Fake Emotion Detection

1st Lalit Meena , 2nd Naman Solanki , 3rd Prajyot Kore
I'mBesideYou Inc.,IIT Bombay, IIT Delhi

***Abstract*—This document is a model and instructions for LaTeX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Emotions are fundamental to human communication, influencing decision-making, social interactions, and overall well-being. The ability to accurately analyze and interpret emotions has led to significant advancements in fields such as sentiment analysis, affective computing, and more recently, fake emotion detection.

Recent research has made substantial progress in the realm of emotion analysis. Notably, the work by [1] provides insights into text-based emotion detection, discussing state-of-the-art proposals that have advanced the field. Furthermore, facial emotion recognition has been explored extensively using deep learning techniques, as outlined by authors in it [2]. Their review emphasizes the importance of deep learning in capturing intricate facial expressions.

In the pursuit of multimodal emotion analysis, researchers have proposed innovative approaches. A notable example is the work by [3], who propose luna2021proposal using aural transformers and action units. This showcases the growing interest in combining modalities to achieve a comprehensive understanding of emotions.

Deception detection, closely related to our endeavor, has been a subject of significant research. Intelligent techniques for deception detection are discussed by authors in it [4], highlighting the need for robust methods in various applications. Additionally, multimodal fusion approaches for deception detection have been explored by researchers [5], showcasing the potential of combining multiple cues.

In our project, we focus on fake emotion detection, where individuals display facial expressions incongruent with their inner emotional state. This challenging task finds applications in psychology, human-computer interaction, and security. Our approach capitalizes on the disparities between facial expressions and inner emotions, aiming to enhance the accuracy of fake emotion detection.

Our research addresses the limitations of existing methods by integrating information from multiple modalities. We develop tailored models for capturing inner emotions and facial expressions, incorporating cutting-edge AI techniques like deep neural networks and natural language processing techniques. These individual models are then integrated into a multimodal AI framework, enabling nuanced and robust emotion predictions.

Our contributions extend to practical applications. Improved diagnosis benefits from our approach, enabling more accurate assessments of patients' emotional states, leading to better treatment plans and outcomes. Furthermore, our approach enhances user experiences in human-computer interaction scenarios, enabling more empathetic and intuitive interactions.

In conclusion, this research advances the field of emotion analysis by proposing a multimodal approach to fake emotion detection. By leveraging diverse data sources and advanced AI techniques, we aim to provide a comprehensive and accurate understanding of human emotions. The subsequent sections of this paper will delve into the methodology, experiments, and results that support the effectiveness of our proposed approach.

II. Background

A. *Understanding Emotions: A Multimodal Exploration*

Evolution from Ekman to Plutchik: Human emotions, both overt facial expressions and internal states, intricately shape social interactions. Paul Ekman's seminal work, "An Argument for Basic Emotions" (1992) [6], is a cornerstone in emotion studies, identifying six universally recognized emotions: happiness, sadness, anger, fear, surprise, and disgust. This foundational framework is further enriched by Robert Plutchik's 1982 paper, "A Psychoevolutionary Theory of Emotions" [7], proposing eight primary emotions arranged in opposing pairs.

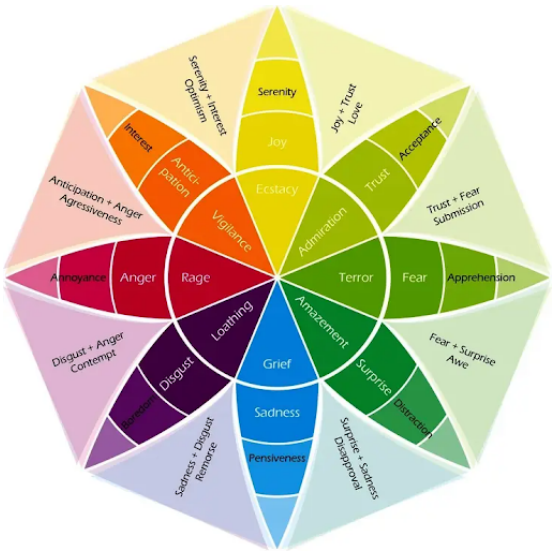

Fig. 1: Dr. Robert Plutchik's Wheel of Emotion

***Emotion Prediction: Affective Computing Milestones**:* In the dynamic field of affective computing, emotion prediction emerges as a pivotal area, drawing insights from notable

datasets such as RAVDESS, IEMOCAP, and ISEAR. Research in text modality-based emotion detection explores embedding-based features, textual feature vectors using pre-trained models like BERT and RoBERTa, and customized rule-based approaches **Exploring Transformers**, [8], [9]. In the speech modality, models like XLSR-Wav2Vec 2.0 and multimodal systems [10] showcase remarkable accuracies.

