Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

CREATION AND ANALYSIS OF MULTIMODAL EMOTION RECOGNITION CORPUS WITH INDIAN ACTORS

Komal Anadkat[1], Dr. Hiteishi Diwanji[2], Dr. Shahid Modasiya[3] , Mihir Mehta[4]

•

[1]Assistant Professor, Information Technology department, G.E.C., Gandhinagar, Gandhinagar, India. komalanadkat@gecg28.ac.in

[2] Professor, Information Technology department, L.D. College of Engineering, Ahmedabad, India. hiteishi.diwanji@gmail.com

[3]Assistant Professor, Electronics & Communication department, G.E.C., Gandhinagar, Gandhinagar, India. shahid@gecg28.ac.in

[4]Assistant Professor, Computer Engineering department, G.E.C., Gandhinagar, Gandhinagar, India. mihir_mehta@gecg28.ac.in

Abstract

Emotion recognition plays an important role in many real-life application areas of artificial intelligence like human-computer interactions, autism detection, stress and depression detection, measuring mental health, and suicide prevention. Emotion state of a person can be decided by the facial expression, tone of voice, words of speech, and body gestures when they are having a face-to-face conversation. People widely use social media platforms to post their feelings and mood through status. So, the status text can be used to identify the emotional state of a person. Physiological signals (EEG, ECG, and EDA) can identify the emotional state more accurately as people cannot be faked during the data collection but it is difficult to collect data. Many unimodal and multimodal datasets are publicly available but still, there is a strong need to create a multimodal dataset that consists of all the important modalities for the identification of emotional state. In this paper, first, we have reviewed all the available unimodal and multimodal datasets, then in the next section, we discuss the method to prepare the multimodal dataset. The data of four different modalities like facial expressions, audio, social media text, and EEG have been collected from seven different actors of different age groups and of different demographic regions. The dataset is non-spontaneous and contains discrete emotion labels like happy, sad, and angry. The procedure to create a dataset of different modalities include steps like capturing data, pre-processing, feature extraction, and storing to the relevant format. In last, to observe effect of different emotions, analysis of proposed multimodal database is carried out using efficient image, speech and text parameters.

Keywords: Unimodal, Emotion recognition, Feature extraction, Multimodal, EEG, Modality.

I. Introduction

Human emotion, sentiment, and feelings [1], emotion identification, and sentiment analysis all fall under the umbrella category of human-computer interaction. It has guided research on computers' abilities to recognize and convey emotions, respond intelligently to human feelings, and manage and exploit human emotions [2]. Nowadays it is highly desirable to build a system that can recognize and understand the emotions of a person, and respond in a way like a human [3]. For instance, if the driver's emotional state can be monitored and appropriate responses generate in a mart vehicle

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

system, then observed findings might effectively minimize the risk of accidents [4]. So, many authors have done significant research and concluded that emotion recognition is the key to promoting human-machine interaction, AI, and many other research fields.

Many Datasets are publicly available for emotion recognition, some of them are unimodal like visual, audio, text, and psychological. Some recent advancements have been done in the field of creating multimodal datasets like audio-textual-visual databases and video-physiological databases. People use to express their emotions through facial expression, speech intensity, and social media status, thus multimodal signals. So, many researchers have developed multimodal databases by collecting on-the-spot emotions, gathering online data, or inducing emotions through videos. Most of these datasets are either multi-physical or physiological. Nowadays people use social media platforms to express their feeling mostly. We have observed that social media status can be used to classify emotions effectively.

We have studied all the available unimodal and multimodal emotion recognition datasets but there is no dataset exist which contains all important dimensions for classification.so, In this paper, We have prepared the multimodal emotion recognition database of 7 demographically different actors which consists of their facial expression, audio, social media status, and EEG signal for three classification categories happy, sad, and angry.

II. Existing Databases for Emotion Recognition

The datasets available for the Emotion recognition task are mainly classified into facial expression, audio, Text, and physiological signals. Some available datasets are multimodal which consist of more than one dimension of emotion recognition. Multimodal datasets are required to make a robust model which identifies the human emotions of different categories.

I. Facial expression databases

Many authors have tried to recognize emotions using facial expressions but people can be faked expressing emotion. So, the dataset used to train the model should be sparse and not limited. The early facial expression databases were created using emotions that were purposefully expressed by people in the lab.

JAFFE [5] is a facial expression dataset, which has 213 images of 256*256 pixels for 7 different emotions (6 basic facial expressions and neutral). The Actors are 10 Japanese female models who acted to create the dataset.

Cohn-Kanade (CK+) [6], which is an extension of CK [7], the dataset contains 593 video sequences from 123 different subjects who were instructed to perform 7 facial expressions (anger, contempt, disgust, fear, happiness, sadness, and surprise). This dataset is lab-controlled yet very extensive to provide comparatively good results for emotion recognition.

Oulu-CASIA [8], the dataset consists of 6 expressions (surprise, happiness, sadness, anger, fear, and disgust) from 80 Subjects and includes 2,880 image sequences captured with one of two kinds of imaging systems. Subjects were asked to make a facial expression according to an expression example shown in picture sequences. The imaging hardware works at the rate of 25 frames per second and the image resolution is 320 × 240 pixels.

BP4D [9], the dataset is a well-annotated 3D video database of spontaneous facial expressions, collected from 41 participants (23 women, 18 men), demographically from a different region. Eight tasks were covered with an interview process and a series of activities to elicit eight emotions.

4DFAB 10] , the dataset consists of at least 1,800,000 dynamic high-resolution 3D faces captured from 180 subjects in four different sessions spanning over a five-year period.

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

So far, the acted databases are listed which were constructed in a particular environment. On the other way, we can create a dataset by collecting images or videos from the internet and it is called In-the-Wild.

