IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 October 2023, accepted 5 November 2023, date of publication 8 November 2023, date of current version 16 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3331226

RESEARCH ARTICLE

Audio and Text Sentiment Analysis of Radio Broadcasts

NAMAN DHARIWAL [1], (Student Member, IEEE), SRI CHANDER AKUNURI [1], SHIVAMA [1], AND K. SHARMILA BANU [2]
[1]Department of Software Systems, Vellore Institute of Technology, Vellore 632014, India
[2]Department of Internet of Things (IoT), Vellore Institute of Technology, Vellore 632014, India
Corresponding author: K. Sharmila Banu (sharmilabanu.k@vit.ac.in)

ABSTRACT The rapid growth of radio broadcast services has created a vast amount of audio data that can provide insights into public opinion and emotions. This research extends the boundaries of sentiment analysis to audio data and aims to propose a computational approach which is termed as bifurcate and mix, for sentiment analysis and emotion detection that leverages advanced natural language processing techniques from audio sentiment analysis tools like Vokaturi, transcription services like AssemblyAI and text sentiment analysis lexicons like VADER to extract and categorize sentiments from the audio data to generate a more efficient real-time sentiment analysis model. The results of the analysis reveal patterns and trends in the sentiments expressed in radio broadcasts by the News Service Division: All India Radio - 'Akashwani'. This research's methodology will contribute to the development of novel applications for sentiment analysis in the media industry and provide valuable insights into public opinion and emotions.

INDEX TERMS Sentiment analysis, radio broadcasts, emotions, news, audio, natural language processing.

I. INTRODUCTION

Being able to use tools for our advantage and advancement is one of humanity's biggest assets. As global technologies move towards artificial intelligence, the ability of computers to understand and analyze various human languages is also bound to evolve in multi-faceted ways. A user is now free to use text, audio, speech, and gestures, which a computer understands using a subfield of artificial intelligence- natural language processing. It gives machines the ability to decode and interpret human languages and perform desired operations on them based on the various procedures pertaining to its many applications.

Sentiment analysis, a subsidiary domain application of natural language processing, is one such tool which is being restricted to the usage of decision making for companies or various algorithms and majorly text data [1].

To understand why sentiment analysis is such a game changer, let's understand what 'sentiment' is and further, what 'sentiment analysis' is. According to the Cambridge Dictionary, 'sentiment' refers to a thought, opinion, or idea based on a feeling about a situation, or a way of thinking about something. The extraction of this feeling and sentiments out of different forms of data using a computer is the goal of sentiment analysis models.

The restriction of sentiment analysis to majorly text data poses a major drawback in the development of sophisticated computer systems. This paper extends the boundaries of sentiment analysis to audio data and will further expand on the concept by proportionally combining the sentiment analysis of text and audio data later in the paper.

The major benefit that speech and audio data has over textual data is the presence of emotions. Communication and interaction between humans are a mixture of emotions like fear, anger, happiness, sadness, etc. These emotions tend to blur out in textual and written formats. Thus, the usage of audio to identify and analyze complex emotions in speech and human interaction acts as an additional feedback mechanism without altering the linguistic contents and meaning intended. The spoken human communication is divided into two channels: the primary channel and the secondary channel [9], [10]. The primary channel is associated with the syntactic and semantic sections of speech. The secondary channel conveys the paralinguistic part associated with speech such

The associate editor coordinating the review of this manuscript and approving it for publication was Maria Chiara Caschera.

as tone, pitch, emotional state, and gestures. Most of the current speech emotion analysis methods convert audio data of speech to textual data and then extract sentiments, thus, only dealing with the primary channel as the secondary channel is lost in translation. It is when the secondary channel is processed that aids in the ability to perform the various functions that help increase convergence, better the degree of judgement, avoid misunderstandings, analyze sarcastic and rhetoric nature of speech, and allow access to additional information about speakers like age, gender, nativity, etc.

The audio data for the model proposed in the paper comes from various sources. This paper also proposes to use two such sources: radio stations and modern day highly connected internet for various purposes that could help illuminate their usage with sentiment analysis.

By the end of 2023, there will be 5.3 billion internet users worldwide. For the period from 2018 to 2023, the compound annual growth rate is 6%. With 300 million additional internet users and a growth rate of 7.7% in 2019 as compared to 2018, 2019 saw the fastest increase throughout this time [2]. It is predicted that audio based social media will be a major part of human interactions in the future.

On the other hand, radio stations have been favored by the masses since the 1920's and drastically faded out of usage as a means of entertainment after the advent of television into the market. Yet, they have beaten all odds by using every possible opportunity and so, are highly numbered as well as distributed compared to the television stations. Though there has been a decline in the number of human radio users, it is still expected to affect around 3.17 billion people with a user penetration of 39.8% by the year 2027 [3]. Radio broadcasts indeed are the best example of how audio data has always been an integral part of human communication and it is thus, reasonable to posit that radio stations are a reliable source of audio data with prolonged consistency for audio analysis that can be broadcasted.

As mentioned earlier, sentiment analysis is an application of natural language processing, thus, there are various methodologies that are being followed. Broadly introducing, a few of them are classification using n-grams, parts-of-speech tagging and machine learning; identification of semantic orientation of words using lexicons and training documents; identification of semantic orientation of sentences and phrases but considering negation words; identification of semantic orientation of documents; objective feature extraction and comparative sentence identification and many more [4]. Of these various procedures, some like neural networks, gaussian mixture models etc. can be appropriately used to classify audio data and extract sentiments [6], [7]. Such procedures are aptly used by emotion detection software like Vokaturi, EmoVoice etc. which will be used in this research to elucidate upon the methodology.

The paper by Devarajan et al. [30] proposes a fake news detection model using deep natural language processing structured in four layers: publisher, social media networking, enabled edge, and cloud layer. The methodology comprises four steps: data acquisition, information retrieval, natural language processing-based data processing with feature extraction, and deep learning-based classification. Utilizing credibility scores of publishers, users, messages, and headlines, the model classifies news articles as fake or real. Evaluated on three datasets: Buzzface, FakeNewsNet, and Twitter, the model exhibited superior performance with an average accuracy of 99.72% and an F1 score of 98.33%, outperforming existing methods in fake news detection.

In 2018, Murugan and Devi [31], a hybrid methodology combining a decision tree, particle swarm optimization, and genetic algorithm is proposed for real-time Twitter spam detection. The research utilized a large dataset of 600 million tweets collected via a URL-based tool. The proposed approach, tested against other hybrid algorithms, demonstrated enhanced spam detection rates, thanks to the incorporation of machine learning and evolutionary algorithms. Central to this method is a decision tree-based classification model, aiding in accurate spam tweet categorization and contributing to a spam-reduced Twitter streaming environment.

The 2021 paper by Linda et al. [32] introduces a CNN-CapsNet model for recognizing viewpoint variations in gait and speech. This model automates feature representation learning, using capsule vectors as neurons to encode image spatial information. It was tested in three scenarios: fixed-view, cross-view, and multi-view, showcasing its effectiveness. Key gait recognition parameters like speed, clothing changes, object carrying, and light intensity were well handled by the system. Compared to other techniques, the CNN-CapsNet model displayed superior performance in gait and speech recognition across various scenarios, marking a significant advancement in the field of intelligent recognition systems.

In the 2016 research by Al-Rawi et al [33], the engagement dynamics on radio stations' social media platforms are explored using the webometric tool NVivo 10-Ncapture to analyze 184,204 comments from 15,163 news stories on Radio Monte Carlo and RNW's Facebook pages. The study revealed a male-dominated engagement and examined multilingual posts and comments to understand engagement patterns. It found higher engagement on posts encouraging discussions on broader issues or clever quotes, and less on breaking news. Notably, users actively engaged with contrasting views. Radio Monte Carlo's page showed frequent use of religious terms and a focus on the Syrian crisis, highlighting the evolving dynamics of radio audience engagement in the digital era.

In 2022, Joloudari et al [34] conducted the sentiment analysis of COVID-19 tweets using BERT and Deep CNN models. The study aims to understand public sentiment during the global health crisis through these advanced machine learning frameworks. The paper highlights the superior performance of BERT models over other deep models in sentiment analysis, a conclusion drawn from a comparative analysis of various research studies. This research underscores

TABLE 1. Comparison of speech emotion detection technologies.

Name	API/SDK	Requires Internet	Information returned	Difficulty of use	Free Software
Beyond Verbal	API	Yes	Temper Arousal Valence Mood (Up to 432 emotions)	Low	No
Vokaturi	SDK	No	Happiness, neutrality, sadness, anger, and fear	Medium	Yes
EmoVoice	SDK	No	Determined by developer	High	Yes
Good Vibrations	SDK	-	Happy level, relaxed level, angry level, scared level, and bored level	Medium	No

Source: Jose, Maria, Garcia-Garcia., Victor, M., R., Penichet., María, Dolores, Lozano. "Emotion detection: a technology review." null (2017).:8-. doi: 10.1145/3123818.3123852

the importance of employing such advanced frameworks in analyzing public sentiment during significant global events like the COVID-19 pandemic, especially on social media platforms where public sentiment is openly shared and can provide insightful analyses.

In another study [35], a novel model named aPSO-FLSADL is introduced to facilitate personalized recommendations in the realm of consumer electronics, focusing on harnessing the sentiment expressed in user reviews. The method involves utilizing SentiWordNet to ascertain sentiment scores of words and employing BERT, a robust pre-trained deep learning model, to convert text data into word embeddings. Subsequently, a CNN-BiLSTM based Federated Learning model is trained to analyze sentiments globally, with an adapted Particle Swarm Optimization (aPSO) and a mutation operator implemented for optimized learning parameter adjustments in the federated learning environment. Testing and evaluation are performed using the Amazon review and CNET datasets, focusing on metrics like accuracy, loss, and hit ratio. As a result, aPSO-FLSADL not only achieved stellar training and testing accuracy on both datasets but also outperformed baseline models by achieving the maximum hit ratio in providing personalized recommendations for consumer electronics products, thus manifesting its potential applicability and effectiveness in enhancing e-commerce platforms through refined, sentiment-aware recommendations.

After due comprehensive study of literature, it was observed that though research on sentiment analysis has been ongoing for many years, its application to social media and communication technologies is much more recent. Machine learning models, BERT architecture and many more sophisticated models have been duly used by many studies to analyze the sentiment in their datasets. However, maximum focus has been towards text-based sentiment analysis and little to none focus has been on sentiment analysis of audio data. Data from social media applications have been widely used, however, the audio data of the same sources are yet unexplored. Also, radio broadcasts are not yet seen as potential data sources to perform sentiment analysis. The evident gap in the current studies, analysis of audio data and especially that of radio broadcasts to create a basis for further prediction models, is explored in this research. This research gives importance to the vast extractable information from the radio news broadcasts and performs sentiment analysis on the same.