Cutting-edge research, exemplified by [10], demonstrates that multimodal models, incorporating both aural and visual cues, outperform unimodal counterparts, achieving an impressive 86.70% accuracy. Transformer-based techniques using BERT or RoBERTa excel in the text modality **Exploring Transformers**, [8], [9]. These insights underscore the potency of combining modalities for comprehensive emotion detection.

B. Hidden Emotions: A Less-Explored Landscape

Why Hide Emotions?: The act of concealing emotions, influenced by social norms, privacy concerns, personal protection, and mental health factors like depression [11], has profound consequences, impacting psychological well-being, particularly in service-oriented professions.

Automatic Hidden Emotion Detection: A New Frontier: Despite substantial progress in emotion detection, the automatic identification of hidden emotions remains underexplored. Early works [11], [12] have explored multi-modal signals and video analysis to detect concealed emotions. Our project aims to contribute to this emerging field by utilizing video, audio, and other modalities for accurate identification.

Applications and Value: The applications of hidden emotion detection are diverse, ranging from interviews and client calls to court hearings and therapy sessions. Additionally, it plays a crucial role in diagnosing mental health issues like depressive disorders, offering potential benefits in HR monitoring, personalized recommendations, and enhancing user experiences.

Challenges and Datasets in Hidden Emotion Detection: The exploration of hidden emotion detection encounters challenges in dataset diversity and annotation complexities. Existing datasets, such as the "Hidden Emotion Dataset" [11], provide valuable resources for model training. However, challenges include the subtle nature of hidden emotions, subjectivity in labeling, and the need for context-aware annotation.

C. Deception Detection: Bridging Modalities

Datasets for Deception Detection: Deception detection, akin to hidden emotion detection in some aspects, relies on datasets like the "gupta2019bag" dataset [13] and the DDPM dataset [14]. These datasets, encompassing audio, video, EEG, gaze data, and more, provide a nuanced understanding of deceptive behavior.

Approaches in Deception Detection: Methodologies for deception detection often involve multimodal fusion [14], achieving a remarkable 100 percentage True Positive Rate. The integration of physiological signals, such as saccadic eye movements and heart rate estimation, showcases the diversity of approaches in this field.

Challenges and Datasets in Deception Detection: Deception detection confronts challenges in the development of diverse and representative datasets. Existing datasets like "Deceptive Speech Dataset" [14] and "Visual Deception Dataset" [13] contribute valuable instances. Challenges encompass the dynamic nature of deceptive behavior, ethical considerations in data collection, and the need for real-world applicability.

Insights and Key Gaps: A clear demarcation exists between the tasks of deception detection and hidden emotion detection. In deception detection, the primary objective is to unveil intentional deceit, often associated with criminal activities. This task emphasizes gathering physiological data, a requirement often impractical in routine applications, serving investigative and security purposes.

Conversely, hidden emotion detection aims to discern whether an individual is concealing their emotions, applicable in diverse contexts driven by societal norms and mental health considerations. Unlike the intensive need for physiological data in deception detection, hidden emotion detection can leverage readily available audio, video, and text data, making it applicable to scenarios like HR monitoring and enhancing user experiences.

Despite distinctions between the two tasks, shared aspects allow for the application of similar approaches. The intricacies of human emotions, whether concealed or deceitful, can be effectively addressed using multimodal models. Thus, despite their divergent focal points, hidden emotion detection can draw insights from the methodologies employed in the domain of deception detection. This nuanced understanding enables the development of robust models capable of discerning concealed emotions across various practical applications.

D. Conclusion and Future Directions

In conclusion, the synthesis of emotion prediction, hidden emotion detection, and deception detection reveals a nuanced landscape. The journey from Ekman's fundamental emotions to Plutchik's paired complexities lays the groundwork. Multimodal approaches, especially those leveraging transformer-based techniques, demonstrate superiority in emotion detection. The less-explored territory of hidden emotion detection beckons, promising applications in diverse fields. Deception detection, with its focus on intentional deceit, provides additional insights. As we navigate this intricate landscape, future research should focus on refining multimodal models, understanding the nuances of hidden emotions, and bridging gaps in the exploration of human affect.

III. METHODOLOGY: UNRAVELING EMOTIONS THROUGH MULTIMODAL TECHNIQUES

Our research methodology is built upon the recognition of the formidable potential inherent in multimodal techniques, specifically designed for unraveling the intricate nuances of emotions, particularly in the domain of deception detection. Extensive background research conclusively underscores the efficacy of multimodal approaches, laying a solid foundation for our exploration.