FER2013 [11], is a large-scale dataset consists 35,887 gray images with 48×48 pixels, collected automatically through the Google image search API. The 07 emotion categories were collected and labeled 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral. SFEW 2.0 [12], is an in-the-wild dataset that is divided into three sets, including Train (891 images) and Val (431 images), labeled as one of six basic expressions (anger, disgust, fear, happiness, sadness, and surprise), as well as the neutral and Test (372 images) without expression labels. EmotioNet [13], dataset consists of one million images with 950,000 automatically annotated AUs and 25,000 manually annotated AUs. AffectNet [14], the dataset contains over 1,000,000 facial images, of which 450,000 images are manually annotated with eight discrete expressions (six basic expressions plus neutral and contempt), and the dimensional intensity of valence and arousal.

II. Speech/Audio databases

There are main two categories of speech datasets: - induced and spontaneous. Induced datasets were created from the professional actor's performance and they have to act in a particular environment and record the dataset which is more authentic.

Berlin Database of Emotional Speech (Emo-DB) [15], the dataset contains about 500 utterances spoken by 10 actors (5 men and 5 women) in a happy, angry, anxious, fearful, bored, and disgusted way.

Belfast Induced Natural Emotion (Belfast) [16], the dataset was recorded from 40 subjects at Queen University in Northern Ireland, UK. Each subject took part in five tests, each of which contained short video recordings 9 (5 to 60 seconds in length) with stereo sound, and related to one of the five emotional tendencies: anger, sadness, happiness, fear, and neutrality.

Ryerson Audio-Visual Database of Emotional Speech and Song (Ravdess), the dataset contains 7356 files, recorded by 24 actors (12 males, 12 females). Each actor vocalized two statements in North American accent. Speech includes calm, happy, sad, angry, fearful, surprise, and disgust expressions, and the song contains calm, happy, sad, angry, and fearful emotions.

III. Textual (Social media) databases

Text data in various levels (e.g., word, phrase, and document) is tagged with emotion or sentiment tags, such as positive, negative, emphatic, general, sad, glad, and so on, in databases for textual emotion analysis.

The multi-domain sentiment (MDS) [17,18], dataset consists of more than 100,000 sentences, which are product reviews acquired from Amazon.com. These sentences are labeled with both two sentiment categories (positive and negative) and five sentiment categories (strong positive, weak positive, neutral, weak negative, and strong negative).

IMDB [19], is a popularly used largest dataset that provides 25,000 highly polar movie reviews for training and 25,000 for testing.   The first line in each file contains headers that describe what is in each column.

Stanford sentiment treebank (SST) [20] is the semantic lexical database annotated by Stanford University. It includes a fine-grained emotional label of 215,154 phrases in a parse tree of 11,855 sentences, and it is the first corpus with fully labelled parse trees.

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

IV. EEG databases

Physiological signals cannot be altered intentionally to hide the emotions which is normally happens with facial expressions, audio and textual emotions. So it is more authentic and reliable. It consists EEG,ECG, EMG and RESP data. EEG is a signal which captures the brain signals so it is highly desirable to use in the task of emotion recognition.

DEAP [21], comprises a 32-channel EEG, a 4-channel EOG, a 4-channel EMG, RESP, plethysmograph, Galvanic Skin Response (GSR), and body temperature, from 32 subjects. Each subject participated in 40 EEG trials, in each of which a specific emotion was elicited by a music video. Immediately after watching each video, subjects were required to rate their truly-felt emotion from five dimensions: valence, arousal, dominance, liking, and familiarity.

SEED [22, 23] is based only on EGG recordings, from 15 subjects. However, it enables repeated experiences to improve data reliability/stability. In their study, participants were asked to experience three EEG recording sessions, with an interval of two weeks between two successive recording sessions. Within each session, each subject was exposed to the same sequence of fifteen movie excerpts, each one approximately four-minute-long, to induce three kinds of emotions: positive, neutral, and negative.

AMIGOS [24] was designed to collect participants' emotions in two social contexts: individual and group. AMIGOS was constructed in 2 experimental settings. First, 40 participants watched 16 short emotional videos. Then, they watched 4 long videos, including a mix of lone and group sessions. These emotions were annotated with not only self-assessment of affective levels but also external assessment of valence and arousal, through GSR and ECG signals.

Wearable devices help to bridge the gap between lab studies and real-life emotions. Wearable Stress and Affect Detection (WESAD) [25] was constructed for stress detection, providing multimodal, high-quality data, including three different affective states (neutral, stress, amusement).

IV. Multimodal databases

Humans mostly express their feelings in through multimodal signals. So, rather than focusing only on a single modal like audio, video, text, or EEG, it is desirable to create a multimodal dataset by collecting spontaneous emotions from available online data.

Interactive Emotional Dyadic Motion Capture (IEMOCAP) [26] is constructed by the Speech Analysis and Interpretation Laboratory. During recording, 10 actors are asked to not only perform selected emotional scripts but also improvised hypothetical scenarios designed to elicit 5 specific types of emotions. The face, head, and hands of actors are marked to provide detailed information about their facial expressions and hand movements while performing.

Harvesting Opinions from the Web database (HOW) [27] is the first publicly available database containing visual, audio, and textual modalities for sentiment analysis. HOW consists of 13 positive, 12 negative, and 22 neutral videos captured from YouTube.

ICT-MMMO (Multimodal Movie Opinion) [28], the dataset consists of 308 YouTube videos and 78 movie review videos from ExpoTV. The dataset has five emotion categories named strongly positive, weakly positive, neutral, strongly negative, and weakly negative.

