Optimized Sentiment Analysis in Tagalog Speech Using PCA and BRNN on Prosodic Suprasegmental and MFCC Features

[1] Ailen B. Garcia
Technological Institute of the Philippines
Aurora Blvd., Cubao, Quezon City
Philippines
qabgarcia01@tip.edu.ph

[2] Bobby D. Gerardo, D.Eng
Northern Iloilo State University
Estancia, Iloilo, Philippines
bgerardo@nisu.edu.ph

[3] Ruji P. Medina, Ph.D
Technological Institute of the Philippines
Aurora Blvd., Cubao, Quezon City
Philippines
ruji.medina@tip.edu.ph

Abstract*—Sentiment analysis in Tagalog Speech is a field that has seen growing interest and development but still faces several challenges and opportunities for enhancement. This study focuses on the development of a Tagalog speech sentiment analysis model utilizing the hybrid prosodic suprasegmental features and Mel Frequency Cepstral Coefficients (MFCCs) optimized by Principal Component Analysis (PCA) and Bidirectional Recurrent Neural Network (BRNN) architecture. Prosodic suprasegmental and MFCC features which capture the spectral content and emotional tone of speech are effectively integrated. Another significant improvement is the use of PCA for dimensionality reduction, which enhanced the model's performance by reducing overfitting. The model achieved a higher accuracy rate of 90.91% compared to the 82% of existing development after being trained to classify sentiments into positive, negative, and neutral categories. The findings of the study show the potential of sophisticated machine learning methods for speech sentiment analysis, especially for Tagalog language where research is relatively limited.

***Keywords*— Tagalog speech sentiment, PCA, Bidirectional RNN, MFCC, prosodic suprasegmental**

I. INTRODUCTION

Speech sentiment is progressing and emerging in different fields of research. It is the computational examination of individuals' views [1], feelings, evaluations, and attitudes toward something[2]. Speech sentiment analysis has garnered a lot of attention lately because of its many uses in a variety of fields, including market research, social media monitoring, student and customer feedback analysis, and more, to find out what people think and desire [3]. It is the process of categorizing the emotional tones that speakers convey through their spoken words [4] important details regarding the views, convictions, and feelings of people speaking can be gleaned from this procedure. Speech or audio sentiment analysis is becoming more and more popular as speech data becomes more readily available and people communicate with one another via speech.

Recently, there are several approaches that have been developed in speech sentiment analysis [1] depending on the quality of data to fulfill the sentiment analysis tasks [5]. In natural language processing tasks, including sentiment analysis, deep learning models have demonstrated remarkable success [6]. The emergence of Deep Learning (DL) has increased the efficiency possibilities for researchers to develop better-performing Speech Emotion Recognition (SER) [7]. Although much existing research showed that a hybrid system performs better than traditional single classifiers used in speech emotion recognition, there are some limitations in each of them.

One of the deep learning used in sentiment analysis is the Bidirectional Recurrent Neural Network (BRNN) [8]. The Bi-directional Recurrent Neural Network (BRNN) is utilized to enhance sentiment analysis efficacy in regional tongues. One of the benefits of the BRNN method [9] is that it represents phrases with high and low resources in a shared space and uses the similarity measure to examine sentiment.

In languages like Tagalog, where there is a lack of annotated voice data and variance in intonation patterns throughout the many regions of the Philippines, creating robust sentiment analysis algorithms is extraordinarily difficult. Filipino or Tagalog is the national language of the Philippines. It belongs to the Malayo-Polynesian group of Austronesian languages, and it is a member of the Central Philippine subgroup of Philippine languages [10]. It is one of the most spoken language forms in the Philippines, with 45 million speakers[11] wherein it can be used for communication among people orally or virtually.

However, analyzing sentiment using audio signals is a significant challenge [12][13] due to the difficulty of accurately determining the robust feature set needed to detect sentiments expressed within the audio signal [14][7]. Also, [12] some models failed to make accurate predictions in emotion recognition task and sentiment analysis tasks with higher numbers of classes. In addition, existing sentiment analysis models primarily focus on English or other widely studied languages, leaving a gap in sentiment analysis research specific to Tagalog speech which is the spoken language of the students. To address these challenges, this study focuses the optimized sentiment analysis particularly in Tagalog speech utilizing Principal Component Analysis (PCA) and Bidirectional Recurrent Neural Network (BRNN) on prosodic suprasegmental and Mel Frequency Cepstral Coefficients (MFCCs).