Year	Classifiers	Pre-Processing	Features	Testing Procedure	Databases	Accuracy
2015 [15]	Bayes	Geometric	MK	CV	ENTERFACE'05	98.00%
2013 [16]	SVM	Smooth	Gabor/PCA	CV	ENTERFACE'05	80.27%
2014 [17]	MLP/RBF	NN	ITMI/QIM	HO (PI)	CK/ENTERFACE'05	75.00%

TABLE I: Results of reviewed works for audiovisual approaches (Table 2).

Year	Classifier	Pre-processing	Features	Testing Procedure	Databases	Accuracy
2016 [18]	ED	Geometric	Landmarks	HO	BU-4DFE/BP4D-S	100.00%
2017 [19]	GMM	Bandlet	LBP/KW	HO	CK/JAFFE	99.80%
2018 [20]	CNN	HE/Geometric	OF	HO	CK+/SAVEE/AFEW	98.77%

TABLE II: Results of reviewed works for video approaches.

A. *Dataset Selection*

To capture real-world complexity and diversity, we selected 100 emotionally diverse videos (Asian) from the "1,003 People - Emotional Video Data" dataset **datatang** as sample shown in fig 2. This handpicked subset goes beyond controlled settings, featuring a multi-lingual tapestry of English and Chinese languages. Meticulously annotated with 11 facial and 15 inner emotions, this dataset reflects the natural ebb and flow of emotions in everyday scenarios, paving the way for sophisticated emotion detection models.

B. *Data Preparation*

*1) **Handling Multi-lingual Texts**:* The complexity of our dataset arises from the inclusion of text transcriptions from different languages, with a significant portion 90% in Chinese and the remainder in English. To facilitate analysis, we used Google Translate to convert the Chinese transcriptions into English, though it exhibited only a 72% accuracy rate for colloquial phrases. **add resource**

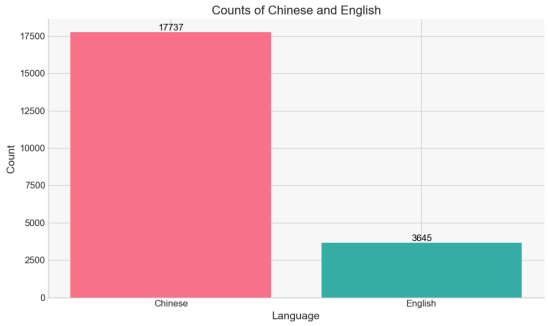

Fig. 2: Language distribution

*2) **Handling Imbalance**:* Another challenge encountered in the dataset was imbalanced data, where certain emotion categories had significantly fewer samples than others. To mitigate this issue and prevent bias in emotion prediction, various techniques were employed. These techniques included undersampling, oversampling, and combined methods like SMOTE (Synthetic Minority Over-sampling Technique), SMOTETomek, Random undersampling, oversampling, and balanced batch generator.

*3) **Handling Multi-labels:Emotion Categorization**:* **Method-1**: The paper [21] introduced a comprehensive emotion categorization scheme that encompassed both facial and inner emotions. Due to the complexity of the emotional labels being multi-labeled, a simplified approach was adopted for categorization. For each instance, the emotion category with the maximum occurrences was selected as the primary label. Facial emotions extracted from video data were meticulously categorized into 11 distinct classes, representing a wide range of emotions such as happiness, sadness, anger, disgust, and surprise, among others. For inner emotions, text transcripts were manually labeled and classified into three fundamental categories: positive, negative, and neutral. Categorization is as shown in figure:

Positive Emotions	Negative Emotions	Neutral Emotions
Positive	Negative	Others
Confident	Embarrassed	Indicative
Approved	Coward	
Surprised	Hesitant	
Imperative	Blamed	
	Depressed	

TABLE III: Inner Emotions Categorization

Positive Emotions	Negative Emotions	Neutral Emotions
Happy	Disappointed	Neutral
Excited	Sad	Other
Surprised	Angry	
	Disgusted	
	Scared	
	Bored	

TABLE IV: Facial Emotions Categorization

Method-2 In second approach, we manually mapped possible 11*15 combinations as faking emotion or not, example as fake or not if given facial and inner emotions are in opposite nature. To identify opposite emotions, we have manually constructed mapping dictionary based on my understanding, related literature, Plutchik's Wheel of Emotions (shown in Fig.12),etc. Fig. 1. Dr. Robert Plutchik's Wheel of Emotions To handle complexity, we avoid solely relying on mapping. Instead, we assess the percentage distribution of positive and negative emotions within inner and facial emotion lists. This approach allows us to identify content as Positive-Negative

based on a set threshold, addressing potential conflicts in labeling.

