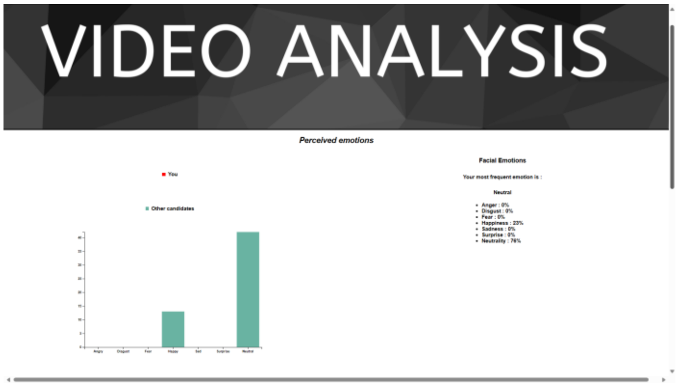


AUDIO ANALYSIS

Perceived emotions

Your most frequent emotion is : Neutral

- Angry : 0%
- Disgust : 0%
- Fear : 0%
- Happy : 0%
- Neutral : 92%
- Sad : 7%
- Surprise : 0%



VIDEO ANALYSIS

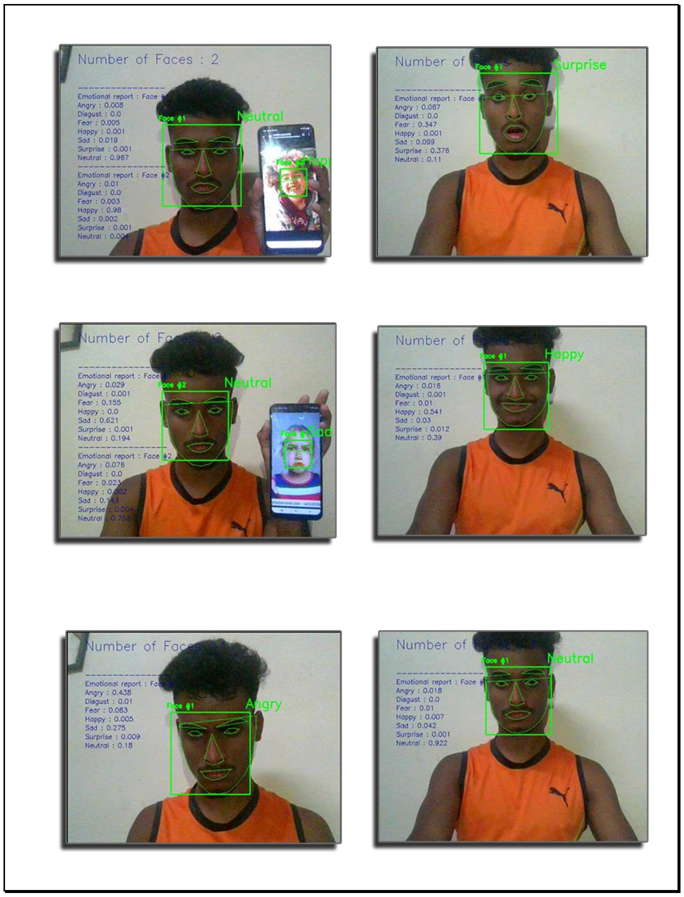
Perceived emotions

Facial Emotions

Your most frequent emotion is :

Neutral

- Anger : 0%
- Disgust : 0%
- Fear : 0%
- Happiness : 23%
- Sadness : 0%
- Surprise : 0%
- Neutrality : 76%



```
Number of Faces : 2
-----------------
Emotional report : Face #1
Angry : 0.008
Disgust : 0.0
Fear : 0.005
Happy : 0.001
Sad : 0.019
Surprise : 0.001
Neutral : 0.967
-----------------
Emotional report : Face #2
Angry : 0.01
Disgust : 0.0
Fear : 0.003
Happy : 0.98
Sad : 0.002
Surprise : 0.001
Neutral : 0.004
```

```
Number of
-----------------
Emotional report : Face
Angry : 0.067
Disgust : 0.0
Fear : 0.347
Happy : 0.001
Sad : 0.099
Surprise : 0.376
Neutral : 0.11
```

```
Number of Faces : 2
-----------------
Emotional report : Face
Angry : 0.029
Disgust : 0.001
Fear : 0.155
Happy : 0.0
Sad : 0.621
Surprise : 0.001
Neutral : 0.194
-----------------
Emotional report : Face #2
Angry : 0.078
Disgust : 0.0
Fear : 0.023
Happy : 0.002
Sad : 0.143
Surprise : 0.004
Neutral : 0.752
```

```
Number of
-----------------
Emotional report : Face
Angry : 0.018
Disgust : 0.001
Fear : 0.01
Happy : 0.541
Sad : 0.03
Surprise : 0.012
Neutral : 0.39
```

```
Number of Fac
-----------------
Emotional report : Fa
Angry : 0.438
Disgust : 0.01
Fear : 0.083
Happy : 0.005
Sad : 0.275
Surprise : 0.009
Neutral : 0.18
```

```
Number of
-----------------
Emotional report : Face
Angry : 0.018
Disgust : 0.0
Fear : 0.01
Happy : 0.007
Sad : 0.042
Surprise : 0.001
Neutral : 0.922
```

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.