CMU-MOSEI (Multimodal Opinion Sentiment and Emotion Intensity)[29], is the largest dataset which consists of 23,453 sentences and 3,228 videos from more than 1,000 online YouTube speakers. Each video contains a manual transcription that aligns audio and phoneme grade.

Remote Collaborative and Affective Interactions (RECOLA) [30], the dataset consists multimodal corpus of spontaneous interactions from 46 participants. These participants worked in pairs to discuss a disaster scenario escape plan and came to an agreement via remote video conferencing.

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

The recordings of the participants' activities were labeled by 6 annotators with two continuous emotional dimensions: arousal and valence, as well as social behavior labels on five dimensions.

III. Dataset Design and Acquisition

Emotion Recognition and prediction is indeed a challenging task, even though many researchers have done significant work in this area. There are many multimodal datasets available for emotion recognition but still there is a need to develop extended multimodal dataset, which include physical and psychological data. In the next section, we discusses the procedure followed in the acquisitions of multimodal facial expression, audio, social media status and EEG data of the subjects for the proposed dataset. The actors depict the wider range 'happy', 'sad', and 'anger' feelings. These discrete feelings of humans are universally accepted as basic human emotions.
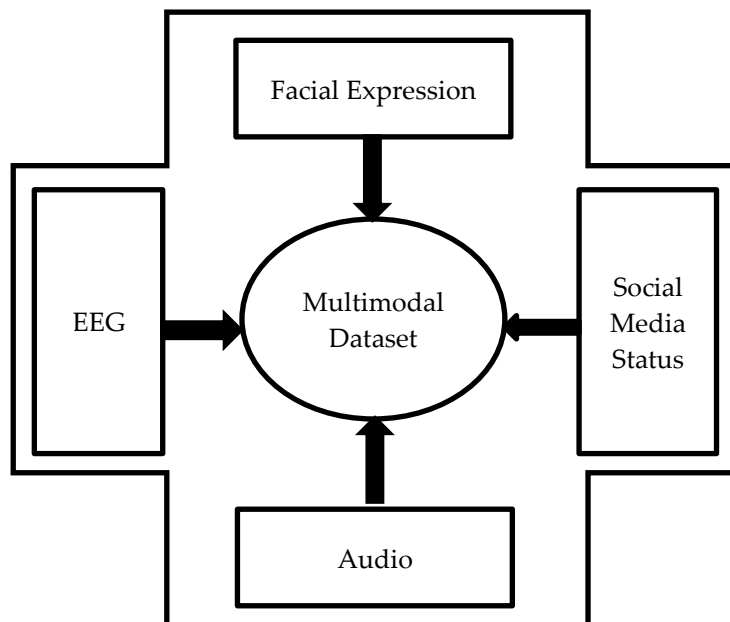

**Figure 1:** *Multimodal Emotion Recognition dataset dimensions*

I. Actor Details

The 07 Professors of Government Engineering College, Gandhinagar have been participated in the acting sessions to create this dataset. The facial expression, voice, social media status and EEG signals are collected for three basic emotions like happy, sad and angry. Actors' mean age = 39.5 years, age range = 33-45, males = 2, and females = 5. Actors are the basically belongs from Rajasthan (1 actor), North Gujarat (2 actors), South Gujarat (1 actor), Saurashtra (2 actors) and Central Gujarat (1 actor). All actors speak English as a foreign language.

II. Facial Expression Dataset

The facial expression dataset was collected for three different emotions happy, sad and angry by 07 different actors. The FER-13 dataset samples are demonstrated to actors for better understanding of the requested target emotions. The dataset FER-2013 contains 35,887 grayscale images of faces with 48*48 pixels and stored in CSV format. As shown in the figure 2, the images of three different facial

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

expression happy, sad and angry are captured using RedMe 9 prime phone for all 7 actors. Then, the collected images have been converted into 48*48 grayscale images and the pixel values of these images are stored in .csv file.
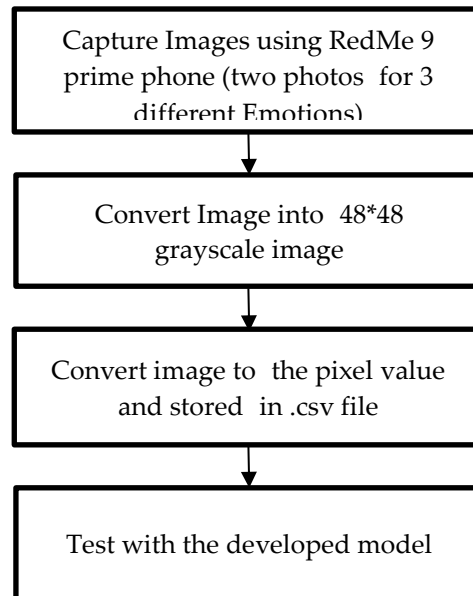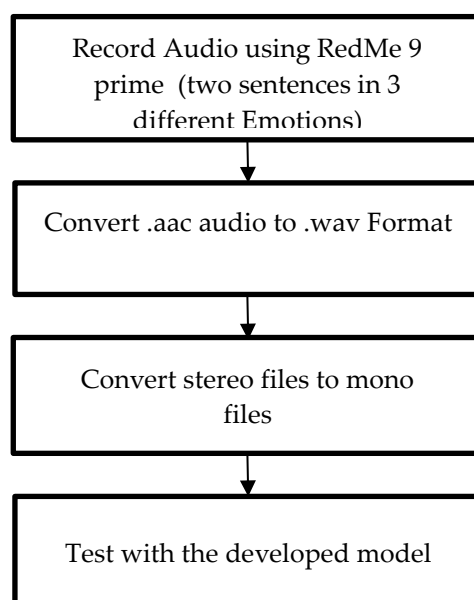