This research aims to:

- Transcribe radio broadcast audio recording into text data and perform sentiment analysis on the generated text.
- Parallelly, also perform emotion detection analysis on the same radio broadcast audio recordings to understand the various emotions in it.
- Combine the findings from the sentiment analysis and emotion detection analysis, in ratio of accuracy of both the analysis models, to generate a unified scale consisting of positive, neutral, and negative sentiments.
- Finally, compare between the various trends observed in sentiment analysis of radio data. Thus, presenting concrete result analysis to prove the importance of combined audio and text sentiment analysis methods.

The resulting sentiment will be a combination of both syntactic-semantics information and paralinguistic information associated with the radio broadcasts. This research will thus help to:

- Analyze the emotions hidden in the radio broadcasts around the world for the identification of possible adverse conditions in various regions.
- Find the true conditions of a region in the most unfiltered ways possible, using sentiment analysis of the local radio broadcasts.
- Understand the importance of incorporating audio sentiment analysis, in addition to text-only sentiment analysis, to improve sentiment analysis.
- Present the importance of non-verbal vocalizations (NVV) and their influence in sentiment analysis.

II. TECHNOLOGY REVIEW

A. SPEECH EMOTION EXTRACTION TECHNOLOGIES

Various technologies have emerged in the recent past that are specifically aimed at the task of extracting emotions and the second channel information from a given audio data source.

TABLE 2. Comparison of speech transcription technologies.

API Models	Accuracy (based on Word Error Rate's)	Models used
CMU SPHINX-4	63%	Hidden Markov Model
Microsoft API	82%	Context Dependent Gaussian Mixture Model Hidden Markov Model
Google API	91%	Deep Learning Neural Networks

Many companies have released API and SDK versions of their software for sentiment analysis. A brief description of the various software is given in table 1.

Beyond Verbal, Vokaturi, Emo Voice and Good Vibrations are a few of the technologies that are available in the market today to facilitate sentiment analysis. Pros and cons of each of these websites have been discussed in [9].

Founded in 2016 and based in Amsterdam, Vokaturi [14] is a speech recognition software company (private) and product that specializes in analyzing and accurately extracting emotions from an audio file.

The software is currently trained to address Paul Ekman's six basic emotions [15] classes namely happiness, sadness, anger, and fear, except surprise and disgust. It also has a fifth class, neutrality. It uses nine acoustic cues such as average pitch (in Hertz), pitch dynamics, average intensity (in decibels), intensity dynamics and spectral scope to compute the mean and standard deviations of the mentioned five emotion classes.

TABLE 3. Accuracy of various transcription API models.

API models	Accuracy (based on Word Error Rate)
AssemblyAI speech-to-text	87.12%
Google API	84.46%
Amazon Transcribe	83.12%
Microsoft Azure	81.01%

Source:https://www.assemblyai.com/blog/speech-to-text-wer-google-aws-may-2020/

The model comprises a pre-trained deep neural network with one input layer, two hidden layers and an output layer. The input layer contains nine nodes, one for each cue, each with the strength of the nine cues. The first hidden layer contains 100 nodes, each having a bias value and weights for each of the nine input nodes (from input layer). Similarly, the second hidden layer consists of 20 nodes, each having their bias values and weights associated with the hundred previous nodes (from the first hidden layer). Finally, the output layer consists of 5 nodes, each representing the 5 output emotions that are analyzed by the model. Each of the 5 nodes have their bias values and weights associated with the 20 previous nodes (from the second hidden layer) [14].

OpenVokaturi is the open-source version of the software. This version does not require the internet to function and runs on the user's system, however, this makes it a bit less powerful than its competitors that run on cloud servers.

This paper uses OpenVokaturi to conduct the paralinguistic component of research.

B. SPEECH RECOGNITION [AUDIO TO TEXT CONVERSION] TECHNOLOGIES

There are various Speech recognition technologies used in place which include APIs from various reputed companies. For the sake of this paper, various comprehensive studies on such API have been reviewed to select an optimum service for the research's specific needs.

The paper [10] has a study where the authors compare the three speech recognition systems and obtain results as in table 2. Whereas as per [26] the accuracy values obtained for various API's are presented in table 3. It's clearly evident that there is a discrepancy in the accuracy data of the Google API, but it can probably be attributed to the small size of data that has been put to test in [10]. It can be clearly inferred from table 3 that AssemblyAI speech-to-text API for natural language processing has a very low word error rate among the other famous technologies available. It also permits audio data as large as 3GB to be used for its transcription services which also is helpful in conversion of radio data at once rather than minuscule chunks. Hence, this API has been deemed to be fit for this paper in converting the audio to text for radio data.

Assembly AI is a natural language processing company that aims at building superhuman systems for transcribing and understanding human speech. Its services boast human level accuracy developed using an assortment of relevant models. It also allows the transcription of audio large data transcription unlike other renowned API's. Its services are used by various companies like Spotify etc. and its models are regularly upgraded.

The Assembly AI's core transcription services have various features. The features that are very prominent for this study are:

- Async transcription: Transcribing pre-recorded audio and/or video files with human accuracy.
- International language support: It supports over 12 languages including English and its dialects.
- Inclusive of all audio and video formats.
- Automatic casing and punctuation.
- Confidence score for each word in the transcript.
- Language detection of audio especially, the dominant language.

C. TEXT SENTIMENT ANALYSIS

Text sentiment analysis is a vastly explored part of the natural language processing domain and many different methodologies and models have been developed pioneering an accurate sentiment detection, the different methodologies to approach this as discussed in table 4.

TABLE 4. Methodologies and approaches to text sentiment analysis.

Methodology	Description	Sub approaches	Characteristics	Examples
Semantic Lexicon Methodology	Semantic Lexicon is a list of lexical features which are generally labelled according to their model.[26]	Semantic Orientation /Polarity Based	Context independent and binary labelling.	LIWC, Harvard GI, Hu-Liu etc.
		Sentiment Intensity / Valance Based	The degree or intensity of various sentiments expressed by a given word is mentioned.	VADER, SentiWordNet ,TextBlob, ANEW, SenticNet etc.
Machine Learning	Machine learning is a subset of Artificial Intelligence used to find and predict required cues from structured data based on the given input data used to train the model	Deep Neural Networks, Supervised, Unsupervised etc.	Very dependent on the trained data.	Naive Bayes Classifier, Support Vector Classifier, Gradient Boosting Classifier, Logistic Regression, Maximum Entropy models, LSTM etc.

VADER is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. It is an open-source sentiment analyzer pre-built library protected under the MIT license and is incorporated into natural language toolkit (NLTK) and is accessible using Python 3 and the accuracy of the newer version of VADER tools is 70% [28]. VADER's unique social context inclination is particularly helpful especially to deal with various types of radio and data, where social media terms could be used.

It is true that machine learning models when trained properly, achieve accuracies greater than 90% as compared to the 80% to near 90%'s of lexicon-based models like VADER. It is to be noted that compared to machine learning models, VADER isn't based on meagre training data. It's more general and can give optimum results in various scenarios that radio broadcasts comprise. Hence for the text sentiment analysis, the paper thus justifiably makes use of VADER for a purely defined output.

III. RADIO BROADCAST DATASET SOURCE

In an attempt to find the most credible and accurate source of radio broadcast recording, various sources like the BBC archives, Archive.org's English audio broadcast data, Library of Congress's audio recording sections, WBAI Radio and many were studied. Following the current trend, most online radio services like BBC Sounds, Audioboom, Mixcloud, etc., have emerged in the market however, focused mainly on subscription-based podcast systems. Hence, alternative sources had to be found. Internet Archive [16] is one of the most popular databases for audio recordings. However, since the entire organization depends solely on its account holders and partners to upload the archives, it is highly likely that one may not find all the desired resources. WBAI [17] is a free speech radio 99.5fm, a noncommercial, open source, user supported radio station broadcasting in New York, New York. However, even though it has a wide range of archives, it has its own limitations. The biggest drawback is the locality of the station only to a small region in New York state and lack of reach to major parts of the world.

A. NEWS SERVICE DIVISION: ALL INDIA RADIO

For the sake of this research, after thorough assessment, the data from India's Public Service Broadcaster, News Service Division: All India Radio (NSD: AIR) has been decided to be an apropos source for the radio broadcast audio recording data [15].

First broadcasted in 1936, The News Service Division of AIR aspires to broadcast news to all parts of the vast country of India across all terrains, 24×7. The service is broadcasted in 22 languages (scheduled under the Constitution of India, eighth schedule). Moreover, they also service the bulletins in 5 other languages including English. The paramount is extended by the 51 dialects and 14 foreign languages such as French, Thai, Chinese, etc., that are also broadcasted to the remotest parts of the nation. The news is broadcasted across 21 bulletins in English (broadcasted for a combined duration of 135 minutes daily) and 23 bulletins in Hindi (broadcasted for a combined duration of 150 minutes daily), the highest consumed language of the service. Similarly, 23 foreign language bulletins are compiled by the NSD, AIR and broadcasted by ESD, AIR.

The main audio archive is divided into two sections: main audio broadcast program and main audio bulletins. Each news bulletin is scheduled from 5 minutes to 15 minutes depending on the broadcasting channel. To add to the ease of use, the NSD, AIR portal provides filters such as bulletin type, where the language can be selected, followed by a subcategory filter which allows the user to select from a drop menu of morning news, midday news, evening news, hourly and parliament bulletin. It further allows the user to select the required date range by accepting start and end date inputs along with the time stamps.

This vast diversity in languages and easy availability of audio samples is the main advantage of NSD, AIR over the other prominent sources listed earlier.

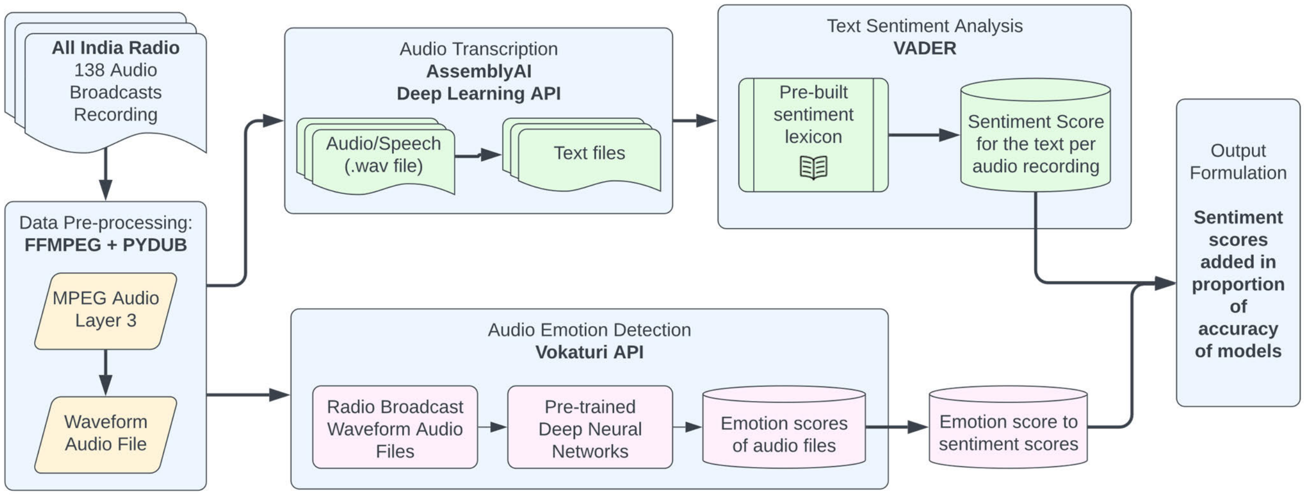

FIGURE 1. **Workflow of proposed methodology.**

IV. PROPOSED METHODOLOGY

For this research, the radio broadcasts' audio data has been sourced from the News Service Division, AIR. For better understanding and streamlining the process, the paper proposes a methodology divided into 5 stages (excluding the final resultant stage). A diagrammatic representation of the systematic model has been presented in figure 1.

