

## Score Level Fusion Of Bimodal Emotion Recognition System Using Text And Speech

A.Shunmuga Sundari<sup>1</sup>, Dr.R.Shenbagavalli<sup>2</sup>

<sup>1</sup> Part Time Ph.D Research Scholar (internal) (17221172162018) - PG and Research Department of Computer Science, Rani Anna Government College for Women.

<sup>2</sup> Assistant Professor, PG and Research Department of Computer Science, Rani Anna Government College for Women, Affiliated to Manonmaniam Sundaranar University, Abhishekapatti, Tirunelveli-627412, Tamilnadu, India.

**Article History:** Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

**ABSTRACT:** The Meteoric blooming of social media services has created freakish opportunities for people to publicly express their views, opinions, and attitudes through various mediums such as text, voice and video. These opinions, views, and attitudes contain user emotion. Emotions sculpt a very indispensable and elementary aspect of people's existence. With the emerging social media analyzing the emotion depicted by the user through single emotion recognition system is much hard to satisfy the demands of emotional recognition system. Aiming to optimize the performance of the emotional recognition system, a score level fusion of bi-modal emotional recognition system SFBM\_TS from text and speech was proposed in this paper. It focuses to predict the emotion into four class joy, sad, anger and fear. Initially text is analyzed through selective lexicon based BI-LSTM method and in parallel the speech is analyzed through deep learning network. Finally it employs the score level fusion of both text and speech by combining each class resulted scores of both method and takes the weighted average score for each emotion class. The emotion class which has the highest average score is finalized as the result. To precise the emotion classification process text data is assigned with 0.6 weights and speech is assigned with 0.4 weights. The definitive emotional state was determined by the output of both audio and text emotion analysis. As a result the accuracy of proposed model SFBM\_TS is higher than that of the single emotion recognition modals by 5%.

**Keywords:** Emotion Recognition, Speech Analysis, BI-LSTM, Deep Neural Network

### 1. INTRODUCTION

The World Wide Web 2.0 has created an efficient platform and also opened a lot of new ways for interactions among the people. Online communities, where the users can express their feelings, opinions, and views about various topics, events, movie and people etc. These opinions, views, and attitudes contain user emotion. Analyzing the emotion depicted by these social media texts has given rich valuable information for both government and companies.

Whereas there are several modalities available for people to express the emotion such as text, speech etc. so it's really hard to satisfy the demands of emotion recognition system to analyze the emotion by a single emotion recognition model. Therefore to optimize the accuracy and performance of emotion recognition system this paper propose a novel score level fusion of bi- model emotional recognition system SFBM\_TS, which comprises of both text and speech. It predicts the emotion into four class joy, anger, sad and surprise. It contains two branches for text and speech.

Selective Lexicon based BI-LSTM (SL+BILSTM) neural network method is used to predict the emotion depicted by the text. BI- LSTM network is a very powerful classifier for sentiment analysis because it makes use of a memory in the network. Having a memory in the network is useful because when dealing with sequenced data such as a text, the meaning of a word depends on the context of the previous text.

Whereas the speech is analyzed by the deep neural network (DNN).Predicting the emotion in speech data is a very challenging task. The deep neural network (DNN) is a feed-forward neural network that has more than one hidden layers between its inputs and outputs. It is capable of learning high- level representation from the raw features and effectively classifying data. After parallel processing the text and speech through respective mechanism, the weighted average of each emotion class is calculated. Finally the emotion class which contains the highest average value is depicted as a result.

### 2. RELATED WORKS

Fukuhura. T, Nakagawa.H and Nishida.T[1] proposed a technique called Temporal Sentiment analysis

produces two types of graphs topic graph and sentiment graph by analyzing the tendency of people's sentiment along with its timeline. To do this analyze the former method used are Mood views, Topic Detection and Tracking, and Theme River. Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vann iZavarella, Erik Van Der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva [2] proposed a technique to analyze the sentiment occurred in a news article from the point of view of author, reader and the text. It provides a way to separate the good and bad news from the domain and also provides the exact sentiment depicted by the text without the need of any other intuitive knowledge.

Doddi, K.S. Haribhakta, M.Y.V. Kulkarni et al, [3] proposed a method to create a positive environment by affording a platform for serving only good news. It is achieved by analyzing the sentiments of the news article and then filtering out the articles which contains only the positive emotion. For classification they used Naïve Bayes, Support Vector Machine, and Maximum Entropy algorithm. E.Cambria, D.Olsher, D.Rajagopal [4] proposes a publicly available semantic and affective resource called SenticNet which is used for concept level sentiment analysis to depict the sentiment.

Ubale Swati, Chilekar Pranali, Sonkamble Pragati[5] delivers the conceptual framework of the News Sentiment Analysis. The core objective of this paper is to analyze the sentiment of text from online published news articles that mentions company news, and identifying positive and negative sentiment that exist in that article and further summarizing the article polarity. H. Saif, Y. He, H. Alani, and M. Fernandez [6] delivers a method which investigates whether removing stop-words helps or hampers the effectiveness of sentiment classification. It uses famous Method to reduce the noisy nature is to remove stop words by using pre-compiled lists of stop- words or by using dynamic stop-word identification by using more sophisticated methods.

