Journal of Theoretical and Applied Information Technology
15th January 2025. Vol.103. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TOWARDS ROBUST TAI LANGUAGE ASPECT-BASED SENTIMENT ANALYSIS: A WEAK-SUPERVISION APPROACH

KANKANA DUTTA[1], RIZWAN REHMAN[2]

[1,2]Assistant Professor, Dibrugarh University, Centre for Computer Science and Applications, Assam, India

E-mail: [1]kankanadutta@dibru.ac.in, [2]rizwan@dibru.ac.in

ABSTRACT

Speech signals carry extensive information about the speaker, including various non-linguistic elements such as sentiment, emotion, and intent. Analyzing these aspects has garnered significant attention due to its wide-ranging applications in fields such as human-computer interaction, mental health assessment, and social media analytics. This paper presents a novel approach to sentiment analysis for the Tai-Phake language, spoken by the Tai-Phake community of Assam, a language with limited linguistic resources. We propose a Convolutional Neural Network (CNN) model that classifies sentiments by leveraging Mel-Frequency Cepstral Coefficients (MFCC) features extracted from speech data. The scarcity of linguistic resources for Tai-Phake posed substantial challenges, necessitating innovative solutions to develop an effective sentiment analysis tool. Our study also benchmarks the performance of the proposed CNN model against two other popular methods: a Multi-Layer Perceptron (MLP) classifier and a Long Short-Term Memory (LSTM) network. Extensive experiments reveal that the CNN model achieves superior performance, with higher accuracy and robustness in classifying sentiments compared to the MLP and LSTM classifiers. This indicates the effectiveness of convolutional architectures in capturing the intricate patterns in speech signals relevant to sentiment analysis. The findings underscore the potential of CNNs in handling sentiment classification tasks in resource-constrained languages like Tai-Phake, paving the way for further advancements in the field of speech-based sentiment analysis.

Keywords: *Machine Learning methods, Neural Networks, CNN, Sentiment Analysis, MFCC*

1. INTRODUCTION

Sentiment analysis is a critical task in the field of Natural Language Processing (NLP) which is used to determine the sentiment or the polarity of the statements for the given context. Sentiment analysis determines whether the context is positive, negative, or neutral [1]. Sentiment analysis can be applied to data of different forms such as text, speech, video, and image. With the growth of Artificial Intelligence, several systems are implemented using voice-based technologies that involve human-machine interaction. The human voice can carry information related to the speaker as well as the information being conveyed. The speech signal can convey different meanings if the emotional status of the speaker is different. In this case, recognizing non-linguistic pieces of information such as sentiment, intention, emotion, etc. can greatly enhance the performance of the machine significantly. Since sentiment and emotion are close (sentiments are often represented as valence, -which is an emotional attribute), these two terms are often used interchangeably. Furthermore, the same model can be used to evaluate both problems [2].

An attempt to recognize sentiment from speech data has been performed in this study using different supervised machine-learning methods. Due to the development of neural networks in recent years, different neural network models have become popular models for sentiment analysis and emotion detection.

In this experiment, speech data from the language Tai-Phake has been used. Tai-Phake, also known as Phakial or Phake is spoken by the Phake people in the North-Eastern part of India, mainly in Assam and Arunachal Pradesh. The Tai Phake language belongs to the Tai-Kadai language family, which is a group of languages spoken in Southeast Asia. This family includes Thai, Lao, Shan, and various other languages. The Phake community is relatively small, with an estimated population of 2 thousand [3]. The Tai-Phake language is a tonal language meaning that pitch or intonation used while speaking a word can change the meaning. Due to that, experiments on this language are a challenging

task. The tools to be used in these experiments need to be trained on a large dataset, which is again a challenging task due to a small number of speakers.

In this study, we have tried to develop a model for sentiment analysis for the language Tai Phake. The primary contributions of this paper are

1. Development of a dataset with speech data in three different sentiment classes.
2. Development of a Neural Network model for speech sentiment analysis.
3. Analyze the result by comparing the result with other supervised methods.