Figure 2: *Procedure in the creation of Facial Expression dataset*

Figure 3*: Actors performed in the proposed Facial expression dataset*

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023
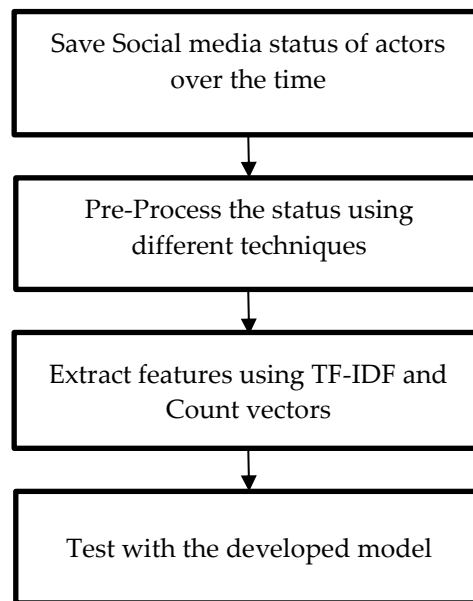
III. Audio Dataset

The Audio dataset was collected for three different emotions happy, sad and angry by 07 different actors. The RAVDESS dataset samples are demonstrated to actors for better understanding of the requested target emotions. The two sentences are given and explained to each actor. They are requested to practice script sentences to evoke the target feelings. The RAVDESS dataset contains 7356 files. The database contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements in a neutral North American accent. As shown in figure 4 ,The recording of all actors have been done by RedMe 9 prime phone. Two sentences Dogs are sitting by the door and Kids are talking by the door are recited by the actors in all three emotions. Then recorded audios are converted to .wav format and stereo files are converted to mono files.

Figure 4: *Procedure in the creation of Audio dataset*

IV. Social Media Status Dataset

The Social media status dataset was collected for three different emotions happy, sad and angry by considering status of 07 different actors. Three basic emotion status sad, happy and angry has been scraped and prepared which is available on kaggle. Every data set contains two columns one the status and another sentiment of that status. So, we pre-process our collected samples accordingly. As shown in Figure 5. ,we have developed a system that could find the emotion label like happy, sad, and anger for any piece of text, especially status and stories on social media with their probability for each emotion using its textual features, Preprocessing techniques, and various machine learning and deep learning algorithms[31]. Samples of collected status from different actors are listed with sentiments in table 1.

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

```
┌─────────────────────────────┐
│   Save Social media status  │
│    of actors over the time  │
└─────────────────────────────┘
               │
               ▼
┌─────────────────────────────┐
│   Pre-Process the status     │
│   using different techniques │
└─────────────────────────────┘
               │
               ▼
┌─────────────────────────────┐
│   Extract features using     │
│   TF-IDF and Count vectors   │
└─────────────────────────────┘
               │
               ▼
┌─────────────────────────────┐
│   Test with the developed    │
│          model               │
└─────────────────────────────┘
```

Figure 5: *Procedure in the creation of Social Media Status dataset*
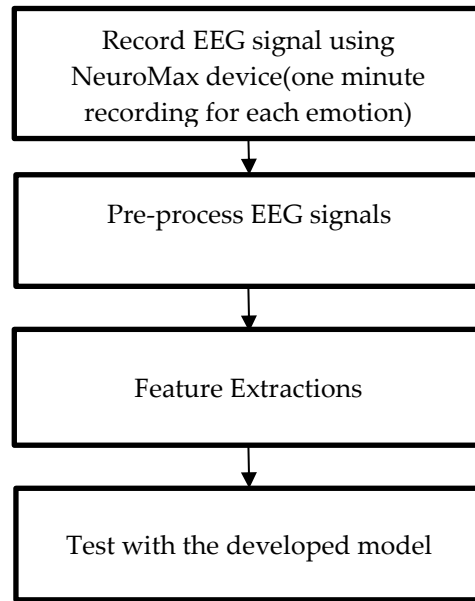
Table 1: *Samples collected from different Actors*

Social Media Status	Sentiment
Happiness is where we find it, but very rarely where we seek it.	Happy
Beware, I'm not in my greatest mood today.	Angry
My silence is just another word for my pain.	Sad
Happiness is when what you think, what you say, and what you do are in harmony.	Happy

V. EEG Dataset

EEG is an electrophysiological monitoring method to record the electrical activity of the brain. We have collected EEG data from 07 people while they are made to concentrate on a particular thought. We have used NeuroMax Nmx-32(channels) series portable device for the collection of data. The subjects were asked to visualize the incident which can make them feel happy, sad, and angry respectively for 60 seconds each. The other activities like eyes blinking, movement, and eyes open are recorded in the data set. The recorded signals contain 32 electrodes data, so the important electrodes for emotion recognition were identified and then prepared data for the analysis task.

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

Figure 6: *Procedure in the creation of EEG dataset*

Figure 7: Actors performed in the proposed EEG dataset

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

IV. Experiment Results

The multimodal emotion recognition dataset contains image, audio, social media text and EEG signal. In this paper, different evaluation parameters are used for different modal. Complete evaluation is performed in Python.

I. Facial Expression Dataset Evaluation

In this paper, Blind/reference less image spatial quality evaluator (BRISQUE) assessment method used to evaluate image quality of facial expression dataset, which only uses the image pixels to calculate features. BRISQUE relies on spatial Natural Scene Statistics (NSS) model of locally normalized luminance coefficients in the spatial domain, as well as the model for pairwise products of these coefficients. First, the MSCN coefficients and the pairwise products are calculated and then the distribution is verified using plot which is shown in Figure 8.