D. Jiang, X. Luo, J. Xuan, and Z. Xu [7] proposes a paper which finds out the multidimensional emotions like joy, anger, sorrow, love, surprise, and fear with the help of the semantics of the text written. For that it uses word emotion association network which is of like association link network. Vineet John, Olga Vechtomova [8] delivers a paper which describes the UWaterloo affect prediction system developed for EMOINT 2017. They tend to find out emotion intensity signals from tweets in an ensemble learning approach. This System utilizes gradient boosted regression as the primary learning technique to predict the final emotion intensities.

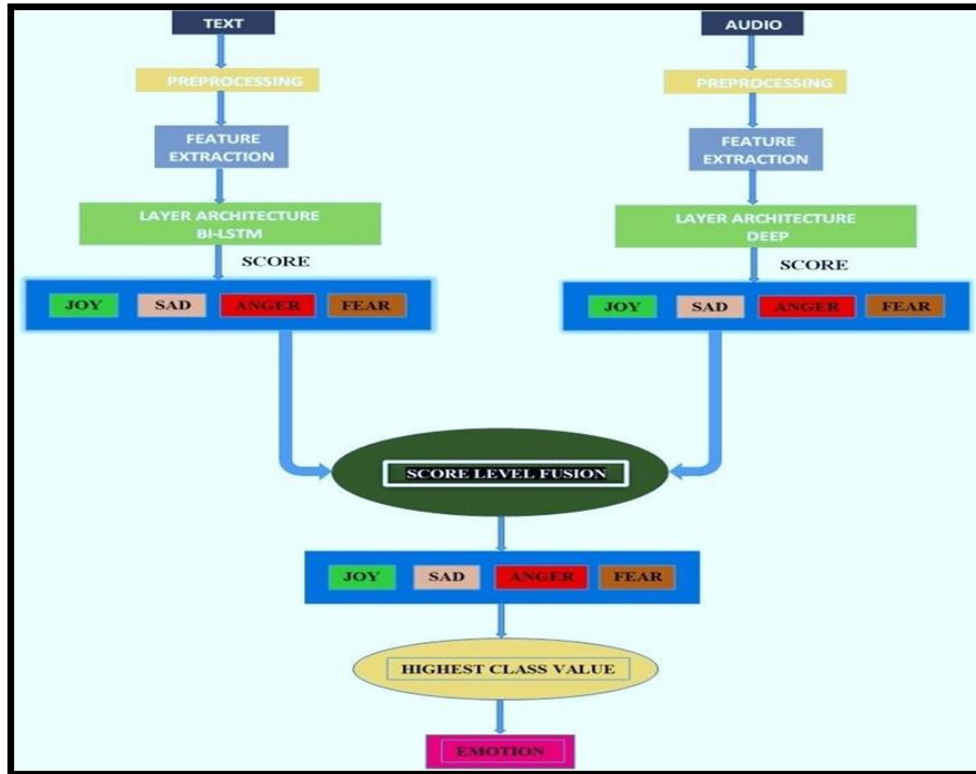
H. Saif, Y. He, M. Fernandez, and H. Alani T [9] proposed a novel semantic based sentiment representation of words called SentiCircle. It takes the co-occurrence patterns of words in various contexts in tweets to seizure their semantics and also updates their pre-assigned strength and polarity in sentiment lexicons accordingly. K.Thendral, S.Chitrakala [10] proposes a technique which uses the emotion term model, topic model and, emotion topic model to find the correlation between social emotions and affective terms to predict the sentiment from a text.

Yong – sooseol and Dong - jookim [11] proposed a hybrid system which uses keyword based and machine learning based method to identify the emotion depicted by the text. Christos Troussas, Maria Virvou [12] presented the model for sentiment analysis of Facebook status using Naive Bayes Classifier. The proposed system predicts whether the depicted emotion is positive, negative or neutral by using sentence level classification. Tejasvinipatil and SachinPatil [13] the author proposed a novel approach for emotion estimation which provides a visual image generation approach that generates images according to emotion in text.

R. A. Calix, S. A. Mallepudi, B. C. B. Chen, and G. M. Knapp [14] proposed a system which uses supervised machine learning technique to automatically identify the emotional state in the text which can be used to render respective facial expressions. J. Kaur and J. R. Saini [15] proposed a paper which analyses the classification of emotions depicted in both formal and informal text writings. For classification it uses different machine learning based methods SVM (support vector machine), NB (Naïve Bayes), Decision Tree. Wojtkowski W, Wojtkowski G, Wrycza S, Zupancic J [16] contrives a survey paper that delves an investigation on emotion recognition from audio channels. It automatizes various techniques of classification, databases procurable for test, adequate features needed to be extracted for analysis and evaluate the performace of various selection methods.

Batliner A, Fischer K, Huber R, Spilker J, Nolth E [17] proposes a paper which addresses the issue of various troubles in communication. It examines the Automatic dialogue systems which is used to understand the emotions in voice using prosodic features and find out it was not accurate in it predictions. So that they propose the module which comes up with a combination of prosodic feature and knowledge sources called “Monitoring of User State Emotion” which results in a most enduring modeling of trouble in communication.