II. RELATED WORKS

An emerging field of study is audio sentiment analysis, which uses automatic speech recognition to extract a speaker's opinion or sentiment from genuine audio [15]. Sentiment analysis is the task of automatically determining the sentiment or emotional tone of text or speech to identify and extract information present in the source materials using natural language processing, computational linguistics, and text analytics [16] [17][18]. The rapid growth of Internet-based applications, such as social media platforms and blogs, has resulted in comments and reviews concerning day-to-day activities [14] [19]. Extracting speaker sentiment from

natural audio sources is a challenging problem. Generic methods for sentiment extraction generally use transcripts from a speech recognition system [20], and process the transcript using text-based sentiment classifiers [12][21].

In speech emotion recognition, there are many methodologies that can be used in order to reduce the dimensionality of data such as principal component analysis (PCA), Linear discriminant analysis (LDA), Random forests, etc. PCA seems to be one of the most popular methodologies. PCA is a preprocessing linear transformation technique [22].

The unsupervised method incorporates visual cues that are indicative of the semi-periodic articulatory production surrounding the orofacial region. Principal component analysis (PCA) is used to merge the visual components into a single feature [23].

Recent advancement of social media which is an enormous ever-growing source has led people to share their views through various modalities such as audio, text and video [24]. Furthermore, in this last few years, sentiment analysis (SA) has attracted increasing interest in the text mining area. It increasingly [25] becomes a popular research area for opinion mining in education that analyses and understands students' opinions toward their institutions for improving the quality of decision-making.

Thus, many machine learning and natural language processing-based algorithms have been utilized previously to examine these feelings [26]. However, because of their recent outstanding performance, deep learning-based techniques are rapidly gaining popularity.

Deep learning applies a multilayer strategy [27] to the neural network's hidden layers. Conventional machine learning techniques involve the human definition and extraction of features using feature selection techniques. Deep learning models, on the other hand, achieve higher accuracy and performance since features are automatically learned and extracted. In recent years, Deep Learning models have shown impressive achievements in computer vision and speech recognition [28].

Deep Learning (DL) methods were introduced to NLP after they achieved successful object recognition via ImageNet. DL methods improved statistical learning results in many fields. At present, a neural network-based NLP framework has achieved new levels of quality and become the dominating technology for NLP tasks [29], such as sentiment analysis machine translation, and question answering systems. Numerous Deep Learning (DL) and machine learning [30] techniques have emerged quickly and shown promise in a variety of applications, including audio, image, and natural language processing.

On the other hand, neural network-based models have outperformed traditional semantic techniques, such as Latent Semantic Analysis, in terms of word representations. Recurrent neural networks (RNNs) and CNNs have outperformed other neural networks in several tasks, particularly text classification tasks involving sentiment recognition [30] [31].

Prosody, the suprasegmental aspects of speech including intonation, rhythm, stress, and pitch, plays a crucial role in communication across languages and cultures [32]. The essential role of prosody in shaping the richness and complexity of human communication, emphasizing its relevance across diverse linguistic contexts and communicative settings [33]. Prosodic features on the other hand, like (pitch, intensity and speech rate) [12] are generally the most commonly implemented features for Speech Emotion Recognition (SER) as they are considered highly correlated with speaker emotion. Also, [34] prosody research has had a significant impact in improving the naturalness of speech synthesis, and has found some successes improving information extraction from speech, speech assessment and extracting affect.

III. METHODS

Analyzing sentiment using audio signals is a significant challenge due to the difficulty of accurately determining the robust feature set needed to detect sentiments expressed within the audio signal [14]. In this study, the development of the proposed optimized Tagalog speech sentiment model is shown in the following figure.

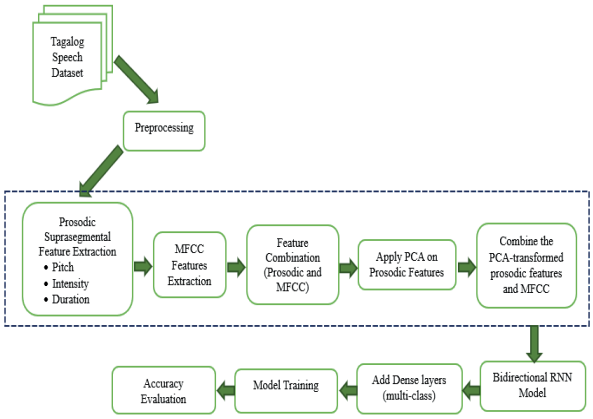

Fig. 1. The Proposed Tagalog Speech Sentiment Model.