Then pre-trained SVR model is used to calculate the quality assessment. However, in order to have good results, we scaled the features to [-1, 1]. The scale used to represent image quality goes from 0 to 100. An image quality of 100 means that the image's quality is very bad. In the case of the analyzed image, we get that it is a good quality image as the image quality score is 3.889.
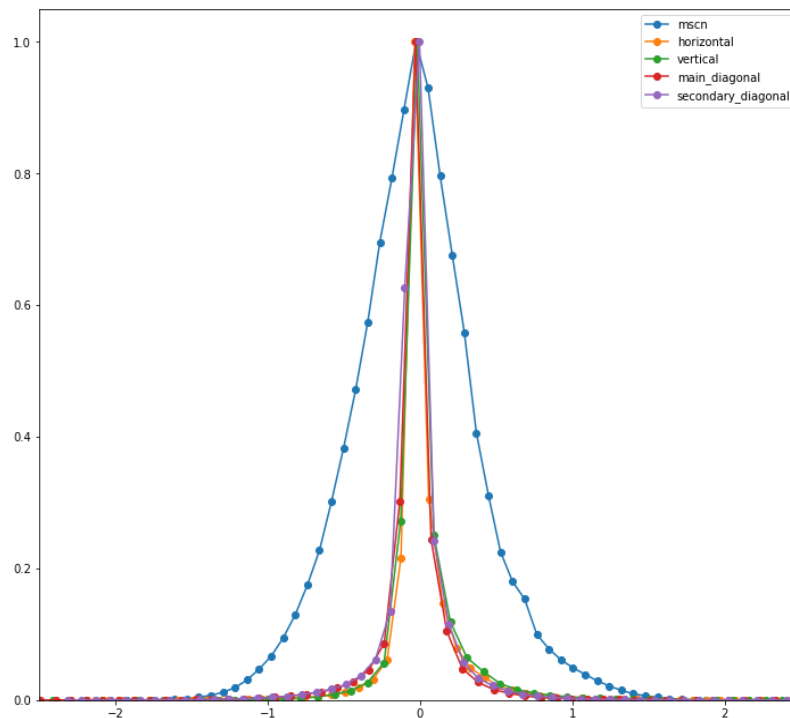

Figure 8. *Plot of image coefficient distribution*

```
%%time

calculate_image_quality_score(brisque_features)
```

```
CPU times: user 10.8 ms, sys: 0 ns, total: 10.8 ms
Wall time: 11.3 ms
3.8893297529106974
```

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

II. Audio Dataset Evaluation

Speech signal is a denouement of time varying vocal tract system agitated by the time varying excitation source signal. In this paper, the converted audio signal is observed by different parameters shown below. Figure 9 is the Plot of amplitude envelope of a sample waveform. A Mel spectrogram is a spectrogram where the frequencies are converted to the Mel scale is show in in Figure 10. The FFT is computed on overlapping windowed segments of the signal, and we get what is called the spectrogram which is shown in Figure 11.The fast Fourier transform (FFT) is an algorithm that can efficiently compute the Fourier transform which is shown in Figure 12. MFCC and the creation of filter banks are all motivated by the nature of audio signals and impacted by the way in which humans perceive sound .Figure 13 shown the MFCC features of test audio signal. The first horizontal yellow lines below every segment are the fundamental frequency and at their strongest. Above the yellow line are the harmonics that share the same frequency distance between them. The Window Count = 67 and Individual Feature Length = 13 are the MFCC parameters.
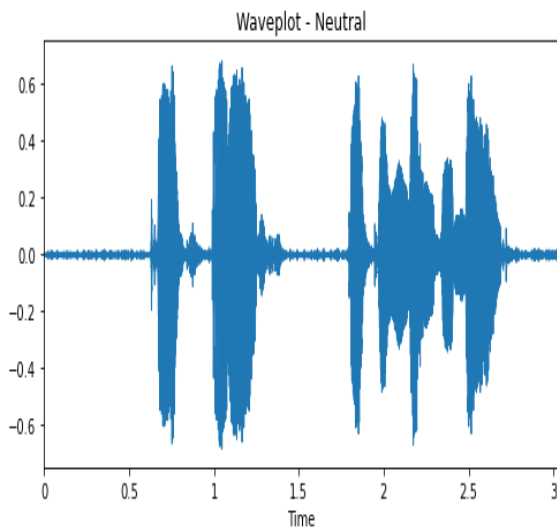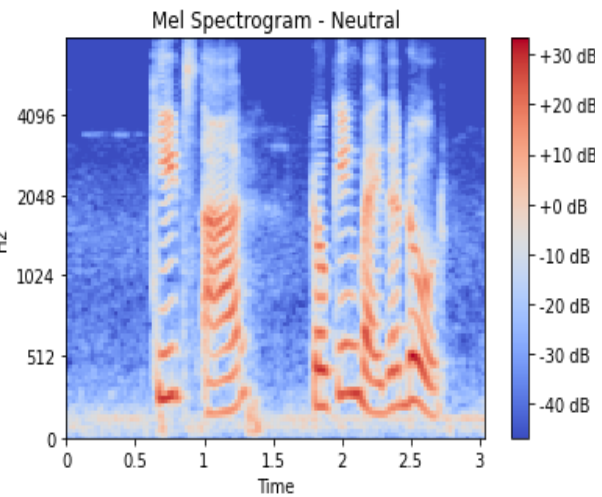