A. DATA PRE-PROCESSING

The first stage, after sourcing the radio broadcasts' audio files, is to sort and arrange the samples in chronological order in which they were aired. The NSD, AIR allows the user to download audio files in the MPEG Audio Layer 3 (.mp3) formats only. However, the various components of the proposed methodology either exclusively work with waveform audio file (.wav) format or have the capacity to process various formats. Thus, it is necessary to convert the sourced.mp3 files to equivalent.wav formats for further processing. The two technical steps involved in this process are as follows:

1) FAST FORWARD MOVING PICTURE EXPERTS GROUP (FFMPEG)

Ffmpeg is an open-source command-line tool consisting of an array of libraries and programs that are widely popular for handling batches of audio, video, and many other media files. Its functionalities consist of the ability to convert audio and video files to various respective formats, compress and decompress media files, extracting audio from video files, creating custom media files and many more. It acts as a platform for libraries like "pydub" to be built on, which acts as the primary interface between the functionalities of ffmpeg and python interface. This well documented tool supports various operating system platforms like Windows, MacOS and Linux, thus, proves to be viable in projects of such scale where the data is very large and parallel computing may be required.

2) PYDUB

Pydub is an open-source library, built over the ffmpeg tool, used for audio manipulation in Python. Pydub offers a high-level API, that abstracts details from ffmpeg, that can be used to play, record, edit, extract, and convert audio files from one file format to another. It can also retrieve duration, bitrate, and sample rate of the audio clip. Pydub can prove to be CPU and memory intensive, especially when used to process large audio files, in works similar to this research.

Both pydub and ffmpeg, are used together for the conversion of the radio audio recordings from.mp3 format to.wav format. Using the method, pydub. AudioSegmant.from_file(), a desired.mp3 file can be loaded into the pydub. Then, using the export() method, which takes two attributes: the file name to be exported as and the file type to be converted into, the audio file can be exported as.wav, which is then further processed in the succeeding stages. Following this process, a total of 138 files (46 days and 3 files per day) were converted to.wav format and arranged in the database chronologically.

B. AUDIO TRANSCRIPTION

After generation of the.wav format database the files are parallelly subjected to two different streams of computation, as shown in figure 1. Conversion of audio to its equivalent text transcription is one of the parallel processes where the processed.wav data is directly fed to the Assembly AI transcription service. A comprehensive and dedicated python and JSON script is used to make API calls between the Assembly AI and the main database for results, via POST and GET methods. The transcribed file is of JSON format with 'text:' containing the transcribed data.

In python, the audio transcription is done directly using the help of 'assemblyai' library and its transcriber() function to generate/initialize an 'assemblyai' transcriber. This transcriber is provided with the input audio file to be transcribed.

AssemblyAI used a structure of deep learning models to perform the speech to text operation on an audio file. The output text of all the audio samples is stored in chronological order and pushed to the next stage to facilitate sentiment analysis of the text.

C. TEXT SENTIMENT ANALYSIS

The text files generated using Assembly AI's transcription are now analyzed using the VADER sentiment analysis tool to extract the sentiment of the words spoken by the radio host. This step is crucial as it will help the model understand if negative words were spoken or sentences that indicate to positive events are spoken, which could be missed or wrongly heard in audio sentiment analysis process later on.

As mentioned earlier, VADER is a highly accessible tool that runs a pre-defined lexicon rule-based model to perform sentiment analysis on words. It can be accessed as an imported library on the Python 3 environment, using the SentimentIntensityAnalyzer method from the vaderSentiment class of the vaderSentiment library. The model matches the word with pre-defined lexicon standards and decides its sentiment score in turn generates polarity_scores of the transcribed text from the analyzer to analyze the final sentiment of the audio text. The result is given in four component scores of positive, negative, neutral, and compound of all the three, where the former three lie between 0-1 and the latter lies between -1 to 1. The first three sentiment scores are tabulated and stored for further calculation of the final sentiment score after combining with audio sentiment analysis scores.

D. AUDIO EMOTION DETECTION

The.wav formatted audio files of the radio broadcasts from stage A are passed through the Vokaturi API tool for emotion detection of the audio files, and eventually sentiment score generation. First, the tool extracts the acoustic features from the speech, and then loads the audio file to its deep neural networks that are trained on two databases that can be accessed in [19] and [20]. At this stage the Vokaturi API will extract the five emotions from the audio files and calculate the strength of each of these emotions from a range of 0 to 1, with precision of up to three decimal places. The output of this stage, the emotion scores, are then passed on to the intermediate result computation stage.

E. OUTPUT FORMULATION

After successful text sentiment analysis, three results are obtained, pertaining to each of the three sentiment classes: positive, negative, and neutral. Similarly, five results are produced, pertaining to the 5 emotions classes: happiness, neutral, sadness, fear, and anger. To equate the sentiment class and emotion classes, [21], [22] have been referred to categorise the 5 emotions into the three sentiment classes.

For the entirety of this paper, assume that in the audio emotion detection the result of happiness, neutral, sadness, fear and anger are a, b, c, d and e respectively, where a, b, c, d and e are positive real numbers. The result of text sentiment analysis for classes positive, neutral, and negative are x, y and z respectively, where x, y and z are positive real numbers. The accuracy of the text sentiment analysis is assumed to be m and that of audio emotion detection to be n, where n and m are positive real numbers. Neutrality to the neutral class is matched, represented by C^n; happiness to the positive class, represented by C^+; and fear, anger and sadness are matched to the negative class, represented by C^- where $C^n = b$ and $C^+ = a$. C^- is thus obtained as the arithmetic mean of c, d, and e, thus $C^- = (c+d+e)/3$.

Calculation of the final three sentiment classes positive, negative, and neutral, represented as F^+, F^- and F^n respectively, is done as

$$F^+ = (mC^+ + nx)/(m+n),$$
$$F^- = (mC^- + nz)/(m+n),$$
$$F^n = (mC^n + ny)/(m+n).$$

Thus, by the computation of the final scores we have combined the sentiment scores of the text sentiment analysis and the converted emotion-to-sentiment scores of the audio recording in the ratio of their respective accuracies, giving the more accurate model higher contribution to the final result.

V. RESULTS AND DISCUSSIONS

Following the proposed methodology, three distinct datasets for a total of 46 days from May 1, 2023, to June 15, 2023, were generated. In these datasets, an analysis of the audio recordings obtained from the main audio bulletin of the NSD: AIR is done and the data for the first 6 days, from May 1, 2023, to May 6, 2023, have been presented in table 5. Each of the three datasets contain 10 columns associated to date, x, y, z, Cn, C+, C−, F+, F− and Fn respectively. The first dataset contains the readings from the morning audio news broadcasts, while the second dataset contains readings from the afternoon (midday) audio news broadcasts and the third dataset contains readings from the evening audio news broadcasts.

A. MORNING AUDIO NEWS RADIO BROADCASTS

Figure 2 illustrates the sentiment analysis of the text that was extracted from the morning audio news broadcasts. A major observation from the generated graph is the similarity in low magnitude of negative and positive emotional trends throughout the duration of the 46 days. This leads us to understand that the written content of the news is majorly aimed towards neutrality with a little extra weight given to positivity. The negative and positive emotions show small fluctuations, however, remain under 20% throughout. The average, maximum and minimum of the positive and negative sentiment scores, rounded to 3 decimal places have been presented in table 6. The average of the neutral sentiment is 23.42447294247967 (approximately 23.42) times and 9.341349491136682 (approximately 9.34) times the score of the negative and positive sentiments respectively.

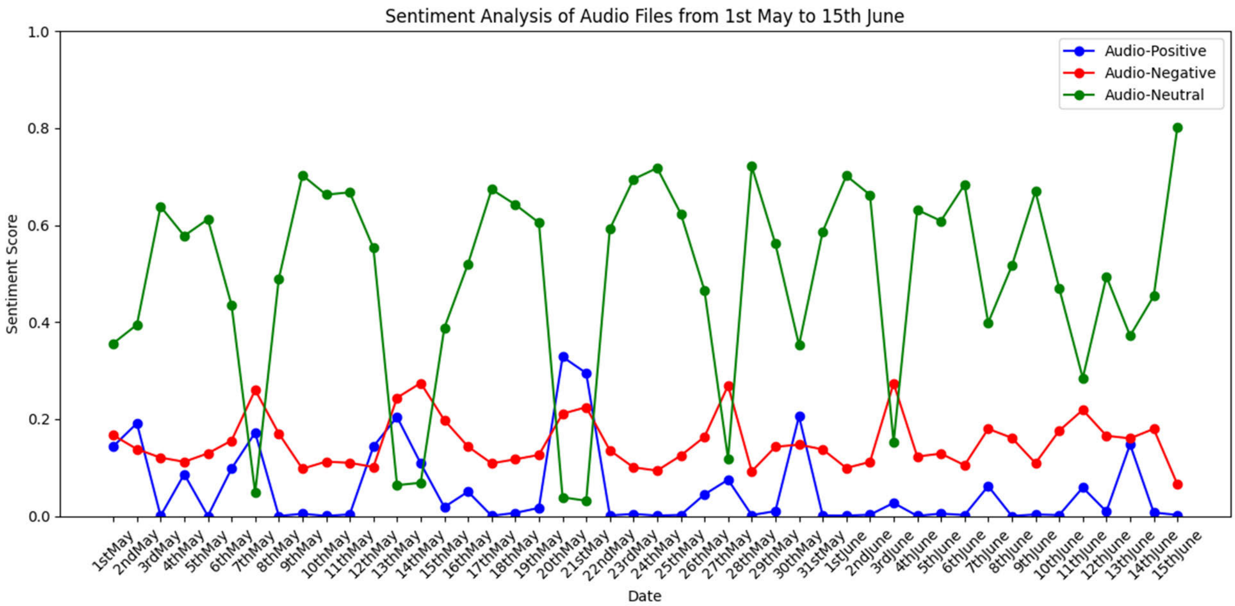

FIGURE 2. Line graph that shows the sentiment scores of positive, negative, and neutral emotions for the audio news samples of morning news broadcasts from May 1, 2023, to June 15, 2023.