Athanaselis T, Bakamidis S, Dologlou I and et al. [18], delivers a language model constructed by the emotional utterance to increase the recognition rate of emotion in spontaneous speech. This model is derived from an already existing corpus called British National Corpus and it increases the recognition rate by about 20%. Han, K.; Yu, D.; and Tashev, I. [19] proposes a paper to solve the issues of what are the features needs to be extracted for predicting emotion in a speech. It utilizes deep neural network (DNN) to predict the emotion in a speech segment. Then it uses the extreme learning machine (ELM) to identify the sentiment in speech with help of



utterance-level features constructed based on segment level feature. The results indicate that this technique substantially boosts the performance of emotion recognition from speech signals.

Bertero, D., and Fung, P., [20] delivers a real-time CNN model to detect the emotion in speech. The proposed CNN model detects the emotion more accurate and very faster. It classifies the predicted emotion into three main classes of emotions: “Angry”, “Happy”, and “Sad”. In future they tend to increase its CNN model speed to detect emotion in speech for real time human machine interaction application. Liu, Z.-T.; Wu, M.; Cao, W.-H.; Mao.,

and et al.[21], r proposes a speech emotion recognition framework to classify the emotion into six class called neutral, happy, sad, fear, angry and surprise. It uses ELM (Extreme Learning Machine) decision tree for emotion classification and it also uses fisher criterion to extract the redundant features from the extracted feature set. As a result it performs very efficient and fast to identify the emotion from the speech of different speakers.

Atassi H, Esposito [22] A proposes a technique to classify the emotions in speech independent of the speaker. It uses two levels of steps to classify the emotion. At first it uses Gaussian Mixture Model (GMM) classifier to classify six emotions with the help of selected acoustic features. Then it uses Sequential Floating Forward Selection (SFFS) algorithm to find out the highest rated emotion among all others.

### 3. METHODOLOGY

This paper proposed a framework of bi- modal emotional recognition system (SFBM\_TS) based on score level fusion of text and speech. The proposed fusion model contains two branches and works parallel to classify the emotion into four class named joy, sad, fear and anger. One branch is for text analysis and another one is for speech analysis. Each branch contains multiple features to examine. Text based branch is analyzed through selective lexicon based BI-LSTM (SL+BILSTM) method. On the other side speech based branch is analyzed parallelly through deep neural network (DNN). Each branch analyzed and predicted the emotion into four classes through its respective method. Then the weighted average of each emotion class from both sides is calculated. To precise the emotion classification, text has given 0.6 weights and speech has given 0.4 weights as shown in Eq.(1).

$$\text{Average\_Score} = (w1*\text{text}+w2*\text{speech})/2 \text{ whereas } (w1= 0.6 \text{ and } w2=0.4) \quad (1)$$

#### [Fig-1: SFBM\_TS MODEL WORK FLOW]

##### 3.1TEXT

Selective lexicon based BI-LSTM [24] method is used to predict the emotion depicted by the text. SL+BI-LSTM model works as follow.

- At first the text is undergone through the most indispensable step preprocessing.
- Then it selects the most affective word called selective lexica among all the others in the text.
- And then it extracts the features for that selective lexicon and builds the final vector representation.
- This final vector is given as a input to the BI-LSTM network to predict the emotion depicted by the text into four class labels called joy, fear, anger and sad.

##### 3.2AUDIO

In deep learning, the multi-step feature transformation is used to transform the original data feature into a feature representation. Further this feature representation is given as an input into the classification function to get the final result. Here Emotion prediction in Speech is done through deep neural network model.

###### 3.2.1Preprocessing

At first we preprocess the raw audio signal from the speech data. For that we use Fast Fourier Transformation (FTT) transformation to convert the audio signal from time domain into frequency domain. Then we split the speech signal into small window of size 2. Windowing is very important to segment the speech data. Now the signal is divided into its frequency components.

###### 3.2.2Feature Extraction

Audio is a very complex type of data to be understood by the machine. To capture the emotion depicted by the speech data then it's very necessary to extract the robust features from the speech. Feature extraction is a process of extracting the most of the data to make an understandable feature representation for processing by the machine. The following features are extracted from the audio signal.

###### 1. Zero Crossing Rate

ZCR is the foremost parameter for voiced/unvoiced classification. ZCR is low for the voiced part and high for the unvoiced part. It determines the rate at which the signal changes from positive to zero to negative or from negative to zero to positive.

###### 2. Energy

Energy is another key parameter that classifies the voiced/unvoiced parts of the speech. The amplitude of the voiced segment is apparently higher than that of the unvoiced segment. In the speech, the unvoiced part has

lower energy whereas the voiced part has higher energy.

### 3. Energy Entropy

Entropy states the measure of energy dispersal. The Voiced region of the speech induces lower entropy whereas the unvoiced part ends up in higher entropy. Low entropy specifies the voice is of high quality without any noise when in fact the high entropy determines the noise in speech.

### 4. Spectral Centroid

It is a measure to characterize a spectrum in digital signal processing. It determines the location of the center of mass of the spectrum. Eventually, it has a robust connection with the impression of the brightness of a sound. It is the finest predictor of the brightness of a sound henceforth it is vastly utilized as an automatic measure of musical timbre in digital audio and music processing.

### 5. Spectral Spread

The spectral spread feature is used to spread the signal as of its name. It is used for transmitting the signal. This technique is used to make a signal with particular bandwidth into a wider bandwidth signal by diligently spreading the signal in the frequency domain.