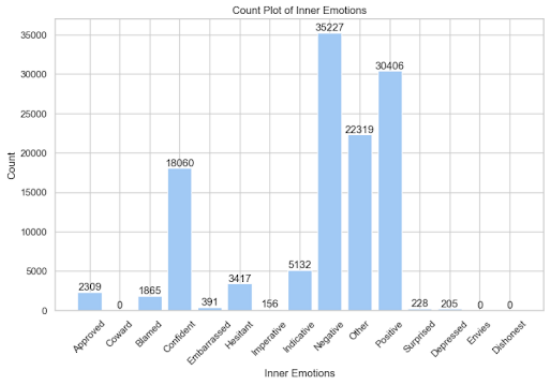

Fig. 3: Inner Emotion Countplot

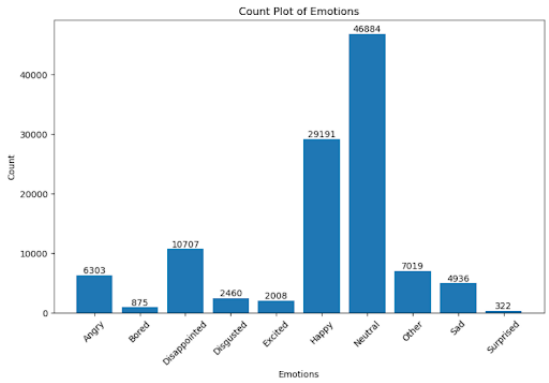

Fig. 4: Facial Emotion Countplot

C. *Evaluation Metrics*

Micro F1 Score

Micro F1 score is the harmonic mean of precision and recall for all classes in the dataset.

$$\text{F1} = \frac{2 \cdot \text{Micro Precision} \cdot \text{Micro Recall}}{\text{Micro Precision} + \text{Micro Recall}}$$

MICRO-AVERAGE F1 SCORE

The micro-average F1 score is calculated as follows:

$$\text{Precision}_{\text{micro}} = \frac{\sum_i \text{True Positives}_i}{\sum_i \text{True Positives}_i + \sum_i \text{False Positives}_i}$$

$$\text{Recall}_{\text{micro}} = \frac{\sum_i \text{True Positives}_i}{\sum_i \text{True Positives}_i + \sum_i \text{False Negatives}_i}$$

$$\text{F1}_{\text{micro}} = 2 \times \frac{\text{Precision}_{\text{micro}} \times \text{Recall}_{\text{micro}}}{\text{Precision}_{\text{micro}} + \text{Recall}_{\text{micro}}}$$

where i represents each class.

Accuracy of Each Class

The accuracy of each class is the ratio of correctly predicted instances of that class to the total instances of that class.

$$\text{Accuracy}_{\text{class}} = \frac{\text{True Positives}_{\text{class}}}{\text{True Positives}_{\text{class}} + \text{False Positives}_{\text{class}}}$$

Receiver Operating Characteristic - Area Under the Curve (ROC-AUC)

ROC-AUC is a measure of the area under the receiver operating characteristic curve, which plots the true positive rate against the false positive rate.

$$\text{ROC-AUC} = \int_0^1 \text{True Positive Rate}\, d(\text{False Positive Rate})$$

D. *Modality-based Methodology: Implicit Overview of Multimodal AI*

Our study revolves around the seamless integration of multimodal AI, strategically harnessing the distinctive strengths of different modalities for a nuanced comprehension of emotions. The synergistic fusion of audio, video, and text modalities forms the foundation of our project, contributing to its overall efficacy.

*1) **Speech Modality**:* In the speech modality-based segment of our methodology, we leveraged the MoviePy library [20] to split videos into subclips based on predefined timestamps. Our experiments utilized the LARGE version of xlsr-Wav2Vec2.0, fine-tuned on the English dataset from Common Voice. Two models were employed for English and Chinese languages, accessible in the Hugging Face repository [21] and [22].

For feature extraction, we employed the pre-trained network to obtain latent speech representations. Initially, we subsampled the audios to 16 kHz and converted them into mono-channels using FFmpeg [23]. The transformer generated sequences of 512-dimensional embeddings from the convolutional feature encoder, and the average of these embeddings along the temporal dimension was calculated. This averaged 512-dimensional representation was then utilized to train static speech emotion recognizers through the sklearn library [24]. Due to the dataset's imbalance and complexities like multi-label and multi-class, we addressed them using sampling techniques such as SMOTE and smote-tomek to balance the dataset before further training.