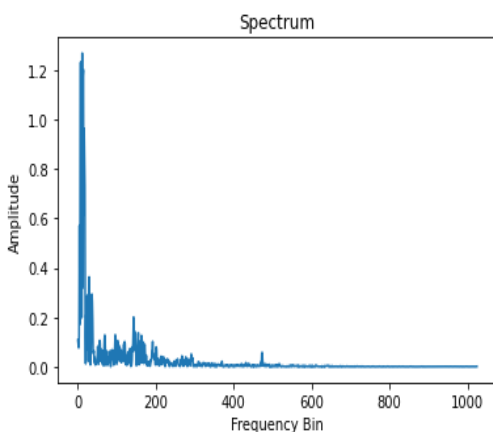*Figure 9*: *Plot of Signal Wave plot*

Figure 10: *Plot of Mel Spectrogram*

Figure 11: *Plot of Spectrogram*

Figure 12 : *Plot of FFT spectrum*

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
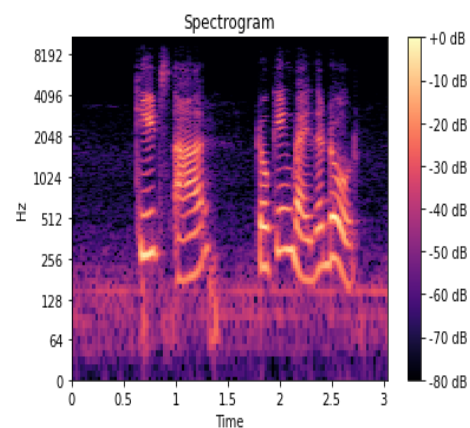DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
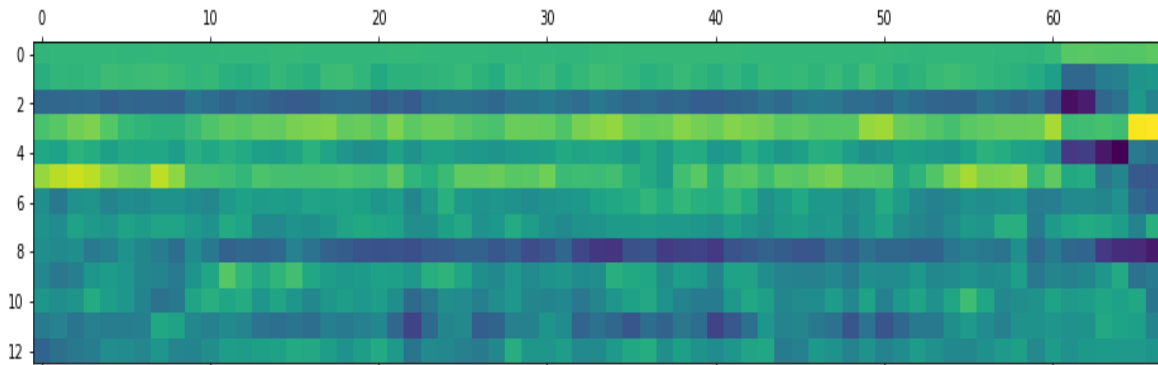Volume 18, September 2023

Figure 13: *Plot of MFCC Features*

III. Social Media Status Dataset Evaluation

In this paper, we explored various tools to evaluate social media status dataset to explore and visualize text data efficiently. The number of characters present in each sentence is shown in figure 14 and the histogram shows that status content range from 45 to 130 characters. The average word length of each status is ranges between 3.50 to 5.25 and that is shown in figure 15. The stop words are the words that are most commonly used in any language such as "the"," a"," an" etc. We can evidently see in Figure 16 that stop words such as "you"," the" and "to" dominate in Status contents. Wordcloud is a great way to represent text data. The size and color of each word that appears in the wordcloud indicate its frequency or importance. Figure 17 shows that the terms associated with the emotions are highlighted which indicates that these words occurred frequently in the social media status.
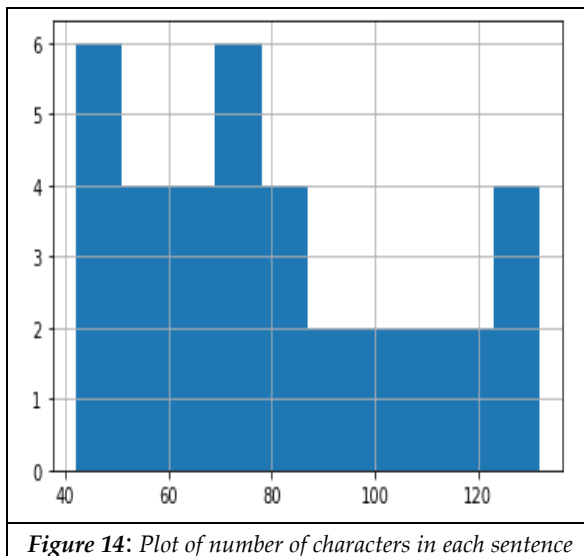

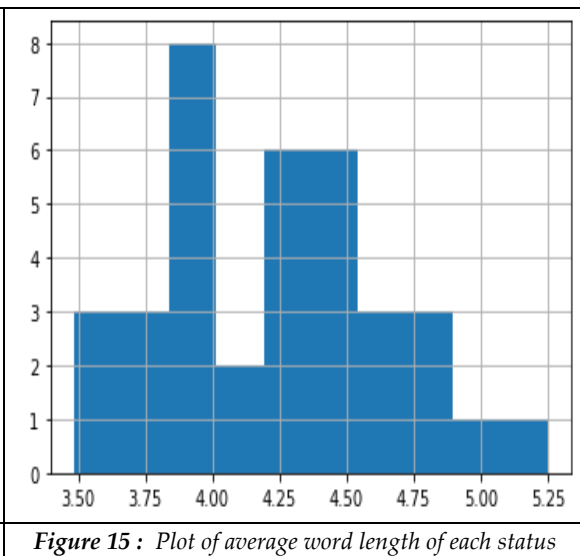| **Figure 14**: *Plot of number of characters in each sentence* | **Figure 15 :** *Plot of average word length of each status* |