A. Tagalog Speech Datasets

The student surveys and interviews through audio or speech provided the datasets. A condenser microphone was used for the recording, which took place in a closed space. As required by the research ethics requirements, the respondents were given an informed consent form to fill out.

B. Preprocessing

During the preprocessing stage, the elimination of any extraneous or noisy components from the speech data were done, such as background noise, sounds other than speech, or disruptions. Ensuring that the amplitude and scale are consistent throughout all recordings requires normalizing the voice data. Adjusting the gain or volume levels of the audio samples can be necessary for this. The Audacity application is utilized to edit the audion and to facilitate feature extraction and analysis by dividing the speech data into smaller chunks, such as words or phrases. An appropriate annotation was made in the appropriate directories when compiling Tagalog speech. Lastly, a librosa python library was used during this stage.

C. Prosodic Suprasegmental Feature Extraction

Feature extraction is one of the most important parts of a sentiment analysis. The speech signal has several different

properties. With the use of this technology, the best features from speech signal which include a plethora of information about speech are then extracted and selected Selecting the right components to enable the voice signal to transmit more emotional information. It's uncertain whether aspects of speech are best for identifying different emotions. Based on the prosodic suprasegmental features, we examined four different feature categories in this study: pitch, energy/intensity, duration, and intonation.

C.1 Pitch.

Pitch is one of the fundamental prosodic features used in speech sentiment analysis. It is vital to the expression of emotions through speech and represents the frequency of vibration of the vocal cords. Emotions such as happiness or enthusiasm are frequently correlated with higher pitches, whereas calmness or sad might be indicated by lower pitches. Below is the mathematical representation of pitch extraction process [35] [36].

C.2 Energy or Intensity

In audio processing, Root Mean Square (RMS) Energy is a metric used to estimate the strength or volume of an audio signal within a given time window. It is also helpful for examining the signal's intensity over time, which may indicate various characteristics including speech stress patterns, music dynamics, or the presence of speech.

C.3 Duration

An audio signal's duration, which is typically expressed in seconds, is the entire amount of time the audio file has been recorded. Knowing the duration is important for many tasks in audio processing, including segmentation, time-based analysis, and just figuring out how long the audio content is.

C.4 Intonation

Intonation refers to the variation in pitch while speaking, which can convey different meanings, emotions, or emphasis. It is how the pitch of the voice rises and falls throughout speech. Emotions like surprise, delight, grief, and sarcasm can all be expressed through intonation patterns. A sentence's finality or confidence, for instance, may be indicated by a falling intonation, whereas an inquiry or uncertainty may be shown by a rising intonation. These pitch changes can be captured in sentiment analysis to aid in differentiating between neutral, positive, and negative feelings.

D. Compute Feature Over Segments

To better collect prosodic information, analyze attributes across specific audio segments. Divide the audio into smaller segments and compute the mean features for each segment. The following are the formula for the computation of features:

D.1 Pitch

$$\text{Average\_Pitch} = \frac{1}{N}\sum_{i=1}^{N}\text{Pitch}_i \tag{1}$$

Where N is the number of pitch values in the segment

D.2 *Energy or Intensity*

$$\text{Average\_Intensity} = \frac{1}{M}\sum_{j=1}^{M}\text{RMS}_{t_j} \tag{2}$$

Where M is the number of segments or frames, and the RMS_{tj} is the RMS value for the *j-th* segment.

D.3 Duration

$$\text{Segment\_Duration} = \frac{\text{Number\_of\_Samples}}{\text{Sampling\_Rate}} \tag{3}$$

Where: Number_of_Samples is the number of samples in the segment

Sampling_Rate is the number of samples per second (Hz)

D.4 *Intonation*

$$\text{Mean Pitch} = \frac{1}{N}\sum_{i=1}^{N}F_0(t_i) \tag{4}$$

Where N is the number of pitch samples

Mean pitch gives the average pitch over a segment of speech, which can help identify the overall intonation pattern (e.g., whether the speech is generally high-pitched or low-pitched).