### 6. Spectral flux

It specifies the measure of how instantly the power spectrum of a signal is changing. Euclidean distance is used to calculate the spectral flux measure. Euclidean distance is calculated between the powers spectrum of one frame with the power spectrum of the previous frame. It can be used to assess the timbre of an audio signal.

### 7. Spectral roll-off

It is the degree of the amount of the right skewedness of the power spectrum. The spectral roll-off can be used to discriminate between noisy and harmonic sounds. Below roll-off Frequency indicates the harmonic sound and above roll-off frequency determines the noisy sounds.

### 8. MFCCs

It is used to identify the texture of the speech. The complete shape of a spectral envelope is illustrated by Mel Frequency Cepstral Coefficients. A signal's MFCC's is composed of the small set of features conventionally the first 13 coefficients of MFCC are taken as features that designate the envelope of the spectra.

### 9. Chroma

The diction Chromagram or Chroma feature closely relates to the twelve different pitch classes. It specifies the quality of a pitch class which consigns to the "color" of the musical pitch, which can be disintegrated into a "pitch height" that indicates the octave pitch and an octave-invariant value called "Chroma". There are two main Chroma features. They are Chroma vector and Chroma deviation.

#### 9.1 Chroma Vector

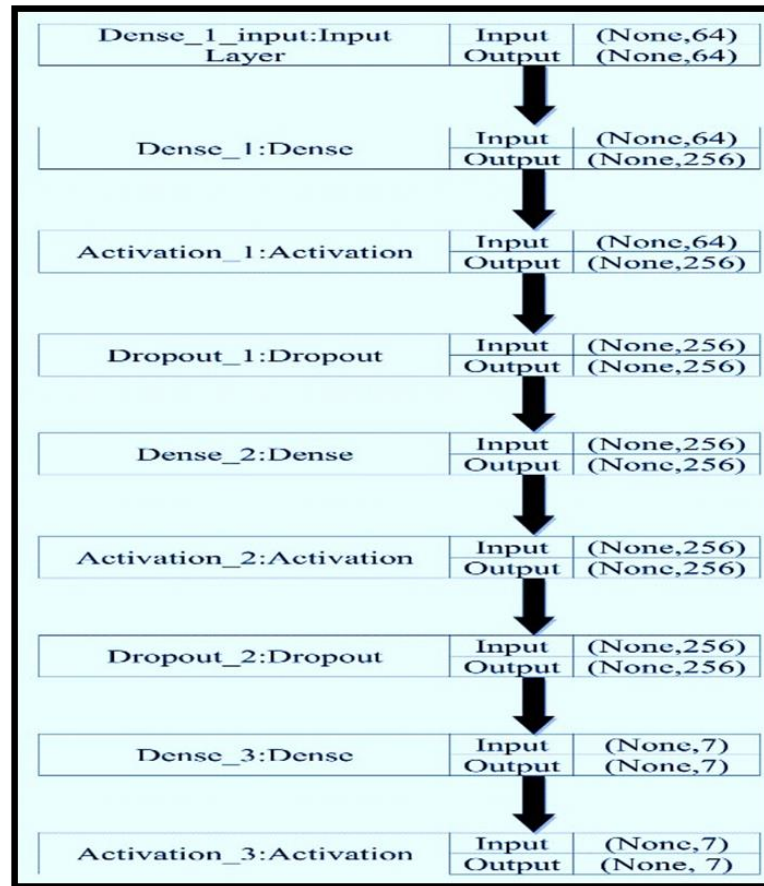
A Chroma vector is indicatively a 12- element feature vector that specifies the amount of energy of each pitch class that appears in the signal. In a standard chromatic scale, the Chroma vector represents magnitudes in twelve pitch classes.

#### 9.2 Chroma deviation

It is the standard deviation of 12 Chroma coefficients.

### 3.2.3 Layer Architecture

At premiere dense input layer is used of units 64. Then it is followed by 1 dense layer of units 64 with ReLU activation. Next this is followed by 0.2 dropout layer. Then the input is given to the dense layer of size 256 with ReLU activation. Next it is go through the 0.2 drop out layer. Now the input is given to the dense layer



[Fig-2: DNN Layer Specification]

of size 7 with ReLU activation. Next it is go through the 0.2 drop out layer. At the last it goes through the soft max layer for classification. Finally the model is compiled with stochastic gradient descent (SGD) optimizer and cross entropy loss function. From the DNN network model the result came out with containing values for all the four emotional classes. Now the fusion model SFBM\_TS picks the classified scores of all the four emotion labels from both the branch and calculate the weighted average scores of each emotion class. To calculate the emotion more precisely SFBM\_TS modal assigns 0.6 weights to text and 0.4 weights to speech. Finally the emotion class which has the highest score is depicted as the emotion delivered by the given bimodal data.

#### 4. DATASET

MELD [23] is a dataset used here for experiment. It is an emotion recognition dataset comprises of multimodal multiparty conversations. It contains audio segments transcripts and raw videos for multimodal processing. MELD includes more than 13000 emotion labeled utterances, from which we take 5000 utterances on joy, sad, anger and fear emotions for experiment. Out of which 3800 utterances are taken for training and 1200 utterances are used for testing.

#### 5. PERFORMANCE METRICS

Performance metrics is used to evaluate the classification model used in the proposed method. Accuracy, Precision, Recall and F1 Score are the specific performance metric used in this proposed work. The procedures used to evaluate the performance metric are listed below.