The extracted features were employed in two ways: one for training a Speech Emotion Recognizer using MLP layers, and another for direct hidden emotion detection. Among the compared models, we employed XGBoost and a multilayer perceptron (MLP) with one or two layers of 80 neurons each for hidden emotion detection. For Speech Emotion Recognition, we used a multilayer perceptron (MLP) with one or two layers of 80 neurons each and an output layer with 2-3 neurons. This approach allowed us to reuse the original features of the speech-to-text task, transferring learned knowledge from the embeddings to the new models.

*2) **Video Modality**:* The video modality is meticulously processed to preserve temporal integrity, leveraging video frames. Cleaning involves handling missing frames, standardizing video lengths, and scaling to normalize pixel values. Feature engineering and selection employ pre-trained Convolutional Neural Networks (CNNs) for extraction, focusing on relevant facial landmarks and expressions. Modeling encompasses 3D CNNs and LSTM-based models.

*a) **Method 1: Emotion Vectors and Blinks**:* DeepFace pipelines were used to extract 7 emotion vectors, blink, gaze, and eye-offset. These features were utilized with Xgboost (cv=10) after oversampling.

*b) **Method 2: Facial Landmarks as Features**:* Using the Dlib library, 68 facial landmarks were extracted as features.

*c) **Method 3: Region of Interest (ROI) Analysis**:* Considered direct frames from videos, this method achieved notable results on a subset of 5 videos with 3 individuals.

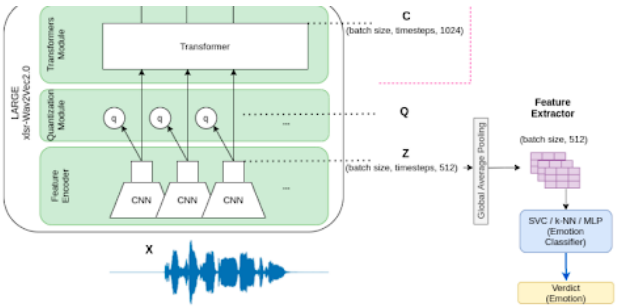

Figure 2. Proposed pipelines for speech emotion recognition.

Fig. 5

*3) **Text modality**:* In the text modality-based section of our methodology, our initial step involved the translation and preprocessing of textual data, which was provided in the form of a JSON file then imported as a Pandas Dataframe. To address multilingual challenges, Google Translate played a crucial role, facilitating the translation of significant Chinese language text into English. Emphasizing language-specific handling, we applied tokenization and embedding based methods to the processed data.

We utilized pre-trained transformer-based BERT models for extracting context-aware embeddings with features represented in 768 dimensions, corresponding to the CLS token, a valuable component in classification tasks. In our architectural overview, we considered two distinct settings. Firstly, focusing on multimodal inner emotion recognition, we predicted the posteriors corresponding to the Text-based Emotion Detector. For this, we fine-tuned the BERT model for 10 epochs specifically for inner emotion detection in text classification.

In the second architecture, we directly classified for the main task using a single text modality, leveraging transformer-based methods. Various settings were applied to the BERT model for text classification in the context of fake emotion detection.

Throughout our modeling phase, we employed different variants of transformer based BERT models for emotion detection. To assess performance rigorously, metrics such as F1-score, precision, and recall were meticulously tracked. This comprehensive approach allowed us to harness the capabilities of advanced language models in the pursuit of hidden emotion detection within textual data.

*4) **Multimodality**:* For hidden emotion detection, our main goal was to identify concealed emotions by examining the differences between emotions detected by the Facial model and the Inner emotion model. Following sentiment analysis on Inner Emotion from text and speech modalities, we sought to refine our outcomes through a decision-level fusion approach.To achieve this, we built a multimodal Inner emotion recognizer by combining the outputs of the textual model and the Speech-based Emotion Detection (ED) Model. This involved using the MLPClassifier on the extracted posteriors to create a robust multimodal model for Inner Emotion Detection.

To compare the knowledge learned from the Facial and Inner emotion models, we implemented a late fusion strategy. Starting by extracting posteriors from the last fully-connected layer of the models, we obtained a 15-dimensional vector for inner emotions and a 7-dimensional vector for facial emotions. These embeddings from each modality were then concatenated. We further explored the hidden emotion detection task by applying additional computation models on top of the Facial Emotions model and Inner emotion Model, utilizing decision-level fusion approaches for the multimodal model.In the end, we trained a multilayer perceptron (MLP) from the sklearn library, incorporating two layers of 80 neurons.

This methodology serves as a robust framework, strategically leveraging the unique strengths of each modality to facilitate effective multimodal emotion detection, particularly within the realm of deception identification.