E. Mel-Frequency Cepstral Coefficients (MFCCs)

In the field of speech and audio processing, one popular feature extraction method is the Mel-Frequency Cepstral Coefficients (MFCC) function. MFCCs, which are produced by performing a linear cosine transform on a log power spectrum on a nonlinear Mel scale of frequency, depict the short-term power spectrum of a sound. The primary goal of MFCC is to replicate how the human auditory system perceives sound in order to increase its resilience to changes in the surrounding acoustic environment. Also, MFCCs offer a condensed depiction of the speech signal, reducing the number of dimensions in the data while maintaining the crucial details. The following is the formula:

where α is typically around 0.95 to 0.97

F. Principal Component Analysis

Principal component analysis, or PCA, is one statistical technique for minimizing the dimensionality of data while preserving as much variance as feasible. It transforms the data into a new coordinate system where the largest deviations by any projection of the data end up on the initial axes (principal components).

A set of related variables can be transformed into a new, uncorrelated set of variables known as prime elements using principal component analysis (PCA). However, if the data is already irrelevant, the PCA is meaningless. The principal components are placed orthogonally based on the variability they reflect, together with the irrelevant data. That is, for a given dimension, the primary principal part represents the ideal level of variability found in the original data set.

The algorithm and formula of PCA are the following:

- Center the Data: Subtract the mean of each feature from the data to center it around the origin:

$$\tilde{X} = X - \bar{X} \tag{5}$$

- Compute the Covariance Matrix – Calculate the covariance matrix of the centered data, where n is the number of samples.
- Compute the Eigenvalues and Eigenvectors –

$$C = \frac{1}{n-1}\tilde{X}^T\tilde{X} \tag{6}$$

Solve the eigenvalue problem for the covariance

$$Cv_i = \lambda_i v_i \tag{7}$$

matrix:
- Sort and select principal components
- Project the original data into the principal components where W is the matrix of the top k eigenvectors.

$$X_{\text{pca}} = \tilde{X}W \tag{8}$$

where λ_i are the eigenvalues and v_i are the eigenvectors

G. PCA-Transformed Prosodic and MFCC Features Combination

To generate a comprehensive feature vector for every speech sample, the transformed features are combined with the MFCC features following the application of PCA to the prosodic features. The model may access emotional and content-related data by combining prosodic features (after PCA) with MFCCs, which improves the model's ability to distinguish between various emotions. Together, these components enable the model to comprehend both the content (MFCCs) and the manner (prosodic characteristics) of the speech.

The speech signal is robustly and thoroughly represented by combining MFCC characteristics with PCA-transformed prosodic features. By utilizing the linguistic information from MFCCs and the emotional cues from prosodic features, this combined feature set improves the accuracy and efficiency of sentiment classification in speech-based models.

H. Bidirectional Recurrent Neural Network (BRNN)

A Bidirectional Recurrent Neural Network is a kind of neural network capable of concurrently processing sequences in both forward and backward orientations.

Speech, text, and time series are examples of sequential data that may be processed very well with a Bidirectional Recurrent Neural Network (BRNN), an advanced neural network implementation. BRNNs are utilized in Tagalog voice sentiment analysis to improve sentiment classification accuracy by capturing past and future contextual information in a speech sequence. An RNN that is bidirectional processes input data by passing it through two independent RNNs, one for each direction—forward and backward. To create the final result, processing is applied to both of these outputs simultaneously.

In figure 2 two separate unidirectional RNNs are combined to create a bidirectional RNN. The input sequence is processed by one of these unidirectional RNNs from left to right, and by the other from right to left.