- **True Positives (TP)** - It specifies the correctly predicted positive values. That is the value of both the actual class and predicted class is also yes.
  - **True Negatives (TN)** - It specifies the correctly predicted negative values. That is the value of both the actual class and predicted class is also no.
- False positives and false negatives occur when the values of both actual and predicted class are contradicted with

each other.

- **False Positives (FP)** - It specifies when the value of actual class is no and the value of predicted class is yes.
- **False Negatives (FN)** -It specifies when the value of actual class is yes and the value of predicted class is no.

**Accuracy** - It is the most inherent performance measure. It is the ratio of correctly predicted observation to the overall observations.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

**Precision** – It is the ratio of correctly predicted positive observations to the overall predicted positive observations.

$$\text{Precision} = \frac{TP}{TP+FP}$$

**Recall** – It is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

$$\text{Recall} = \frac{TP}{TP+FN}$$

**F1 score** - F1 is generally one of the most useful metric. The weighted average between Precision and Recall is the F1 score.

$$\text{F1 Score} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

## 6. RESULT ANALYSIS

Table 1.1 shows the results of the performance of all the models

| S.NO | MODELS         | ACCURACY | PRECISION | RECALL | F1 SCORE |
|------|----------------|----------|-----------|--------|----------|
| 1.   | Text_CNN       | 0.7257   | 0.7156    | 0.7171 | 0.7164   |
| 2.   | Text_LSTM      | 0.742    | 0.7295    | 0.7417 | 0.7356   |
| 3.   | Text_SL+BILSTM | 0.783    | 0.751     | 0.7588 | 0.7549   |
| 4.   | Audio_Deep     | 0.7889   | 0.7574    | 0.7648 | 0.7611   |
| 5.   | SFBM_TS        | 0.827    | 0.7906    | 0.8089 | 0.7997   |

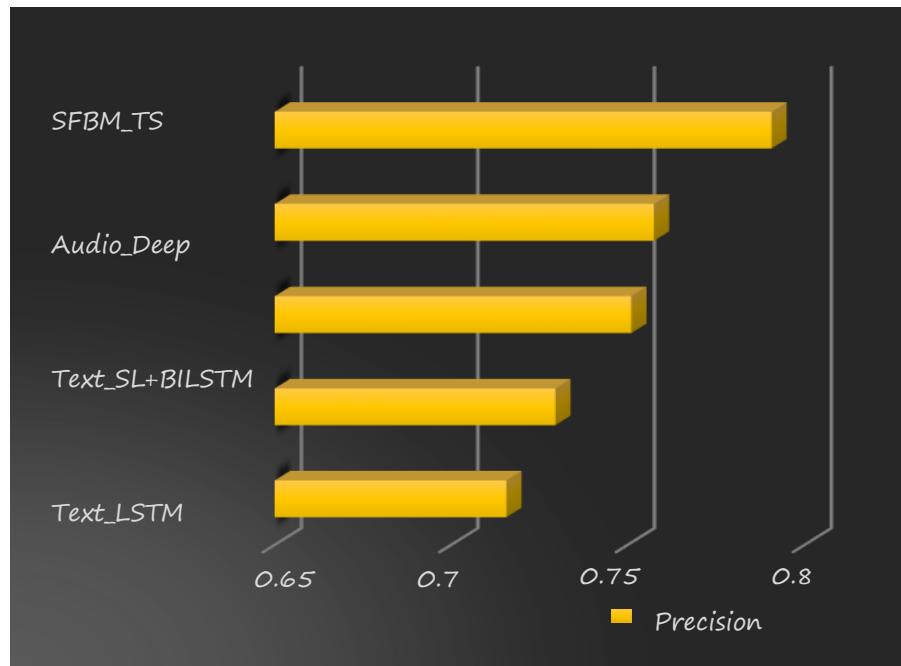
[TABLE-1: Performance Metrics of the Models]

Graph [1] shows the accuracy of Text\_CNN, Text\_LSTM,Text\_SL+BILSTM,Audio\_Deep and SFBM\_TS models. In terms of accuracy the model Text\_CNN scores 0.7257, Text\_LSTM scores 0.7420, Text\_SL+BILSTM scores 0.7830, Audio\_Deep scores 0.7889 and SFBM\_TS scores 0.8270. Compared to all model the fusion model SFBM\_TS comes up with high score than others. It scores 4% greater than Text\_SL+BILSTM model.