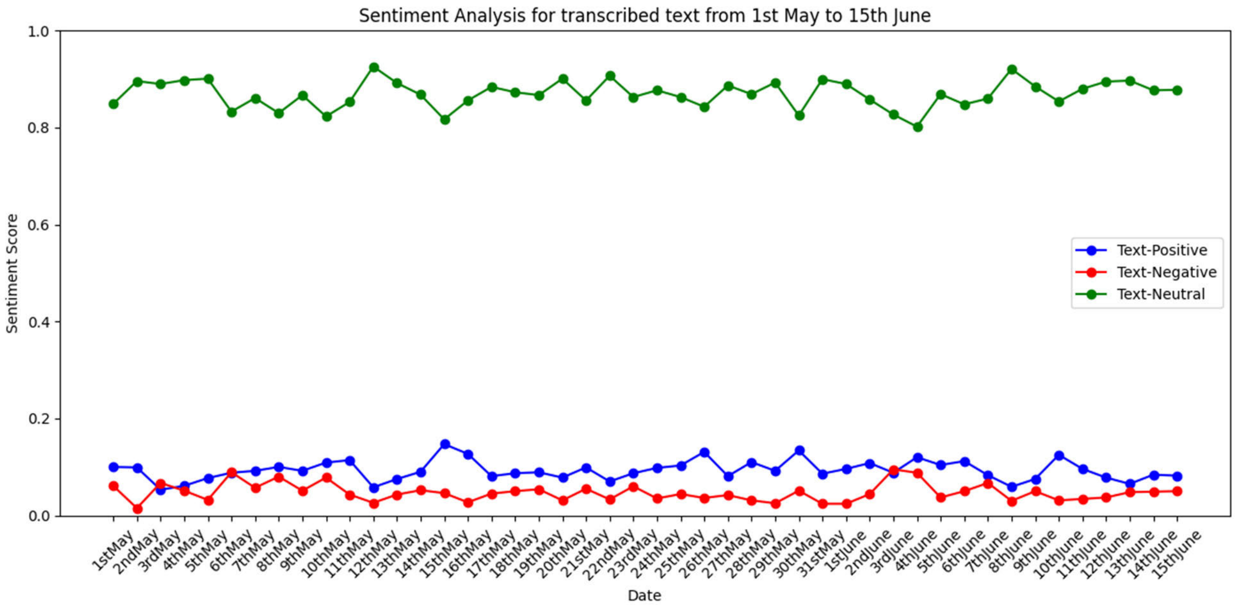

FIGURE 3. Line graph that shows the sentiment scores of positive, negative, and neutral emotions for the transcribed text of morning news broadcasts from May 1, 2023, to June 15, 2023.

Local maxima are observed randomly throughout the trends. The local maximum in the positive emotion scores is seen on Monday, May 15, 2023. On the other hand, the local maximum for negative emotion is observed on June 3, 2023. It is also observed that the negative emotion score crosses the positive emotion score only once on May 3, 2023. After manually analyzing the audio news recording of May 3, 2023, it is observed that the news featured topics such as refugee crisis, terrorism in parts of the nation and fatal encounters, which rightfully justifies the observed sentiment scores. The green line representing neutral emotion in the news remains between the score values 0.802 and 0.923. This is further indication towards the neutrality of the written news broadcasted by All India Radio. The sum of the three scores, for each of the three emotions, at any given date approaches 1.0 (never attaining 1.0). Thus, figure 2 clearly indicates that if the emotional analysis of audio news broadcasts was done solely by sentiment analysis of transcribed text, the results would show neutral emotion by large margins with little influence of positive or negative emotions.

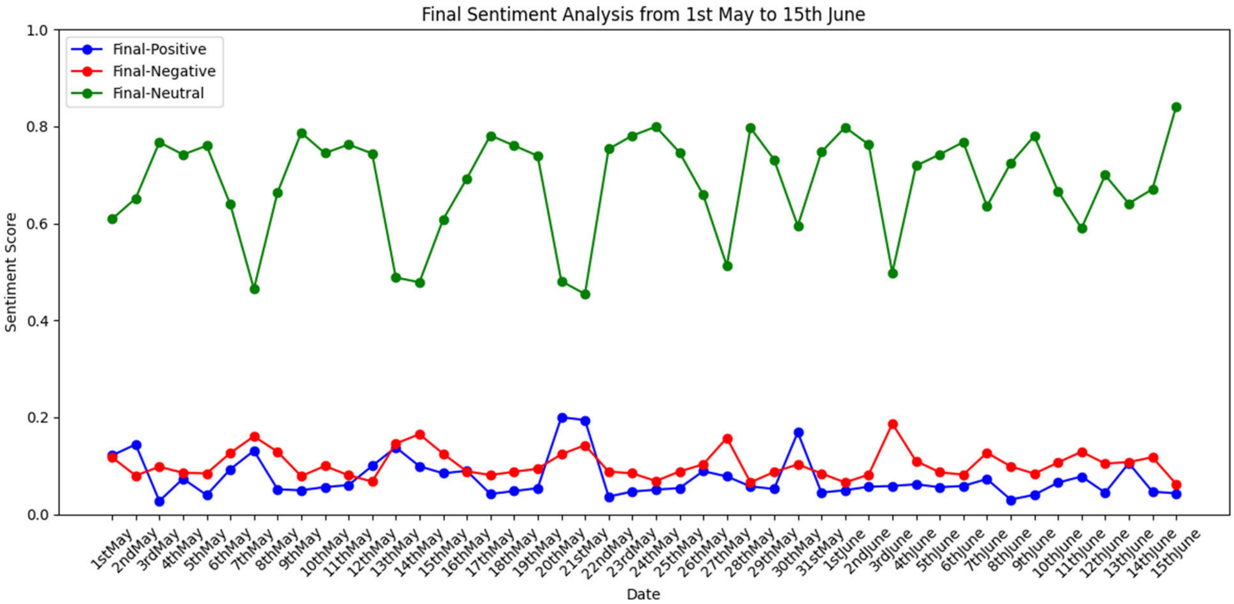

FIGURE 4. Line graph that shows the combined sentiment scores (transcribed text and audio sample) of positive, negative, and neutral emotions for the news broadcast samples of morning news broadcasts from May 1, 2023, to June 15, 2023.

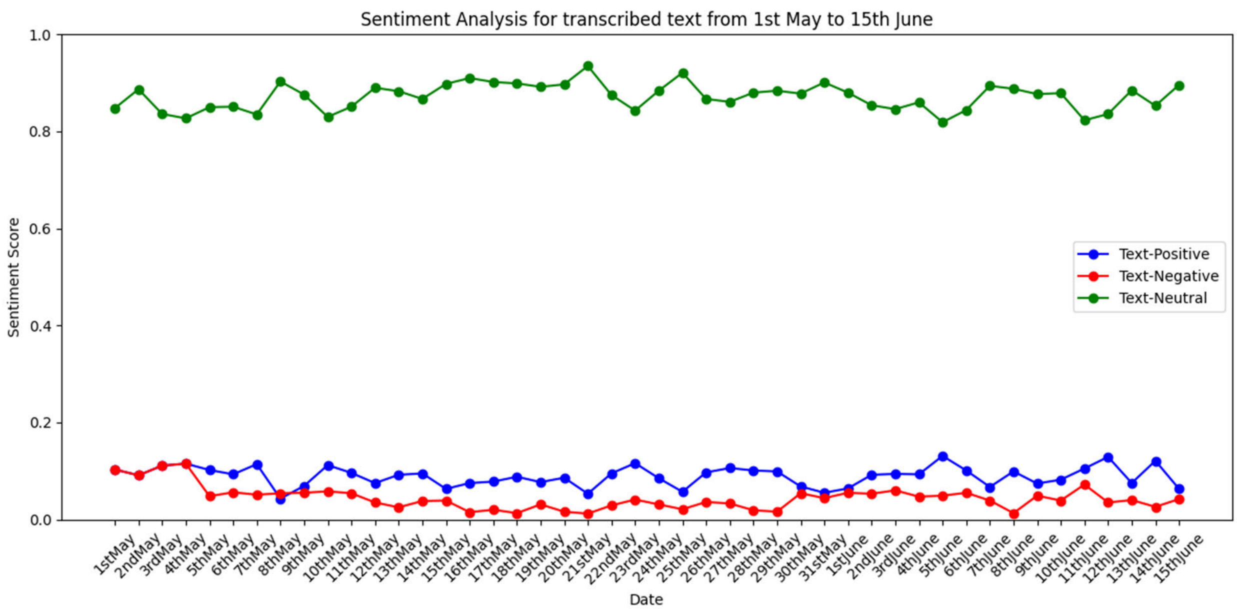

FIGURE 5. Line graph that shows the sentiment scores of positive, negative, and neutral emotions for the transcribed text of midday news broadcasts from May 1, 2023, to June 15, 2023.

On the other hand, when figure 3 is examined, which is a line graph representation of audio sentiment analysis of the news broadcasts, it visualizes very sharp steep fluctuations in the trends of the three emotions. The emotions tend to display higher levels of randomness and lesser similarities or patterns are seen between trends. The negative emotion shows the highest score of 0.274289609 on May 14, 2023, which was almost double that of the positive emotion score on the day, and neutral sentiment score was least of the three. The audio sentiment scores, therefore, clearly show the change in the trend of the sentiments and add greater depth to the analysis and provide a higher degree of score range. Where in figure 2 the neutral sentiment score was maximum in the case of all 46 days, figure 3 shows that the 2 days, May 20, 2023 (Saturday) and May 21, 2023 (Sunday), had maximum sentiment scores for positive emotions. When the news broadcasted on May 20, 2023, is manually analyzed, it's found that the news was based mostly on the inaugural of the global G20 and G7 summits and their hopeful goals and the enhancement of friendship between India and Japan, thus justifying the trend. These events were deemed neutral when only text was analyzed, thus, we see

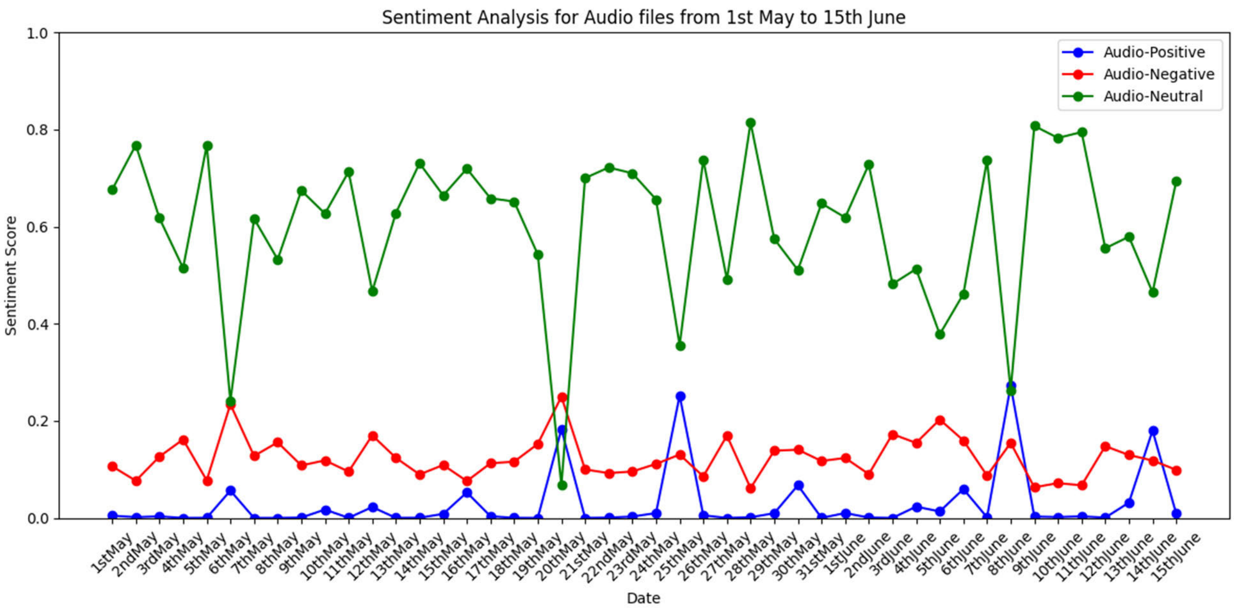

FIGURE 6. Line graph that shows the sentiment scores of positive, negative, and neutral emotions for the transcribed text of midday news broadcasts from May 1, 2023, to June 15, 2023.