IV. Experiments

Our research journey has unearthed key insights into the intricate landscape of emotion detection. The multimodal approach emerges as pivotal, surpassing the limitations of single modality techniques. The fusion of audio, video, and text modalities not only improves accuracy but also provides a more comprehensive understanding of hidden emotions, unraveling the intricacies of human emotional expression.

Video Modality:

In this research, we created a video dataset by incorporating start and end timings for transcriptions, along with inner and facial emotions. Using OpenCV (https://opencv.org/), we calculated Frames Per Second (FPS) and extracted start and end frames for each transcription entry. This comprehensive dataset enables precise analysis of emotional expressions in various video segments, enhancing emotion recognition research.

***Method 1: Emotion Vectors and Blinks**:*

- DeepFace (https://github.com/serengil/deepface) pipelines were employed to extract 7 emotion vectors namely 'angry', 'disgust', 'fear', 'happy', 'sad', 'surprise', 'neutral', as well as blink, gaze, and eye-offset features.
- These features were utilized in conjunction with XGBoost (cv=10) after oversampling using SMOTE. The method demonstrated results in terms of performance, achieving

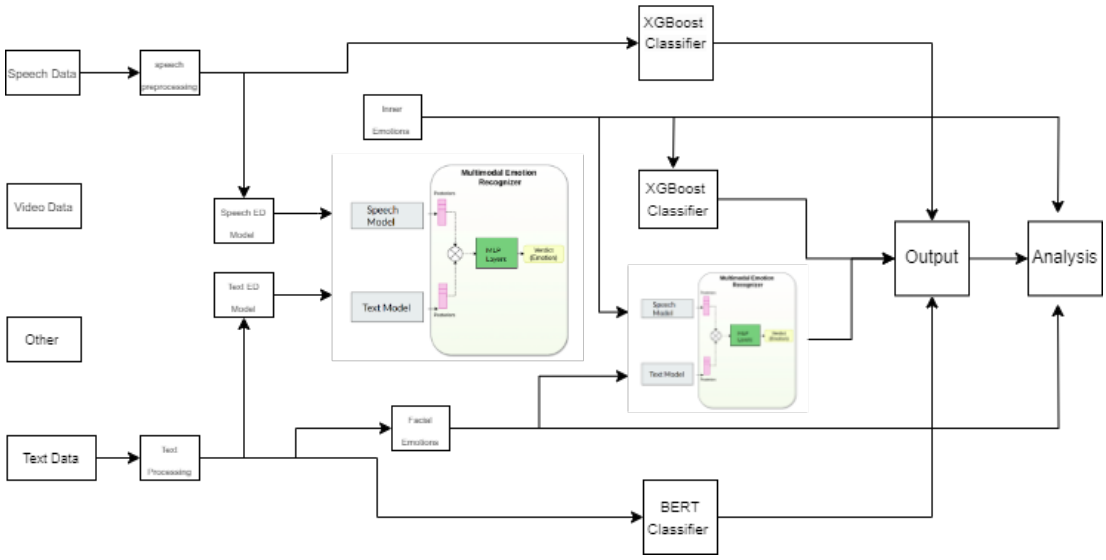

Fig. 6: Whole architecture of the Project

an accuracy of 49% and a micro-average F1 score of 0.15 for facial emotion prediction, and AUC scores for each emotions are given below:

TABLE V: ROC Scores using Emotion Vectors and Blinks

Class	ROC AUC
Neutral	0.70
Sad	0.65
Disappointed	0.587
Happy	0.739
Disgusted	0.58
Excited	0.59
Other	0.69
Surprised	0.68
Angry	0.66
Bored	0.63
Scared	0.543

Method 2: Facial Landmarks as Features: We considered the MediaPipe pose estimation model to extract the initial 22 landmarks, encompassing both facial and hand landmarks. Subsequently, we performed feature engineering inspired by the approach detailed in [22]. For each instance, we computed statistical features including mean, standard deviation, minimum, and maximum across timestamps. Emotions were categorized into new labels for modeling, with the class distribution depicted in Fig. 7.

Dimensionality reduction was conducted using Principal Component Analysis (PCA). Various classifiers were applied for classification, employing weighted class entropy and hyperparameter tuning, the results of which are summarized in Table VI.

TABLE VI: Mean ROC Scores for MediaPipe Pose Estimation Landmarks using different models

Models	Mean ROC AUC
Random Forest	0.512
XGBoost	0.508
Logistic Regression	0.480

From Table VI, it is evident that the landmark-based features yielded suboptimal performance.