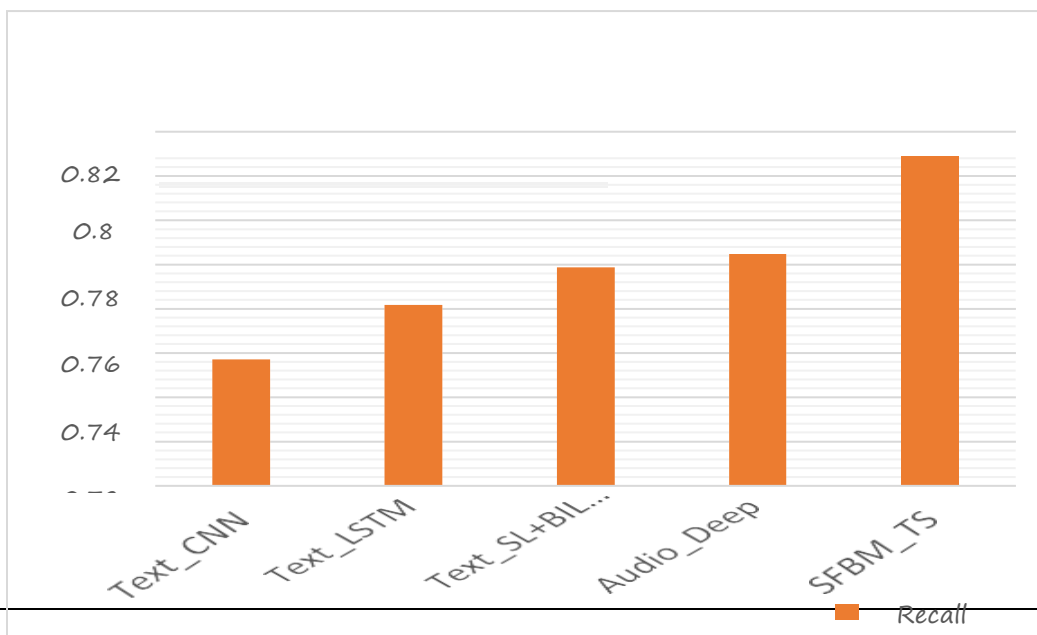
**[Graph 1: Representation of Accuracy of all models]**

Graph [2] shows the precision of Text\_CNN, Text\_LSTM, Text\_SL+BILSTM, Audio\_Deep and SFBM\_TS models. In terms of accuracy the model Text\_CNN scores 0.7156, Text\_LSTM scores 0.7295, Text\_SL+BILSTM scores 0.7510, Audio\_Deep scores 0.7554 and SFBM\_TS scores 0.7906. Compared to all model the fusion model SFBM\_TS comes up with high score than others. It scores 4% greater than Text\_SL+BILSTM model.



**[Graph 2: Representation of Precision of all models]**

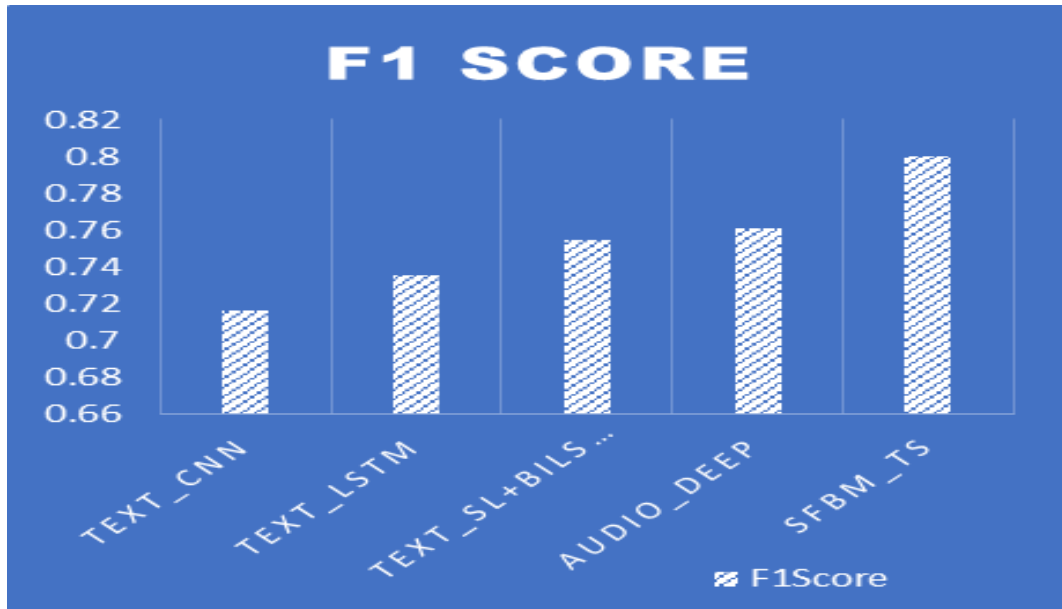
Graph [3] shows the recall of Text\_CNN, Text\_LSTM, Text\_SL+BILSTM, Audio\_Deep and SFBM\_TS models. In recall metrics the model Text\_CNN scores 0.7171, Text\_LSTM scores 0.7417, Text\_SL+BILSTM scores 0.7588, Audio\_Deep scores 0.7647 and SFBM\_TS scores 0.8089. Compared to all model SFBM\_TS comes up with high score than others. The model SFBM\_TS scores 5% greater than Text\_SL+BILSTM model.