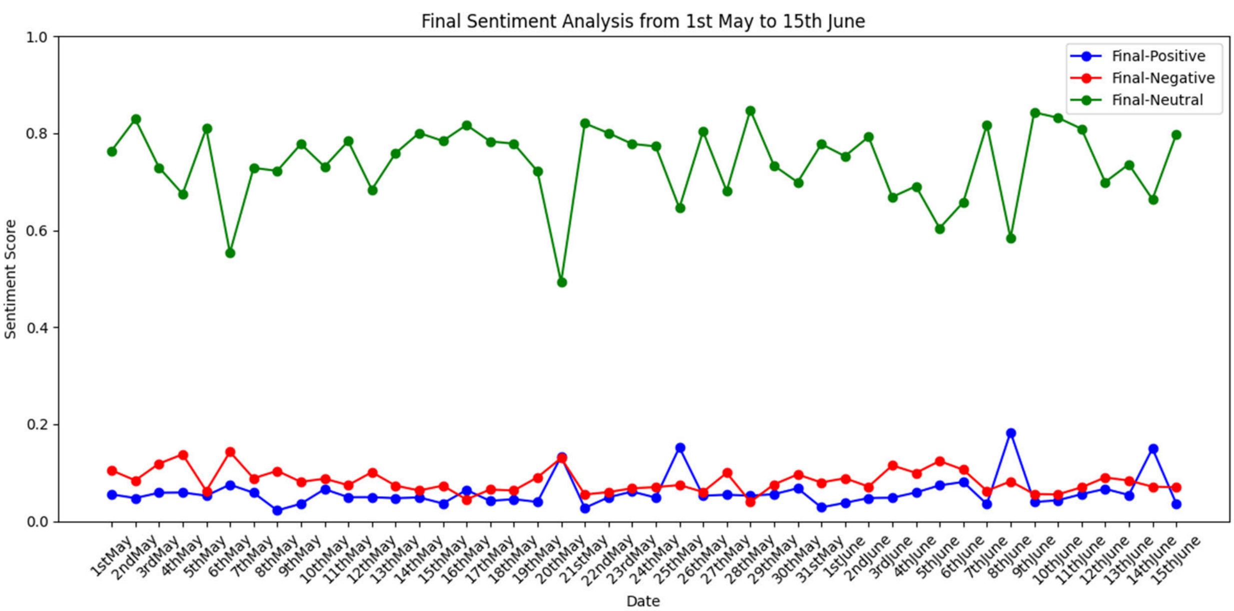

FIGURE 7. Line graph that shows the combined sentiment scores (transcribed text and audio sample) of positive, negative and neutral emotions for the news broadcast samples of midday news broadcasts from May 1, 2023, to June 15, 2023.

the audio sentiment analysis adding dimensions in understanding the sentiments. On the other hand, May 7, 2023 (Sunday), May 13, 2023 (Saturday), May 14, 2023 (Sunday), May 27, 2023 (Saturday) and June 3, 2023 (Saturday) had the maximum sentiment scores for the negative emotions. The aforementioned observations show that both positive and negative emotions attain maximum scores on weekends, that is Saturdays and Sundays, which was missed in only textual analysis.

Figure 4 shows the final sentiment scores of the audio news broadcasts which are computed using the given formula in the output formulation section. Higher fluctuation as compared to figure 2 and lesser fluctuations as compared to figure 3 in all the three emotion classes are observed. The major drawback with figure 2 was the extremely small values generated in the positive and negative emotions, thus, the scores failed to comment on which of the two is influential on a particular day in complementing the neutral sentiment. The major drawback with figure 3 was its very high variance which increased randomness in the results. The final values as depicted by figure 4, which counter both the above drawbacks as it allows us to see higher scores in trends of positive and negative emotions with a moderate variance of the three sentiments. It is observed that the neutral sentiment stays as

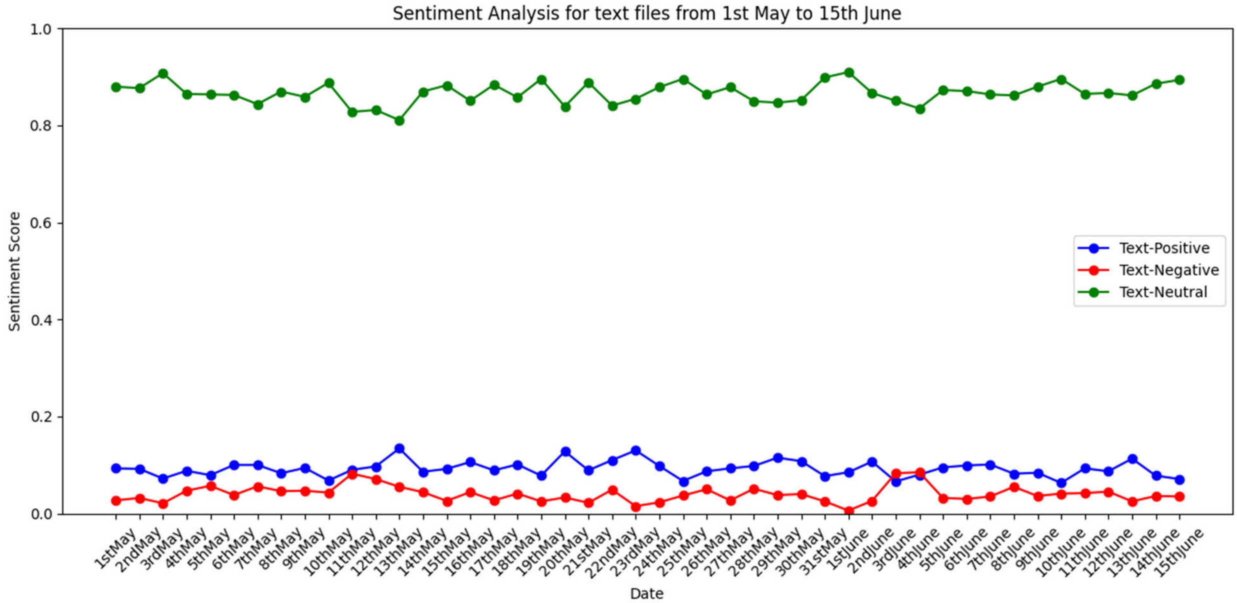

FIGURE 8. Line graph that shows the sentiment scores of positive, negative and neutral emotions for the transcribed text of evening news broadcasts from May 1, 2023, to June 15, 2023.

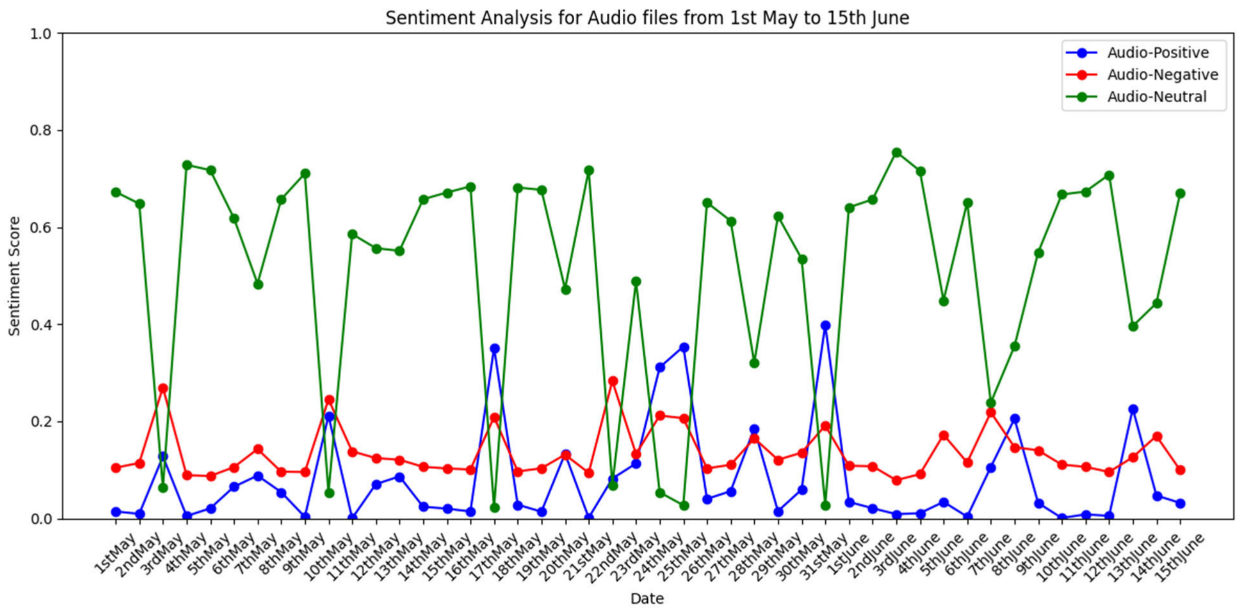

FIGURE 9. Line graph that shows the sentiment scores of positive, negative and neutral emotions for the audio news samples of evening news broadcasts from May 1, 2023, to June 15, 2023.

the maximum scored emotion on all days, which is due to the text sentiment analysis scores having higher influence on the overall score due to the higher accuracy of the text sentiment analysis model. The negative emotion score succeeded the positive emotion score by a small margin on some days like June 3, 2023, while the opposite is observed in the case of a few other days like May 30, 2023. From figure 4 it can be clearly seen that on May 3, 2023, the negative emotion score is significantly more than the positive emotion score, which could not be analyzed in such detail by figure 2 alone. It is observed that trends in figure 4 as a combination from figure 2 and figure 3. In figure 2 it is seen that the positive emotion score is greater in the majority section as compared to negative emotion score and the exact opposite trend is seen in figure 3 where the negative score is greater in most of the region, thus, these two analyses thus produce contrasting results. Figure 4 presents a solution to this conflict as by combining the results of the previous two analyses a trend is generated where positive and negative sentiment score lines cross each other multiple times alternating the higher scored emotion class between positive and negative (neutral class remains highest overall). Hence, it's inferred from figure 4