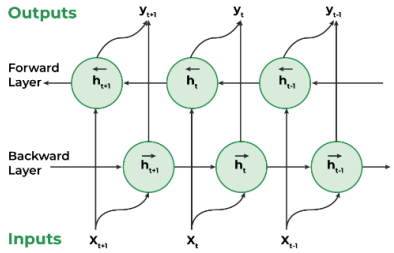

Fig. 2 A Bidirectional Recurrent Neural Network (BRNN) Model

H.1 Confusion Matrix - Accuracy

The ratio of correctly predicted instances (both positive and negative) to the total number of occurrences is used to determine accuracy, which assesses the overall correctness of the model:

$$\text{Accuracy} = \frac{TP + TN + \text{True Neutral}}{\text{Total Instances}} \tag{9}$$

H.2 Confusion Matrix – Precision

The precision measure is the percentage of positively (or negatively/neutrally) anticipated observations that are accurate compared to the total number of predicted observations that are accurately predicted as shown below. It shows the proportion of actual instances of a given sentiment class that were expected to be that sentiment.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

H.3 Confusion Matrix – Recall

Recall is the ratio of accurately predicted positive observations to all observations in the actual class. It is often referred to as Sensitivity or True Positive Rate:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

H.4 Confusion Matrix - F1-Score

The F1-Score, which offers a balance between precision and recall, is the harmonic mean of the two, particularly in cases where the distribution of classes is not similar:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

IV. RESULTS AND DISCUSSIONS

The simulation of optimized sentiment analysis of Tagalog speech model includes the following: python programming using the Google Colab, Audacity for editing the audio or speech samples, Librosa library in python, tensorflow, scikitLearn, and an Intel (R) Core i7 65000U CPU @ 2.50 GHz with 8 GB of RAM. Following the proposed Tagalog Speech Sentiment architecture or model the following results were obtained:

A. *Sentiment Analysis Confusion Matrix Results*

Figure 3 shows the result of the Tagalog speech sentiment analysis where majority of the instances were correctly classified as positive, negative, and neutral which got the accuracy rate of 90.91%.

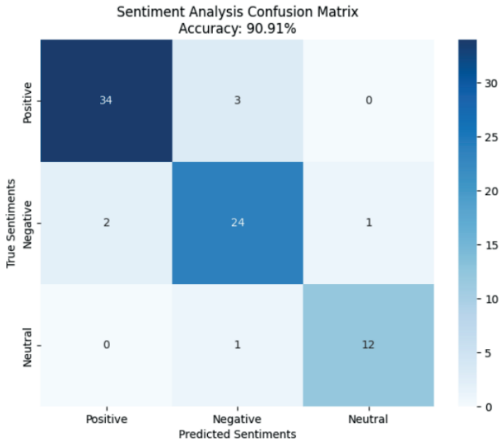

Fig. 3. Tagalog Speech Sentiment Analysis Confusion Matrix

B. *Model Accuracy*

The model accuracy plot illustrates a model's performance over training in terms of sentiment classification of Tagalog speech. In particular, a high percentage of accurate predictions provided by the model on the training or validation dataset is shown by an accuracy rate of 90.91%. The curve starts around 50% and steadily climbs to 90% over the progression of 50 epochs. Similarly, the curve start to progress in the validation accuracy and stabilize to 90.91%. Effective learning and high generalization are indicated by the consistency of the training and validation accuracy curves. Due to its outstanding accuracy, the model appears to be suitable for practical use, as it consistently captures the speech elements required for sentiment analysis.

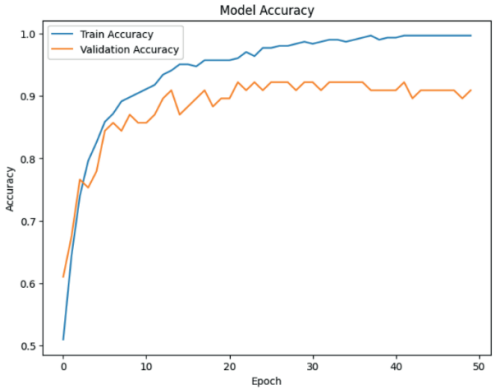

Fig. 4. Tagalog Speech Sentiment Accuracy Results

C. *Model Loss*

In this study, a model loss representation is another important tool for evaluating the performance of Tagalog speech sentiment. The declining and stabilizing loss curve represents effective learning and good generalization making accurate predictions on new and unseen data.