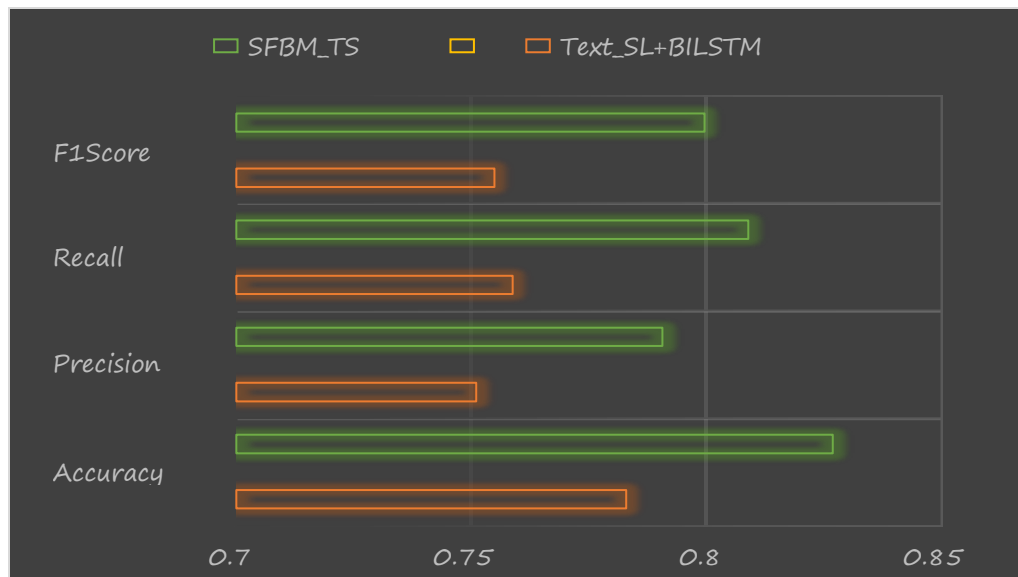
**[Graph 3: Representation of Recall of all models]**

Graph [4] shows the F1 score of Text\_CNN, Text\_LSTM, Text\_SL+BILSTM, Audio\_Deep and SFBM\_TS models. In metrics F1, the model Text\_CNN scores 0.7164, Text\_LSTM scores 0.7356, Text\_SL+BILSTM scores 0.7549, Audio\_Deep scores 0.7610 and SFBM\_TS scores 0.7997. Compared to all models the fusion model comes up with high score than others. F1 score of SFBM\_TS model is 5% greater than Text\_SL+BILSTM models F1 score.



**[Graph 4: Representation of F1-SCORE of all models]**

Graph [5] shows the performance comparison of SFBM\_TS model and Text\_SL+BILSTM. It shows the proposed fusion model SFBM\_TS secure high scores in performance metrics than the Text\_SL+BILSTM model. It shows that the emotion is classified precisely when text and speech both are used in classification process than the text alone is used.



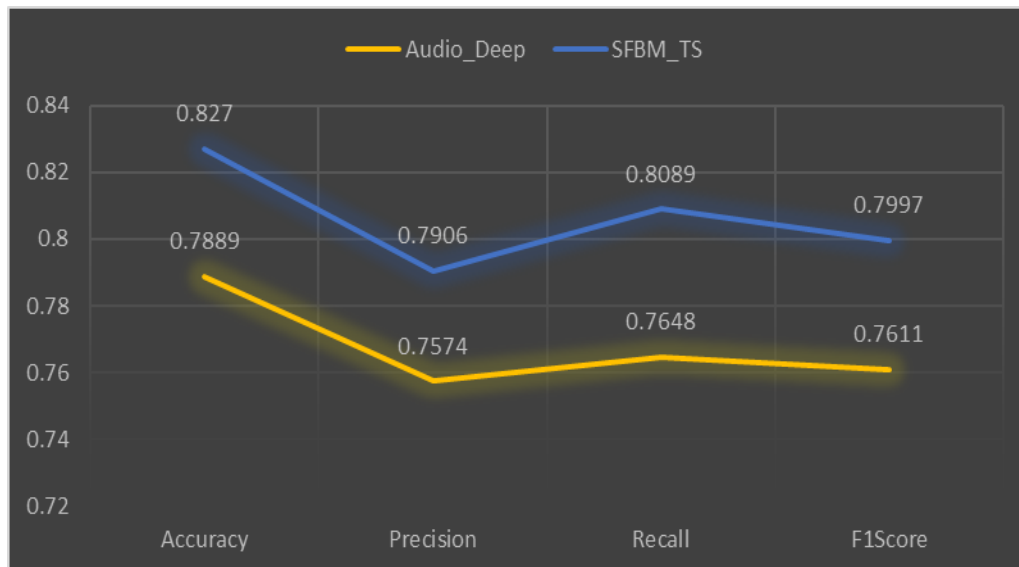
**[Graph 5: Comparison between SFBM\_TS and Text\_SL+BILSTM models]**

As a result SFBM\_TS model scores 4% in accuracy, 4% in precision, 5% in recall and 5% in f1 score more than the Text\_SL+BILSTM model.

Graph [6] shows the comparison of Audio\_Deep model and proposed fusion model SFBM\_TS. As a result SFBM\_TS model secure high scores in performance metrics than the Audio\_Deep model. SFBM\_TS model scores 4%in accuracy, 4% in precision,5%in recall and 5% in f1 score than Audio\_Deep model. It shows that emotion classification can be done more accurately by using both text and speech than speech alone.

**[Graph 6: Comparison between SFBM\_TS and Audio\_Deep models]**

6.