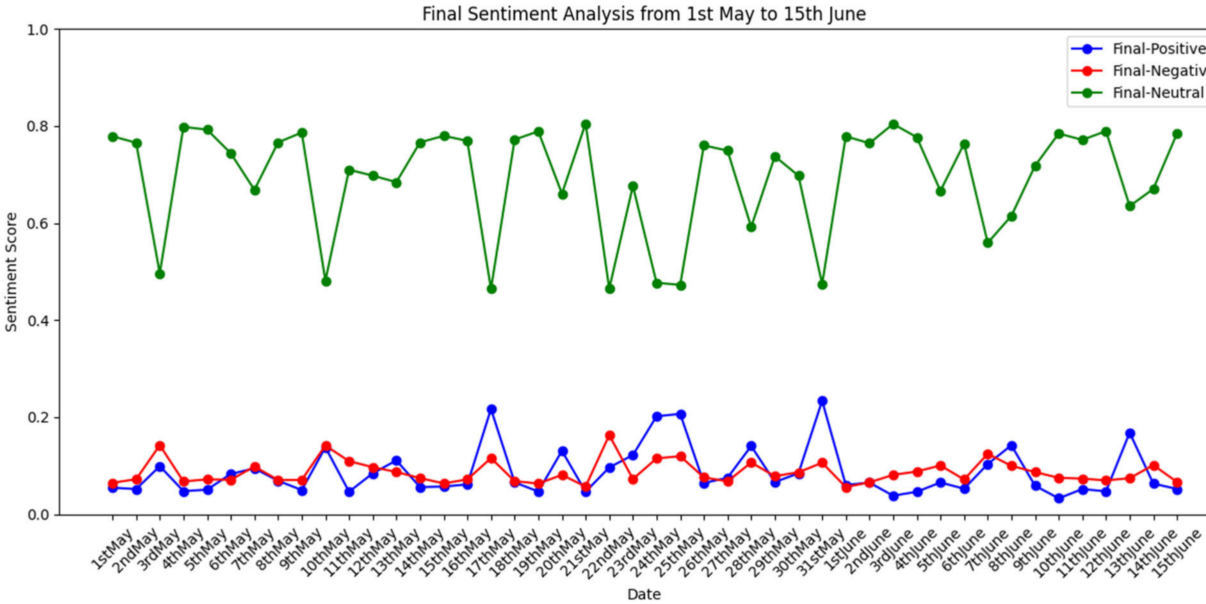

FIGURE 10. **Line graph that shows the combined sentiment scores (transcribed text and audio sample) of positive, negative and neutral emotions for the news broadcast samples of evening news broadcasts from May 1, 2023, to June 15, 2023.**

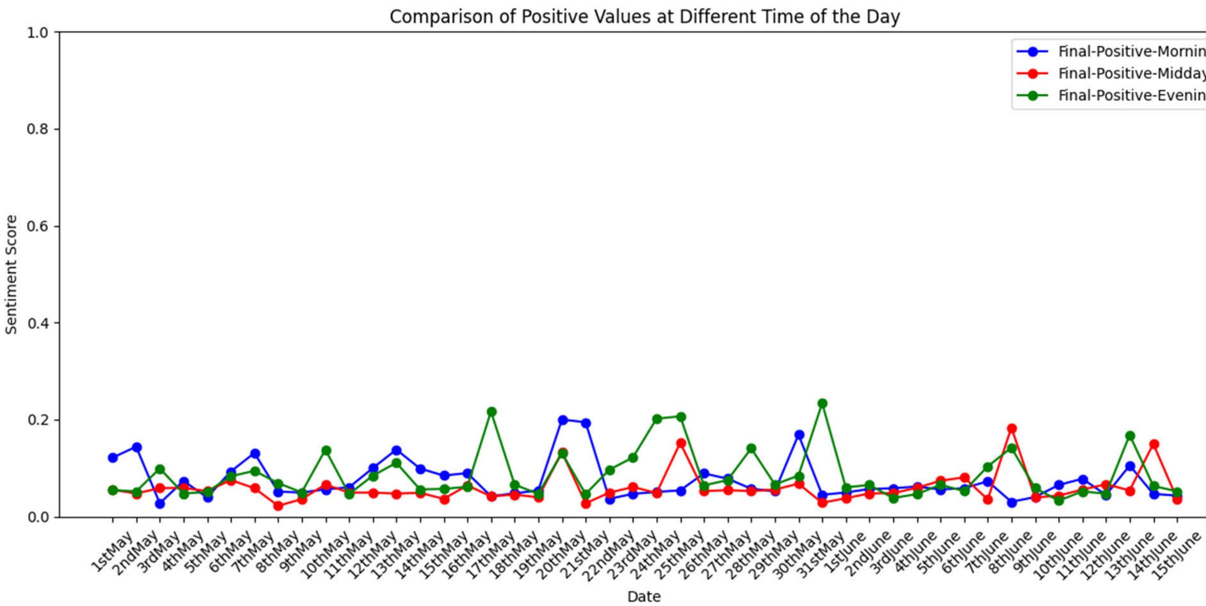

FIGURE 11. **Line plot that compares the final positive sentiment scores of the morning, midday, and evening news audio broadcasts from May 1, 2023, to June 15, 2023.**

that though neutrality is the base emotion on all days, positive and negative emotions are also present in varying amounts which give depth to the analysis.

B. MIDDAY AUDIO NEWS BROADCASTS

After the morning news broadcast analysis, the paper analyses the midday news broadcasts by All India Radio from May 1, 2023, to June 15, 2023. First let's discuss the scores computed by the transcribed text of the midday news which is plotted in figure 5. In figure 5, the positive and negative sentiment scores overlap for the first 4 days of the sampled dates and begin separating from then onwards. Similar to the observations in figure 2, the positive sentiment score is seen to lead by small margins in most of the sections compared to the negative sentiment scores. The negative score exceeded the positive scores only on May 8, 2023, the recording of which when manually examined for justification showed that this was caused due to news such as MIG fighter jet crash, boat tragedy and court orders related to gangster's release from jail. Furthermore, both negative and positive sentiment scores remained under 15% throughout while the neutral sentiment score remained above 80% with maximum score

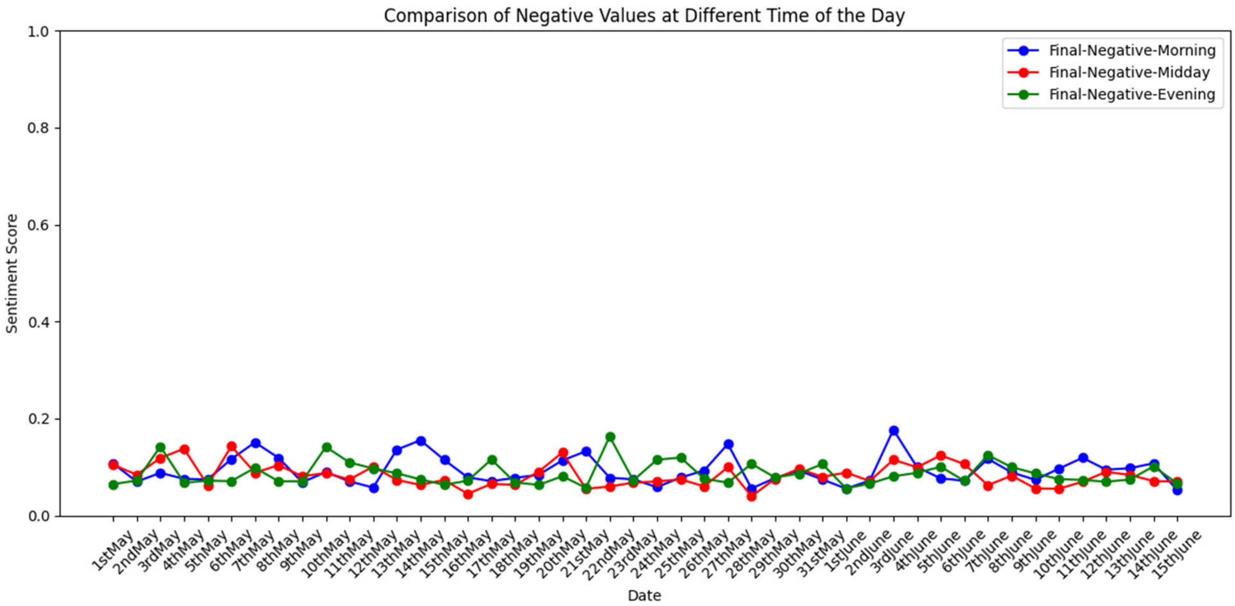

FIGURE 12. Line plot that compares the final negative sentiment scores of the morning, midday, and evening news audio broadcasts from May 1, 2023, to June 15, 2023.

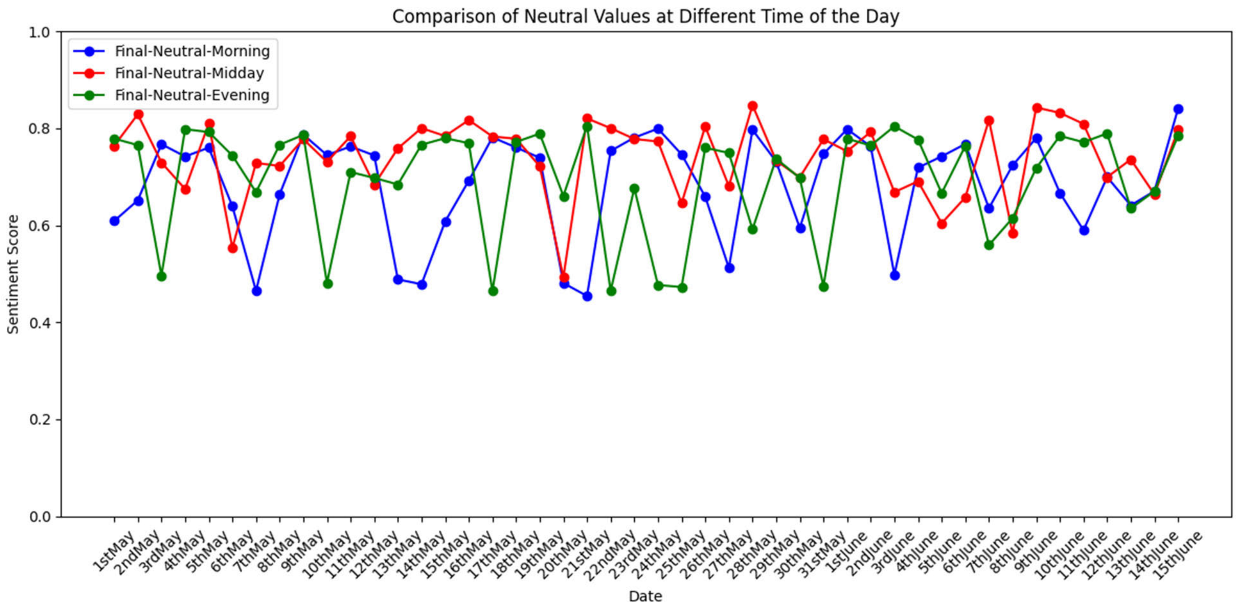

FIGURE 13. Line plot that compares the final neutral sentiment scores of the morning, midday, and evening news audio broadcasts from May 1, 2023, to June 15, 2023.

being 0.935 on May 21, 2023. The sum of the three sentiment scores on a particular day approach (but never attaining) 1.00. It is inferred that the transcribed news text displayed neutrality with large margins having only a small decrease in values when one of the other two sentiment scores increased.