Fig. 7

Textual Modality:

*1) **Method 1:*** In the text modality-based section of our methodology, our initial step involved the translation and preprocessing of textual data, which was provided in the form of a json file then imported as a pandas dataframe. To address multilingual challenges, Google Translate played a crucial role, facilitating the translation of significant Chinese language text into English. Emphasizing language-specific handling, we applied tokenization and embedding based methods to the processed data.

Cleaning procedures were then implemented to eliminate textual noise and enhance data quality. For feature engineering and selection, We utilized pre-trained transformer-based BERT models for extracting context-aware embeddings with features represented in 768 dimensions, corresponding to the CLS token, a valuable component in classification tasks.

In our architectural overview, we considered two distinct settings.Firstly, focusing on multimodal inner emotion recognition, we predicted the posteriors corresponding to the Text-based Emotion Detector. For this, we fine-tuned the BERT

Sr. No.	Settings	Loss Function & valid_eval	Acc	F1-Score
1	Sigmoid units, threshold (0.5) for labels	Custom weighted Loss & no	0.52	0.53
2	Softmax, argmax for labels	Loss custom weighted & Val_loss	0.53	0.54
3	Softmax, argmax for labels	Custom weighted Loss & no	0.51	0.52
4	Softmax, argmax for labels	BCE & val_f1_score	0.56	0.56
5	Softmax, argmax; L2reg; Non-trainable BERT layers	Weighted BCE & val_acc	0.47	0.49
6	Same settings as 5th + 1 dense, 1 dropout layer	Weighted BCE & val_acc	0.48	0.49

TABLE VII: Comparison of Results of Fake ED task, directly from BERT based Text-classification, for different settings

Performance Metric	Score
Average Accuracy	93%
Weighted Avg Accuracy	78.43%
Weighted Avg F1	0.42
Weighted Avg Precision	0.47
Weighted Avg Recall	0.41
Random Model Accuracy	around 7%
Random Individual Emotion Accuracy	around 50%

TABLE VIII: Final results of Inner Emotions prediction

model for 10 epochs specifically for inner emotion detection in text classification.

Table 5 presents a comparison of BERT model settings for Text Classification in Fake Emotion Detection. Metrics such as f1-score, precision, and recall are weighted averages across three classes. Notably, rows 5 and 6 using BCE with class weights experience underfitting during training, possibly due to imbalanced class weights and layer modifications. Conversely, rows 2 and 4 demonstrate superior results compared to others.

	precision	recall	f1-score	support
not_fake	0.97	0.92	0.94	2052
fake	0.11	0.24	0.15	87
accuracy			0.89	2139
macro avg	0.54	0.58	0.55	2139
weighted avg	0.93	0.89	0.91	2139

Fig. 8: Classification report of Results of Fake ED directly from Speech features (Labeling approach-1)

In the second architecture, we directly classified for the main task using a single text modality, leveraging transformer-based methods. Various settings were applied to the BERT model for text classification in the context of fake emotion detection.

Throughout our modeling phase, we employed different variants of transformer based BERT models for emotion detection. To assess performance rigorously, metrics such as F1-score, precision, and recall were meticulously tracked. This comprehensive approach allowed us to harness the capabilities

	precision	recall	f1-score	support
not_fake	0.70	0.67	0.68	1233
fake	0.09	0.20	0.13	89
neutral	0.63	0.59	0.61	817
accuracy			0.62	2139
macro avg	0.48	0.49	0.47	2139
weighted avg	0.65	0.62	0.63	2139

Fig. 9: Classification report of Results of Fake ED directly from Speech features (Labeling approach-2)

of advanced language models in the pursuit of hidden emotion detection within textual data.

*2) **Method 2**:* In our study, we utilized various techniques for textual feature extraction, including GloVe embeddings (200d) , count vectorization, NLP features, and TF-IDF vectors. These methods enriched our data representation, enhancing emotion recognition and providing deeper insights into emotional experiences and personality traits. See results below:

TABLE IX: Model Performance Metrics

Model	Micro-Average F1	Accuracy
Logistic Regression	0.58	0.65
SVM	0.58	0.73
Random Forests	0.39	0.70
Bi-LSTM	0.53	0.75

A. *Fake Emotion Prediction*

using the categorization, described above, we considered two approaches:

*1) **Binary labeling**:* Here we have only considered binary labels:0-Not fake and 1-Fake. After labeling, training data contains only 853 fake samples out of total 17105 samples (4.98 %). And testing data contains only 87 fake samples out of total 2139 samples (4.07 %). There is problem of scarcity of fake samples and resulted imbalance dataset.