**CONCLUSION**

With the procreation in social media, the single emotion prediction system is not sufficient because a variety of domains like text, audio etc., are used to express the emotions. Each domain such as text, speech has its own pros and cons to predict the emotion. The single emotion prediction system is not able to obtain the emotion prediction precisely. Because In text sometimes it may contain weakly labeled data likewise speech data can also

have imbalanced tones in its. In such situation using of both text and audio can balance each other and make the emotion prediction more precisely. To optimize the performance of the emotion recognition this paper proposes a novel technique called a score level fusion of emotion recognition system composed of text and speech data. This model is composed of two branches to work in parallel one is for text data analysis done by selective-lexicon based BI-LSTM model and another branch is for audio data analysis by deep neural network model. Both the branches predicted the result into four emotional classes then the weighted average of each class is taken. The emotion class which has the highest value is depicted as resulting emotion. Experimental results showed that the emotional recognition accuracy of the proposed model is higher than that of all other models.

## REFERENCES

1. Fukuhara, T., Nakagawa, H., Nishida, T., 2007. Understanding Sentiment of People from News Articles: Temporal Sentiment Analysis of Social Events. In: ICWSM.
2. Alexandra Balahur, Ralf Steinberger, MijailKadjov, VanniZavarella, Erik Van Der Goot, MatinaHalkia, Bruno Pouliquen, and JenyaBelyaeva. Sentiment analysis in the news. arXiv preprint arXiv:1309.6202, 2013.
3. Doddi, K.S. Haribhakta, M.Y.V. Kulkarni, P. (2014)“Sentiment Classification of News Articles”, KiranShriniwasDoddi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , pp. 4621-4623
4. E.Cambria, D.Olsher ,D.Rajagopal “Senticnet 3: A Common and common- sense knowledge base for cognition- driven sentiment analysis”, AAAI Conference on Artificial Intelligence (2014).
5. 1Ubale Swati,2Chilekar Pranali, 3Sonkamble Pragati “Sentiment Analysis Of News Articles Using Machine Learning Approach”,International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835 Volume-2, Issue-4, April-2015
6. H. Saif, Y. He, H. Alani, and M. Fernandez, “On stopwords, filtering and data sparsity for sentiment analysis of twitter,” in The International Conference on Language Resources and Evaluation, 2014.
7. D. Jiang, X. Luo, J. Xuan, and Z. Xu,”Sentiment Computing for the News Event Based on the Social Media Big Data”,*IEEE Access* (2017)
8. VineetJohn, Olga Vechtomova, “UWat- Emote at EmoInt-2017: *Emotion Intensity Detection using. Affect Clues, Sentiment Polarity and Word Embeddings*”
9. H. Saif, Y. He, M. Fernandez, and H. Alani, “Contextual semantics for sentiment analysis of twitter,” *Information Processing & Management*, vol. 52, no. 1, pp. 5–19, 20
10. K.Thendral, S.Chitrakala , “*Emotion Recognition System for Affective Text*”, ISSN (Online): 2347 - 2812, Volume-2, Issue -11,12 2014
11. Yong-sooseol and Dong-jookim, ”Emotion recognition from text using Knowledge-based ANN”, ITC-CSCC 2008
12. Christos Troussas, Maria Virvou, ”Sentiment analysis of facebook statuses using Naive Bayes classifier for languagelearning”, 2012.
13. Tejasvinipatil and SachinPatil,”Automatic generation of emotions for social networking websites using text mining”in proceedings of IEEE-2012.
14. R. A. Calix, S. A. Mallepudi, B. C. B. Chen, and G. M. Knapp. Emotion recognition in text for 3-d facial expression rendering, 2010.
15. J. Kaur and J. R. Saini ,”Emotion Detection and Sentiment Analysis in Text Corpus: A Differential Study with Informal and Formal Writing Styles”, 2014
16. Anagnostopoulos CN, Vovoli E (2010) Sound processing features for speaker-dependent and phrase-independent emotion recognition in Berlin Database. In: Papadopoulos GA, Wojtkowski W, Wojtkowski G, Wrycza S, Zupancic J (eds) Information systems development, pp 413–421
17. Batliner A, Fischer K, Huber R, Spilker J, Nolth E (2003) How to find trouble in communication. *Speech Commun* 40: 117–143
18. Athanaselis T, Bakamidis S, Dologlou I, Cowie R, Douglas-Cowie E, Cox C (2005) ASR for emotional speech: clarifying the issues and enhancing performance. *Neural Netw* 18: 437–444
19. [Han, Yu, and Tashev 2014] Han, K.; Yu, D.; and Tashev, I. 2014. Speech emotion recognition using deep neural network and extreme learning machine. In The fifteenth annual conference of the international speech communication association (INTERSPEECH), 223–227.

21. [Bertero and Fung 2017] Bertero, D., and Fung, P. 2017. A first look into a convolutional neural network for speech emotion detection. In 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), 5115– 5119. IEEE.
22. [Liu et al. 2018] Liu, Z.-T.; Wu, M.; Cao, W.-H.; Mao, J.-W.; Xu, J.-P.; and Tan, G.-Z. 2018. Speech emotion recognition based on feature selection and extreme learning machine decision tree. *Neurocomputing* 273:271–280.
23. Atassi H, Esposito A (2008) A speaker independent approach to the classification of emotional vocal expressions. In:Proceedings of 20<sup>th</sup> IEEE international conference on tools with artificial intelligence, pp 147-152.
24. S. Poria, D.Hazarika, N. Majumder, G. Naikand et.al,MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversation. (2018)
25. ShunmugaSundari A, Shenbagavalli P (2019) Dominant Lexicon based BI- LSTM for emotion prediction on a text.