Figure 6, contrary to figure 5, shows large fluctuations in the line plot of the three sentiments. Data with large variance and deviation is achieved using the audio emotion detection techniques. The maximum neutrality score dropped from 0.935 to 0.814387298 and is now achieved on May 28, 2023, instead of May 21, 2023. The positive sentiment score crossed the neutral score line on June 8, 2023, with a difference of 0.009224445. The negative sentiment score crossed the neutral score line on May 20, 2023, with a difference of 0.182766618. The neutral sentiment score is seen to be the maximum throughput with positive and negative sentiment score alternately being second maximum, however this trend fails on May 20, 2023, when negative score crossed both positive and neutral sentiment scores, and neutral sentiment was observed to be even less than positive sentiment score. These observations point out stark differences in the text sentiment analysis and audio sentiment analysis and substantiate

TABLE 5. The APIs' machine learning models' results for all three segments for the first six days of the observed timeframe.

MORNING AUDIO NEWS RADIO BROADCASTS									
	Audio Sentiment Analysis Scores			Text Sentiment Analysis Scores			Final Sentiment Analysis Scores		
Date	POSITIVE	NEGATIVE	NEUTRAL	POSITIVE	NEGATIVE	NEUTRAL	POSITIVE	NEGATIVE	NEUTRAL
1st May 2023	0.143	0.166	0.356	0.001	0.051	0.849	0.121	0.107	0.608
2nd May 2023	0.191	0.138	0.395	0.099	0.005	0.896	0.144	0.070	0.652
3rd May 2023	0.001	0.120	0.639	0.053	0.057	0.890	0.027	0.088	0.767
4th May 2023	0.086	0.112	0.578	0.061	0.041	0.898	0.073	0.076	0.742
5th May 2023	0.001	0.129	0.612	0.077	0.022	0.901	0.040	0.074	0.76
6th May 2023	0.098	0.156	0.436	0.088	0.079	0.833	0.093	0.116	0.639
MIDDAY AUDIO NEWS RADIO BROADCASTS									
	Audio Sentiment Analysis Scores			Text Sentiment Analysis Scores			Final Sentiment Analysis Scores		
Date	POSITIVE	NEGATIVE	NEUTRAL	POSITIVE	NEGATIVE	NEUTRAL	POSITIVE	NEGATIVE	NEUTRAL
1st May 2023	0.005	0.106	0.676	0.103	0.103	0.848	0.055	0.105	0.764
2nd May 2023	0.002	0.077	0.768	0.091	0.091	0.887	0.047	0.084	0.829
3rd May 2023	0.003	0.126	0.618	0.111	0.111	0.836	0.059	0.118	0.730
4th May 2023	0.001	0.162	0.515	0.115	0.115	0.827	0.059	0.138	0.675
5th May 2023	0.001	0.077	0.768	0.102	0.048	0.850	0.052	0.062	0.810
6th May 2023	0.056	0.234	0.241	0.093	0.056	0.851	0.075	0.143	0.554
EVENING AUDIO NEWS RADIO BROADCASTS									
	Audio Sentiment Analysis Scores			Text Sentiment Analysis Scores			Final Sentiment Analysis Scores		
Date	POSITIVE	NEGATIVE	NEUTRAL	POSITIVE	NEGATIVE	NEUTRAL	POSITIVE	NEGATIVE	NEUTRAL
1st May 2023	0.014	0.104	0.673	0.093	0.027	0.880	0.055	0.065	0.779
2nd May 2023	0.009	0.114	0.648	0.092	0.032	0.877	0.052	0.072	0.766
3rd May 2023	0.128	0.270	0.063	0.072	0.021	0.908	0.099	0.142	0.496
4th May 2023	0.005	0.089	0.728	0.088	0.047	0.865	0.047	0.067	0.798
5th May 2023	0.021	0.087	0.717	0.079	0.057	0.864	0.051	0.072	0.793
6th May 2023	0.065	0.105	0.619	0.001	0.038	0.863	0.083	0.071	0.744

this research's claim regarding the importance of using both techniques to get best possible and accurate sentiment results. One sharp observation that was made from the comparison of figure 5 and figure 6 was the drop in positive sentiment score in the latter. Where the average positive sentiment score was 0.089130435 for the transcribed text, it reduced approximately 4 times to merely 0.028590504 when the audio analysis was examined. Now, the neutral sentiment was seen maximum for most of the days, followed by negative sentiment and lastly positive sentiment.

Figure 7 shows the trends when the results of figure 5 and figure 6 were added in the proportion of their accuracies. Thus, we combat the drawbacks of extreme valuation and extreme variance observed in the figures 5 and 6 respectively. The neutral sentiment though stayed the strongest of the three shows more flexibility in values and shows influence of positive and negative sentiments. If only the transcribed text was used for sentiment analysis of news broadcasts, then it would fail to identify the leading sentiment on dates May 1, 2023, to May 4, 2023, as both values overlap. It would simply have considered the negativity and positivity in the news to be the same. However, when the results obtained are combined from the audio sentiment analysis with the results of figure 5, it's noticed that equality was not truly observed, and negativity was rather the dominant sentiment out of the two. This was confirmed by manual analysis of the news audio clip of the four days where topics such as explosion in Ukraine and frivolous politics in Indian state

TABLE 6. The APIs' machine learning models' results for all three segments, morning, midday and evening, for the first six days of the observed timeframe.

Category	Value Type	Audio Sentiment Analysis Scores			Text Sentiment Analysis Scores			Final Sentiment Analysis Scores		
		Positive	Negative	Neutral	Positive	Negative	Neutral	Positive	Negative	Neutral
Morning Radio Broadcasts	Minimum	0.001	0.065	0.031	0.053	0.005	0.802	0.027	0.052	0.454
	Maximum	0.328	0.274	0.802	0.147	0.085	0.923	0.2	0.177	0.841
	Average	0.06	0.153	0.486	0.093	0.037	0.87	0.076	0.094	0.683
Midday Radio Broadcasts	Minimum	0.001	0.061	0.067	0.043	0.012	0.819	0.022	0.04	0.493
	Maximum	0.273	0.25	0.814	0.131	0.115	0.935	0.184	0.143	0.848
	Average	0.029	0.123	0.601	0.089	0.044	0.872	0.06	0.083	0.74
Evening Radio Broadcasts	Minimum	0.001	0.079	0.022	0.063	0.006	0.811	0.033	0.056	0.464
	Maximum	0.399	0.283	0.755	0.134	0.085	0.91	0.234	0.163	0.805
	Average	0.081	0.135	0.513	0.092	0.04	0.867	0.087	0.087	0.695

of Karnataka, which shows that the sentiment was towards a negativity which was not clear from figure 5. Thus, though results were obtained from only transcribed text analysis, their shortcomings were reduced using the audio sentiment analysis results.

C. EVENING AUDIO NEWS BROADCASTS

Lastly, the paper analyses the evening news broadcasts of the NSD: AIR for the same duration as the morning and midday radio broadcasts. Figure 8 represents the sentiment score of the transcribed news obtained from the evening news broadcasts. It is observed that neutral sentiment scores lead the other two by great margins and ranged from 81% to 91% throughout. Minor variations in the neutral sentiment score were observed with the average as presented in table 6. The positive sentiment score is seen to lead the negative sentiment score by small margins daily, with exceptions on June 3, 2023, and June 4, 2023, when the negative score was greater. The negative sentiment score had an average almost 2.29 times less than the average of the positive sentiment score. As observed previously as well, overall neutrality in news's written text is maintained with positive sentiments being second highest almost throughout, with aforementioned exceptions.

Figure 9, depicting the sentiment analysis of the audio, on the contrary shows a high level of fluctuations. Very high variance was observed in data values and sharp deviations. Though neutrality is still seen to be maximum in major sections, negative sentiment score leads the positive sentiment score by margin (8 days being exceptions), opposite to what was observed in figure 8. The negative score had an average of 0.135254226, about 3.36% greater than observed in transcribed text. Taking a specific example, on May 10, 2023, the transcribed text showed a negative sentiment score of 0.043, which was less than the positive and neutral sentiment score on the day. However, when the negative sentiment score from figure 9, for the same day, is observed, it comes out to be 0.245159279 and is higher than the positive and neutral sentiment scores. The reason for this is found on manual inspection of the news audio broadcasts which reveals that though the written news was towards neutrality, the interviews of various personalities in regional languages such as Hindi, show hints of negative sentiment in NVV such as sighs, when carefully examined. The positive sentiment is seen to be the maximum of the three sentiments on May 17, 2023, May 24, 2023, May 25, 2023, and May 31, 2023, which are all mid-weekdays like Wednesday and Thursday. The average, minimum, and maximum values for the positive, negative, and neutral sentiment scores can be studied from table 6.

The combined results of data from figures 8 and 9 are displayed in figure 10. Neutrality is maximum from all 46 days. A more homogeneous pattern is observed with respect to positive and negative sentiment lines. The lines seem to cross each other several times with occasional sharp rises in sentiment scores of the positive sentiment. The figure clearly shows the need for the use of audio sentiment analysis. Using only transcribed text, a model would classify, for example, May 22, 2023, as neutral with positive sentiment effects. However, after adding the data obtained from audio sentiment analysis, May 22, 2023, was in fact observed to be a day with negative sentiments. Figure 10 provides a solution to the high variance observed in figure 9 and supports the establishment of general trend in sentiment values.

D. COMPARING RESULTS

Table 6 tabulates the minimum, maximum and average values of the output of the text sentiment analysis, audio sentiment analysis and final sentiment analysis for the three sentiments, positive, negative, and neutral, for morning, evening, and midday news broadcasts. After the analysis of the sentiment scores for the morning, midday, and evening news audio broadcasts separately, now, an analysis of the daily trends after combining the data of the three aforementioned daily broadcasts is done.

From figure 11 depicting the final scores of positive sentiments, it's clearly observed that morning and evening news broadcasts were more positive as compared to the midday broadcasts. Similarly, when figure 12 is observed, it's seen that negative sentiment was also strongest in the morning and evening news broadcasts. Figure 13 shows that midday news on the other hand is seen to be most focused towards neutrality as compared to morning and evening news. Thus, this study not only presents the overall sentiment analysis of the radio news for a day but also presents a result on the timings when the news broadcasts are mostly neutral or positive/negative.