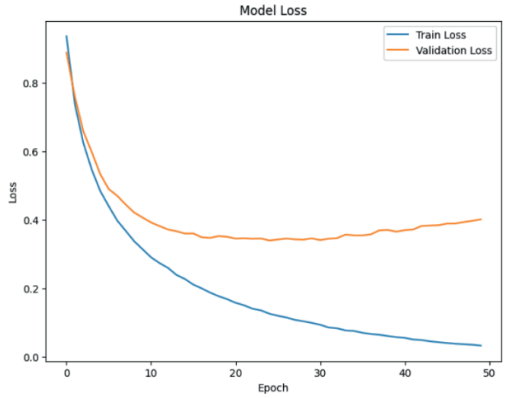

Fig. 5 Model Loss Result

V. CONCLUSION

This study explored the optimized sentiment analysis in Tagalog speech using PCA and Bidirectional RNN on prosodic suprasegmental and MFCC features. The accuracy results of 90.91% shows a significant performance. A key factor in improving model performance was the integration of prosodic and MFCC features, which are essential for capturing the spectral content and emotional tone of speech. Applying PCA to prosodic suprasegmental features reduced dimensionality, retained essential information, and mitigated the risk of overfitting. The Bidirectional RNN model effectively captured the temporal associations of speech data. The results of the study demonstrate the potential of sophisticated machine learning methods for speech sentiment analysis, especially for Tagalog language where research is quite limited. For further enhancement, consider the variability of frequency and amplitude in prosodic suprasegmental feature selection and extraction and extensive hyperparameter tuning on PCA to help the deep learning models with a more reduced feature.

ACKNOWLEDGMENT

The author would like to acknowledge the support provided by the people behind this undertaking. To Occidental Mindoro State College (OMSC) students and teachers for assisting with the collection of data. To Dr. Arnulfo T. Villanueva for his expertise in the field of Filipino Language. To her family, friends, professors, and classmates for the encouragement and guidance. Above all, to our almighty God, for wisdom, strength, and courage to pursue this endeavor.

REFERENCES

[1] Lou, Y. (2023, May). Deep learning-based sentiment analysis of movie reviews. In *Third international conference on machine learning and computer application (ICMLCA 2022)* (Vol. 12636, pp. 177-184). SPIE.

[2] Luo, Z., Xu, H., & Chen, F. (2019, January). Audio Sentiment Analysis by Heterogeneous Signal Features Learned from Utterance-Based Parallel Neural Network. In *AffCon@ AAAI* (pp. 80-87).

[3] Osmanoğlu, U. Ö., Atak, O. N., Çağlar, K., Kayhan, H., & Can, T. (2020). Sentiment analysis for distance education course materials: A machine learning approach. *Journal of Educational Technology and Online Learning*, *3*(1), 31-48.

[4] Baragash, R., & Aldowah, H. (2021, March). Sentiment analysis in higher education: a systematic mapping review. In *Journal of Physics: Conference Series* (Vol. 1860, No. 1, p. 012002). IOP Publishing.

[5] Lin, F., Liu, S., Zhang, C., Fan, J., & Wu, Z. (2023). StyleBERT: Text-audio sentiment analysis with Bi-directional Style Enhancement. *Information Systems*, *114*, 102147.

[6] Liu, G., Cai, S., & Wang, C. (2023). Speech emotion recognition based on emotion perception. *EURASIP Journal on Audio, Speech, and Music Processing*, *2023*(1), 22.

[7] Andayani, F., Theng, L. B., Tsun, M. T., & Chua, C. (2022). Hybrid LSTM-transformer model for emotion recognition from speech audio files. *IEEE Access*, *10*, 36018-36027.

[8] Dongbo, M., Miniaoui, S., Fen, L., Althubiti, S. A., & Alsenani, T. R. (2023). Intelligent chatbot interaction system capable for sentimental analysis using hybrid machine learning algorithms. *Information Processing & Management*, *60*(5), 103440.

[9] Kumar, R. G., & Shriram, R. (2019). Sentiment analysis using bi-directional recurrent neural network for Telugu movies. *International Journal of Innovative Technology and Exploring Engineering*, *9*(2), 241-245.

[10] Schachter, P., & Reid, L. A. (2018). Tagalog. In *The world's major languages* (pp. 852-876). Routledge.

[11] Boquiren, A. J., Garcia, R., Hungria, C. J., & de Goma, J. (2022). Tagalog Sentiment Analysis Using Deep Learning Approach With Backward Slang Inclusion. In *Proceedings of the International Conference on Industrial Engineering and Operations Management Nsukka*.