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

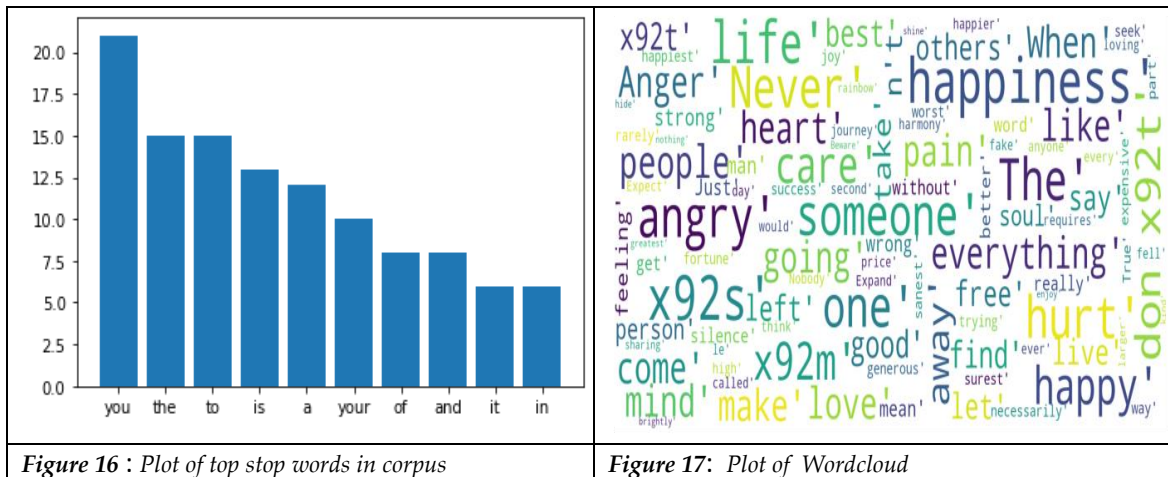| **Figure 16** : *Plot of top stop words in corpus* | **Figure 17**: *Plot of Wordcloud* |

V. Conclusion and Future Work

As Emotion Recognition is a growing research field in today's era, there are many existing unimodal and multimodal datasets available. The multimodal datasets normally consist of audio-visual, visual-audio-text, audio-visual-physiological data or EEG-EDA-ECG data. Still there is a pressing demand to develop a more extended and new multimodal dataset which consist important modalities like facial expression, audio, social media status and EEG signals. To fulfil this requirement, we have developed a new dataset for multimodal emotion recognition in which 7 different Professors of Government Engineering College, Gandhinagar have acted, for four basic modalities. This dataset has developed for three discrete emotion categories named happy, sad and angry. This multimodal dataset can be used to develop an emotion recognition model which is more robust and more accurate. In the future, the dataset can be tested on a pre-trained unimodal emotion recognition model to check whether the model works well on real-world dataset or not. We will try to include more modality like ECG, EDA ,RSA data to make our dataset more dimensional. The audio data is taken in English language only, so in future we will try to include audio data in other languages like Gujarati and Hindi to make our data set multilingual. In addition to the three basic emotion categories, the dataset can be upgraded for other categories like Surprise,Neutral,fear and disgust.

References

[1] K.S. Fleckenstein (1991). Defining Affect in Relation to Cognition: A Response to Susan McLeod. *J. Adv. Compos. 11 447–53.*

[2] R.W. Picard, E. Vyzas, J. Healey (2001). Toward machine emotional intelligence: analysis of affective physiological state, *IEEE Trans. Pattern Anal. Mach. Intell. 23 1175–1191. https://doi.org/10.1109/34.954607.*

[3] M. Scheutz (2012), The Affect Dilemma for Artificial Agents: Should We Develop Affective Artificial Agents?, *IEEE Trans. Affect. Comput. 3 424–433. https://doi.org/10.1109/T-AFFC.2012.29.*

[4] J.A. Healey, R.W. Picard(2005) , Detecting stress during real-world driving tasks using physiological sensors, *IEEE Trans. Intell. Transp. Syst. 6 156–166. https://doi.org/10.1109/TITS.2005.848368.*

[5] M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba (1998) , Coding facial expressions with Gabor wavelets, in: *Proc. Third IEEE Int. Conf. Autom. Face Gesture Recognit.,: pp. 200–205. https://doi.org/10.1109/AFGR.1998.670949.*

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

[6] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews(2010) , The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Workshop, IEEE, San Francisco, CA, USA, 2010: pp. 94–101. https://doi.org/10.1109/CVPRW.2010.5543262.*

[7] T. Kanade, J.F. Cohn, Yingli Tian(2000), Comprehensive database for facial expression analysis, in: Proc. Fourth IEEE Int. Conf. Autom. Face Gesture Recognit. *Cat No PR00580: pp. 46–53. https://doi.org/10.1109/AFGR.2000.840611.*

[8] G. Zhao, X. Huang, M. Taini, S.Z. Li, M. Pietikäinen(2011), Facial expression recognition from near-infrared videos, Image Vis. *Comput.* 29 607–619. *https://doi.org/10.1016/j.imavis.2011.07.002.*

[9] X. Zhang, L. Yin, J.F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, J.M. Girard(2014), BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database, Image Vis. *Comput. 32 (2014) 692–706. https://doi.org/10.1016/j.imavis.2014.06.002.*

[10] S. Cheng, I. Kotsia, M. Pantic, S. Zafeiriou(2018), 4DFAB: A Large Scale 4D Database for Facial Expression Analysis and Biometric Applications, *IEEECVF Conf. Comput. Vis. Pattern Recognit., IEEE, Salt Lake City, UT, USA, 2018: pp. 5117–5126. https://doi.org/10.1109/CVPR.2018.00537.*