VI. CONCLUSION AND FUTURE WORKS

Text sentiment analysis though produces high accuracy results, there are a few drawbacks to the methodology. Spoken audio not only contains words but also gestures and signals for pain, grief, happiness etc. Text sentiment analysis is a good option to analyze the transcribed text; however, when it comes to analyzing the NVV, text analysis is not enough. Technologies like audio sentiment analysis and emotion detection are the way to analyze NVV with great efficiency. The sound like a sigh might be captured during text transcription but their emotions remain vague. Using audio sentiment analysis and studying the audio pattern of the sign, one can determine if it's a sigh of relief or pain. However, the high variance and deviation observed using the audio sentiment analysis alone affects the accuracy. This is in turn supported by the text sentiment analysis providing much better results that can be verified by manual analysis of the audio files.

This research presented analysis of news audio broadcasts based on audio and text sentiment analysis and the combination of the two results helped us extract more information out of the audio and any one of the methods could do alone. This comprehensive study on the recordings by All India Radio, can and must be extended to local news broadcasting stations and international radio stations for a bigger, more detailed picture on the sentiment of the news broadcasted. Thus, in turn, helping evaluate the overall condition of any state or territory.

In the future, it would be an imperative aim to utilize this novel research to find solutions to real world problems and do our part in giving back to society. One probable use of the research is in disaster management. During disasters radio signals are used for communication in remote locations. Using the emotion detection approach that's presented in the paper, development of advanced distress signal analysis tools may be generated. When a person in need signals the resume team using a radio signal, he or she may not necessarily be in a state to vocalize his condition in words, but using the model to understand the NVV, one can identify the situation as a "SOS" call. Thus, saving lives in many countries where radio broadcasts are still the main source of communication. A major limitation towards development of a standalone system that runs these analysis models is the presence of high-quality datasets that can be used to train the models with the most relevant data and increase accuracy of the models. Datasets on NVV are most needed in accurate analysis of disaster management situations.

Another vital plan is to address the areas of the human demographic that were currently outside the scope of this research and propose to extend this model to cater to all local languages. Also, an aim to understand the needs of the differently- abled citizens, sexual minorities, and for the persecuted in a country and provide solutions is a crucial outcome using our designed model which could be extended just from news and a general product to make it personal to whomever using it.

REFERENCES

[1] M. R. Yaakub, M. I. A. Latiffi, and L. S. Zaabar, "A review on sentiment analysis techniques and applications," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 551, no. 1, 2019, Art. no. 012070.

[2] Statista. (Jan. 19, 2023). *Global Internet User Growth 2018–2023*. [Online]. Available: https://www.statista.com/statistics/1190263/internet-users-worldwide/#:~:text=The%20number%20of%20internet%20users,to%202023%20is%20six%20percent

[3] Statista. *Traditional Radio—Worldwide | Statista Market Forecast*. [Online]. Available: https://www.statista.com/outlook/amo/media/music-radio-podcasts/traditional-radio/worldwide

[4] Y. Mejova, "Sentiment analysis: An overview," Dept. Comput. Sci., Univ. of Iowa, Iowa City, IA, USA, 2009.

[5] A. Apturkar, A. I. Iliev, A. Anand, A. Oli, S. R. Siddenki, and V. R. Meka, "Sentiment analysis of speech with application to various languages," *Digit. Presentation Preservation Cultural Sci. Heritage*, vol. 10, pp. 103–118, Sep. 2020.

[6] J. Abeßer, "A review of deep learning based methods for acoustic scene classification," *Appl. Sci.*, vol. 10, no. 6, p. 2020, Mar. 2020.

[7] D. Bamunuarachchi et al., "Review paper on audio signal classification techniques," Tech. Rep., 2012.

[8] J. M. Garcia-Garcia, V. M. R. Penichet, and M. D. Lozano, "Emotion detection: A technology review," in *Proc. 18th Int. Conf.*, 2017, p. 8, doi: 10.1145/3123818.3123852.

[9] S. Casale, A. Russo, G. Scebba, and S. Serrano, "Speech emotion classification using machine learning algorithms," in *Proc. IEEE Int. Conf. Semantic Comput.*, Aug. 2008, pp. 158–165.

[10] G. Bohota and V. Këpuska, "Comparing speech recognition systems (Microsoft API, Google API and CMU Sphinx)," *Int. J. Eng. Res. Appl.*, vol. 7, no. 3, pp. 20–24, Mar. 2017, doi: 10.9790/9622-0703022024.

[11] J. Y. Kim, C. Liu, R. A. Calvo, K. McCabe, S. C. R. Taylor, B. W. Schuller, and K. Wu, "A comparison of online automatic speech recognition systems and the nonverbal responses to unintelligible speech," 2019, *arXiv:1904.12403*.

[12] B. Abu-Salih, M. Alhabashneh, D. Zhu, A. Awajan, Y. Alshamaileh, B. Al-Shboul, and M. Alshraideh, "Emotion detection of social data: APIs comparative study," *Heliyon*, vol. 9, no. 5, p. e15926, 2023.

[13] Vokaturi. (2016). *Vokaturi*. Accessed: May 30, 2023. [Online]. Available: https://vokaturi.com/

[14] "Classification of emotions," in *The Emotion Machine*, Jun. 2021. [Online]. Available: https://www.theemotionmachine.com/classification-of-emotions

[15] *News on AIR—News Services Division, All India Radio News*. [Online]. Available: https://newsonair.gov.in/

[16] *Internet Archive: Digital Library of Free and Borrowable Books, Movies, Music & Wayback Machine*. [Online]. Available: https://archive.org/

[17] *WBAI Radio 99.5FM NYC*. [Online]. Available: https://wbai.org

[18] CMUSphinx, Carnegie Mellon University. (2016). *CMUSphinx Tutorial for Developers*. [Online]. Available: http://www.speech.cs.cmu.edu/sphinx/

[19] *Surrey Audio-Visual Expressed Emotion (SAVEE) Database*. [Online]. Available: http://www.kahlan.eps.surrey.ac.uk/savee

[20] *Emo-DB*. [Online]. Available: http://www.emodb.bilderbar.info/index-1280.html

[21] B. L. Fredrickson and M. A. Cohn, "Positive emotions," Tech. Rep., 2008.

[22] R. C. Solomon and L. D. Stone, "On 'positive' and 'negative' emotions," *J. Theory Social Behav.*, 2002.

[23] H. Nguyen, A. Veluchamy, M. Diop, and R. Iqbal, "Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches," *SMU Data Sci. Rev.*, vol. 1, no. 4, 2018, Art. no. 7.

[24] U. Mehta and D. Verma, "Sentiment analysis of Facebook using textblob and vader," *J. Innov. Eng. Res.*, vol. 4, no. 1, pp. 10–14, 2021.

[25] C. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. Assoc. Adv. Artif. Intell. Conf.*

[26] *Improved WER—Speech-to-Text Accuracy vs. Google*, AWS, Seattle, WA, USA, May 2020.

[27] [Online]. Available: https://www.assemblyai.com/blog/speech-to-text-wer-google-aws-may-2020/

[28] *AWS Transcribe Main Site*. [Online]. Available: https://aws.amazon.com/transcribe/

[29] [Online]. Available: https://vadersentiment.readthedocs.io/en/latest/pages/features_and_updates.html

[30] G. G. Devarajan, S. M. Nagarajan, S. I. Amanullah, S. A. S. A. Mary, and A. K. Bashir, "AI-assisted deep NLP-based approach for prediction of fake news from social media users," *IEEE Trans. Computat. Social Syst.*, early access, Mar. 29, 2023, doi: 10.1109/TCSS.2023.3259480.

[31] N. S. Murugan and G. U. Devi, "Detecting streaming of Twitter spam using hybrid method," *Wireless Pers. Commun.*, vol. 103, no. 2, pp. 1353–1374, Nov. 2018.

[32] G. M. Linda, N. V. S. S. R. Lakshmi, N. S. Murugan, R. P. Mahapatra, V. Muthukumaran, and M. Sivaram, "Intelligent recognition system for viewpoint variations on gait and speech using CNN-CapsNet," *Int. J. Intell. Comput. Cybern.*, vol. 15, no. 3, pp. 363–382, Jul. 2022.

[33] A. Al-Rawi, "Understanding the social media audiences of radio stations," *J. Radio Audio Media*, vol. 23, no. 1, pp. 50–67, Jan. 2016.

[34] J. H. Joloudari, S. Hussain, M. A. Nematollahi, R. Bagheri, F. Fazl, R. Alizadehsani, R. Lashgari, and A. Talukder, "BERT-deep CNN: State of the art for sentiment analysis of COVID-19 tweets," *Social Netw. Anal. Mining*, vol. 13, no. 1, p. 99, Jul. 2023.

[35] G. G. Devarajan, S. M. Nagarajan, A. Daniel, T. Vignesh, and R. Kaluri, "Consumer product recommendation system using adapted PSO with federated learning method," *IEEE Trans. Consum. Electron.*, early access, Sep. 29, 2023, doi: 10.1109/TCE.2023.3319374.

NAMAN DHARIWAL (Student Member, IEEE) is currently pursuing the B.Tech. degree in computer science engineering with the Vellore Institute of Technology, Vellore. He is a Data Science Engineer. He is passionate about helping build a just and sustainable world through knowledge and skill sets. His research interests include artificial intelligence, machine learning, natural language processing, deep learning, and image processing.

SRI CHANDER AKUNURI is currently pursuing the bachelor's degree in computer science with the Vellore Institute of Technology, Vellore. He is a research enthusiast hailing from Vijayawada. He has an undying interest in acquiring knowledge from various domains and ascertaining novel ways of applying it to fields of passion, such as natural language processing, artificial intelligence, data science, and computer vision. His inspirations lie in the profound expressions of philosophy and workings of songs in Carnatic music. He also enjoys Indic and English literature. His research interests include medicine, environment and climate change, biosciences, wider social applicability, and finance.

SHIVAMA was born in Bihar, India. He is currently pursuing the degree in computer science with the Vellore Institute of Technology (VIT), Vellore. This academic journey is a part of the larger vision to make a substantial impact in the technological industry. After entering VIT, he developed a keen interest in machine learning (ML). Since then, he has engaged in research projects in this area. These pursuits act as crucial stepping stones toward fulfilling the high aspirations for the future. With a relentless drive and unwavering focus, he looks forward to facing the challenges that lie ahead in the still-unfolding story.

K. SHARMILA BANU has been with the School of Computer Science and Engineering, Vellore Institute of Technology, since 2009. She has more than 14 years of teaching experience for undergraduate and graduate students and three Ph.D. scholars. She has published more than 15 research articles. She is working on a consultancy project, a seed fund grantee, and a MeitY Fellowship grantee. Her research interests include rough set theory, neighborhood rough set theory, natural language processing, and deep learning.