[12] Atmaja, B. T., & Sasou, A. (2022). Sentiment analysis and emotion recognition from speech using universal speech representations. *Sensors*, *22*(17), 6369.

[13] Bansal, M., Yadav, S., & Vishwakarma, D. K. (2021, April). A language-independent speech sentiment analysis using prosodic features. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1210-1216). IEEE.

[14] Luitel, S., & Anwar, M. (2022, August). Audio Sentiment Analysis using Spectrogram and Bag-of-Visual-Words. In *2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI)* (pp. 200-205). IEEE.

[15] Kaushik, L., Sangwan, A., & Hansen, J. H. (2017). Automatic sentiment detection in naturalistic audio. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *25*(8), 1668-1679.

[16] Raja, J. G. J. S., & Juliet, S. (2023, May). Deep learning-based sentiment analysis of Trip Advisor reviews. In *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 560-565). IEEE.

[17] Anand, S., & Patra, S. R. (2022). Voice and Text Based Sentiment Analysis Using Natural Language Processing. In *Cognitive Informatics and Soft Computing: Proceeding of CISC 2021* (pp. 517-529). Singapore: Springer Nature Singapore.

[18] Swetha, B. C., Divya, S., Kavipriya, J., Kavya, R., & Rasheed, A. A. (2017). A novel voice based sentimental analysis technique to mine the user driven reviews. *International Research Journal of Engineering and Technology*.

[19] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731-5780.

[20] Baragash, R., & Aldowah, H. (2021, March). Sentiment analysis in higher education: a systematic mapping review. In *Journal of Physics: Conference Series* (Vol. 1860, No. 1, p. 012002). IOP Publishing.

[21] Boukabous, M., & Azizi, M. (2022, March). Multimodal sentiment analysis using audio and text for crime detection. In *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)* (pp. 1-5). IEEE

[22] Rosenberg, A. (2018, June). Speech, prosody, and machines: Nine challenges for prosody research. In *Proc. Speech Prosody* (pp. 784-793).

[23] Tao, F. (2018). *Advances in Audiovisual Speech Processing for Robust Voice Activity Detection and Automatic Speech Recognition* (Doctoral dissertation).

[24] Mohanty, M. D., & Mohanty, M. N. (2022). Verbal sentiment analysis and detection using recurrent neural network. In *Advanced Data*

[25] Dongbo, M., Miniaoui, S., Fen, L., Althubiti, S. A., & Alsenani, T. R. (2023). Intelligent chatbot interaction system capable for sentimental analysis using hybrid machine learning algorithms. *Information Processing & Management*, *60*(5), 103440.

[26] Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, *53*(6), 4335-4385.

[27] Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, *9*(3), 483.

[28] Sohangir, S., Wang, D., Pomeranets, A., & Khoshgoftaar, T. M. (2018). Big Data: Deep Learning for financial sentiment analysis. *Journal of Big Data*, *5*(1), 1-25.

[29] Peng, S., Cao, L., Zhou, Y., Ouyang, Z., Yang, A., Li, X., ... & Yu, S. (2022). A survey on deep learning for textual emotion analysis in social networks. *Digital Communications and Networks*, *8*(5), 745-762.

[30] Abid, F., Li, C., & Alam, M. (2020). Multi-source social media data sentiment analysis using bidirectional recurrent convolutional neural networks. *Computer Communications*, *157*, 102-115.

[31] Zahouani, A. L., & Boubaker, H. (2023). Forecasting Crude Oil Price with Hybrid Approaches. *Rev. Econ. Financ.*, *21*, 564-576.

[32] Kaushik, L., Sangwan, A., & Hansen, J. (2017). Automatic Sentiment Detection in Naturalistic Audio. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25, 1668-1679.

[33] Liscombe, J. J. (2007). *Prosody and speaker state: paralinguistics, pragmatics, and proficiency*. Columbia University.

[34] Bektosheva, D. (2024). THE ROLE OF PROSODY IN COMMUNICATION. *Академические исследования в современной науке*, *3*(12), 102-104.

[35] McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015, July). librosa: Audio and music signal analysis in python. In *SciPy* (pp. 18-24).

[36] Koffi, E. (2020). A tutorial on acoustic phonetic feature extraction for automatic speech recognition (ASR) and text-to-speech (TTS) applications in African languages. *Linguistic Portfolios*, *9*(1),1