[11] 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 (2013), Challenges in Representation Learning: A report on three machine learning contests, *ArXiv13070414 Cs Stat. http://arxiv.org/abs/1307.0414.*

[12] A. Dhall, R. Goecke, S. Lucey, T. Gedeon (2011), Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark, *IEEE Int. Conf. Comput. Vis. Workshop ICCV Workshop, 2011: pp. 2106–2112. https://doi.org/10.1109/ICCVW.2011.6130508.*

[13] C.F. Benitez-Quiroz, R. Srinivasan, A.M. Martinez(2016) , EmotioNet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild, *IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, IEEE, Las Vegas, NV, USA, pp. 5562–5570. https://doi.org/10.1109/CVPR.2016.600.*

[14] A. Mollahosseini, B. Hasani, M.H. Mahoor (2019), AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild, *IEEE Trans. Affect. Comput. 10 18–31. https://doi.org/10.1109/TAFFC.2017.2740923.*

[15] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss(2005), A Database of German Emotional Speech, *4.*

[16] I. Sneddon, M. McRorie, G. McKeown, J. Hanratty(2012), The Belfast Induced Natural Emotion Database, *IEEE Trans. Affect. Comput. 3 32–41. https://doi.org/10.1109/T-AFFC.2011.26.*

[17] J. Blitzer, M. Dredze, F. Pereira(2007), Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification, *ACL 2007. (n.d.) 8.*

[18] M. Dredze, K. Crammer, F. Pereira(2008), Confidence-weighted linear classification, *Proc. 25th Int. Conf. Mach. Learn. - ICML 08, ACM Press, Helsinki, Finland, pp. 264–271. https://doi.org/10.1145/1390156.1390190.*

[19] A.L. Maas, R.E. Daly, P.T. Pham, D. Huang, A.Y. Ng, C. Potts(2011), Learning Word Vectors for Sentiment Analysis, *Proc. 49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol., Association for Computational Linguistics, Portland, Oregon, USA,: pp. 142–150. https://www.aclweb.org/anthology/P11-1015.*

[20] R. Socher, A. Perelygin, J. Wu, J. Chuang, C.D. Manning, A. Ng, C. Potts, Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, *(n.d.) 12.*

[21] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras(2012), DEAP: A Database for Emotion Analysis ;Using Physiological Signals, *IEEE Trans. Affect. Comput. 3 18–31. https://doi.org/10.1109/T-AFFC.2011.15.*

Komal D. Anadkat, Hiteishi M. Diwanji , Shahid Modasiya, Mihir Mehta
FACIAL EXPRESSION, AUDIO, SOCIAL MEDIA STATUS, AND EEG
DATASET FOR MULTIMODAL EMOTION RECOGNITION

RT&A, No 3 (74)
Volume 18, September 2023

[22] R.-N. Duan, J.-Y. Zhu, B.-L. Lu(2013), Differential entropy feature for EEG-based emotion classification, *6th Int. IEEEEMBS Conf. Neural Eng. NER, 2013: pp. 81–84. https://doi.org/10.1109/NER.2013.6695876*.

[23] W.-L. Zheng, B.-L. Lu (2015) , Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks, *IEEE Trans. Auton. Ment. Dev. 7 162–175. https://doi.org/10.1109/TAMD.2015.2431497.*

[24] J.A. Miranda Correa, M.K. Abadi, N. Sebe, I. Patras (2018), AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups, *IEEE Trans. Affect. Comput. 1–1. https://doi.org/10.1109/TAFFC.2018.2884461.*

[25] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, K. Van Laerhoven (2018), Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection, *Proc. 20th ACM Int. Conf. Multimodal Interact., Association for Computing Machinery, Boulder, CO, USA,: pp. 400–408. https://doi.org/10.1145/3242969.3242985.*

[26] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J.N. Chang, S. Lee, S.S. Narayanan(2008), IEMOCAP: interactive emotional dyadic motion capture database, Lang. Resour. Eval. *42 335. https://doi.org/10.1007/s10579-008-9076-6*.

[27] L.-P. Morency, R. Mihalcea, P. Doshi (2011), Towards multimodal sentiment analysis: harvesting opinions from the web, *Proc. 13th Int. Conf. Multimodal Interfaces, Association for Computing Machinery, Alicante, Spain, 2011: pp. 169–176. https://doi.org/10.1145/2070481.2070509.*

[28] M. Wollmer, F. Weninger, T. Knaup, B. Schuller, C. Sun, K. Sagae, L.-P. Morency(2013), YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context*, IEEE Intell. Syst. 28 46– 53. https://doi.org/10.1109/MIS.2013.34*

[29] A. Bagher Zadeh, P.P. Liang, S. Poria, E. Cambria, L.-P. Morency(2018), Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph, *Proc. 56th Annu. Meet. Assoc. Comput. Linguist. Vol. 1 Long Pap., Association for Computational Linguistics, Melbourne, Australia, : pp. 2236–2246. https://doi.org/10.18653/v1/P18-1208*.

[30] F. Ringeval, A. Sonderegger, J. Sauer, D. Lalanne(2013), Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions, *10th IEEE Int. Conf. Workshop Autom. Face Gesture Recognit. FG,: pp. 1–8. https://doi.org/10.1109/FG.2013.6553805.*

[31] Komal Anadkat, Hiteishi Diwanji, Shahid Modasiya (2022). Effect of Preprocessing in Human Emotion Analysis Using Social Media Status Dataset. *RT & A ,No 1(67),volume-17*.