*2) **Multi-class labeling**:* Here we tried to improve labeled data distribution by including neutral label:0-Not fake ,1-Fake and 2-Neutral. After labeling, training data contains: 9851 not fake samples, 6395 neutral samples and 859 fake samples. And testing data contains: 1233 not fake samples, 817 neutral samples and 89 fake samples. Now, there is some fair splitted distribution but still fake samples around 5%. There is still problem of scarcity of fake samples and resulted imbalance dataset

REFERENCES

[1] E. Acheampong *et al.*, "Text-based emotion detection: Advances, challenges, and opportunities," *Engineering Reports*, vol. 2, e12189, 2020. DOI: 10.1002/eng2.12189.

[2] M. Mellouka *et al.*, "Facial emotion recognition using deep learning: Review and insights," *Procedia computer science*, vol. 176, pp. 554–563, 2020. DOI: 10.1016/j.procs.2020.07.101.

[3] Y. Luna-Jiménez *et al.*, "A proposal for multimodal emotion recognition using aural transformers and action units on ravdess dataset," *Applied Sciences*, vol. 12, no. 1, p. 327, 2022. DOI: 10.3390/app12010327.

[4] A. Alaskar *et al.*, "Intelligent techniques for deception detection: A survey and critical study," *Soft Computing*, pp. 1–19, 2022. DOI: 10.1007/s00500-022-07603-w.

[5] I. Chebbi *et al.*, "Deception detection using multimodal fusion approaches," *Multimedia Tools and Applications*, pp. 1–27, 2021. DOI: 10.1007/s11042-021-11148-9.

[6] P. Ekman, "An argument for basic emotions," 1992.

[7] R. Plutchik, "A psychoevolutionary theory of emotions," 1982.

[8] M. Polignano, M. de Gemmis, and G. Semeraro, "SWAP at SemEval-2019 task 3: Emotion detection in conversations through tweets, CNN and LSTM deep neural networks," in *Proceedings of the 13th International Workshop on Semantic Evaluation*, J. May, E. Shutova, A. Herbelot, X. Zhu, M. Apidianaki, and S. M. Mohammad, Eds., Jun. 2019.

[9] Z. Li, F. Tang, M. Zhao, and Y. Zhu, "Emocaps: Emotion capsule based model for conversational emotion recognition," *arXiv preprint arXiv:2203.13504*, 2022.

[10] C. Luna-Jiménez, R. Kleinlein, D. Griol, Z. Callejas, J. M. Montero, and F. Fernández-Martınez, "A proposal for multimodal emotion recognition using aural transformers and action units on ravdess dataset," *Applied Sciences*, vol. 12, no. 1, p. 327, 2021.

[11] B. C. Song and D. H. Kim, "Hidden emotion detection using multi-modal signals," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–7.

[12] P. Agrawal and A. Suri, "Nelec at semeval-2019 task 3: Think twice before going deep," *arXiv preprint arXiv:1904.03223*, 2019.

[13] V. Gupta, M. Agarwal, M. Arora, T. Chakraborty, R. Singh, and M. Vatsa, "Bag-of-lies: A multimodal dataset for deception detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[14] J. Speth, N. Vance, A. Czajka, K. W. Bowyer, D. Wright, and P. Flynn, "Deception detection and remote physiological monitoring: A dataset and baseline experimental results," in *2021 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, 2021, pp. 1–8.

[15] J. G. Rázuri, "Decision-making content of an agent affected by emotional feedback provided by capture of human's emotions through a bimodal system," *International Journal of Computer Science Issues*, vol. 12, no. 6, 2015.

[16] M. Rashid, S. Abu-Bakar, and M. Mokji, "Human emotion recognition from videos using spatio-temporal and audio features," *The Visual Computer*, vol. 29, pp. 1269–1275, 2013.

[17] M. Bejani, D. Gharavian, and N. M. Charkari, "Audiovisual emotion recognition using anova feature selection method and multi-classifier neural networks," *Neural Computing and Applications*, vol. 24, pp. 399–412, 2014.

[18] A. Danelakis, T. Theoharis, and I. Pratikakis, "A spatio-temporal wavelet-based descriptor for dynamic 3d facial expression retrieval and recognition," *The visual computer*, vol. 32, pp. 1001–1011, 2016.

[19] M. S. Hossain and G. Muhammad, "An emotion recognition system for mobile applications," *IEEE Access*, vol. 5, pp. 2281–2287, 2017.

[20] J. Zhao, X. Mao, and J. Zhang, "Learning deep facial expression features from image and optical flow sequences using 3d cnn," *The Visual Computer*, vol. 34, pp. 1461–1475, 2018.

[21] D. L. Robinson, "Brain function, emotional experience and personality," *Netherlands Journal of Psychology*, vol. 64, pp. 152–168, 2008.

[22] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for emg signal classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, 2012, ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2012.01.102. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417412001200.