2. LITERATURE REVIEW

Syamala and Nalini experimented with aspect-based sentiment analysis by combining both acoustic features of speech and linguistic features of text. The deep learning model is used for speech feature extraction while different text feature extraction methods are used for aspect extraction and the proposed model proved the effectiveness of the proposed hybrid model compared to different existing models [4].

Aspect-based sentiment analysis was done from product review speech data using a hybrid model where speech emotion was recognized first by using a Deep CNN, then speech aspect was recognized by using BiLSTM, and a rule-based classifier for aspect-wise sentiment classification. MFCC features were extracted from the speech data from four different datasets and obtained accuracy rates of 93.28%, 91.45%, 92.12%, and 90.45% [5].

Nicolini M. and Ntalampiras S worked on a multilingual model that works on multiple levels. It first identifies the speaker's gender in the first level and the speaker's emotional state in the second level. Features such as GTCC, Delta GTCC, Delta-delta MFCC, Male Spectrum, and spectral crest were applied on different classifiers such as kNN, YAMNet, BiLSTM, etc., and found different results for both male and female speakers [6].

A multi-lingual approach for emotion detection was proposed by Padam and Magare for an Indian regional language, the Marathi language. Different speech features such as pitch, formant frequency, and MFCC are used individually for the analysis and detection of six types of emotions, and found that compared to other features, pitch, and MFCC values give better results [7].

Tamulevi˘ciuset *al.*, performed a cross-linguistic speech emotion recognition for the English, Lithuanian, German, Spanish, Serbian, and Polish languages(size 10,000) with two-dimensional acoustic feature spaces, such as cochleagrams, spectrograms, mel-cepstrograms, and fractal dimension-based space, are employed as the representations of speech emotional features on the CNN model and the results show the superiority of cochleagrams over other features and the accuracy rate is close to a monolingual dataset [8].

A method for emotion detection using CNN for continuous real-time data with MFCC features was proposed by Attar *et al.* who found an accuracy of 78.8% [9].

Byun and Lee experimented on a Korean dataset using RNN for emotion detection and they proposed a feature combination to improve efficiency. They found that the method involving harmonic features showed better results than dissonant speech [10].

Choudhary *et al*. used MFCC and Mel Spectrum features in different deep learning models such as CNN, LSTM, and GRU and found that the CNN model works best with a 97.1% accuracy rate [11].

Jacob experimented on Hindi speech data with the first four formant frequencies and bandwidth in a neural network and 97.14% accuracy was achieved [12].

Swain *et al.* worked on an Odia language speech dataset and designed an ensembled classifier using a Deep Convolutional Recurrent Neural Network. The model was used to extract deep features which were represented as 3-D log Mel-spectrograms. Then a bi-directional-gated recurrent unit network was applied to produce utterance-level emotion and finally, ensemble classifiers using SoftMax and Support Vector Machine classifier were used to improve the final recognition rate. This model was used to classify seven emotional states and achieved an accuracy rate of 85.31% [13].

Maghilnan experimented with emotion detection from speaker-specific speech data using a different number of MFCC features. Different supervised machine learning algorithms such as linear SVM, and Naive Bayes along with a proposed model were used for the experiments, and found that for 12 to 14 features, the best results were obtained [14].

Table 1: Comparison of Related Work

Sl. No	Paper Name	Method Used	Features	Accuracy
1.	An Efficient Aspect-based Sentiment Analysis Model by the Hybrid Fusion of Speech and Text Aspects [4].	SVM, Random Forest, Naïve Bayes Logistic Regression Decision Tree K-Nearest Neighbour	Both acoustic and linguistic features	95% with Random Forest
2.	Aspect-Based Sentiment Analysis of Customer Speech Data Using Deep Convolutional Neural Network and BiLSTM [5].	CNN, BiLSTM	MFCC	93.28%, 91.45%, 92.12%, and 90.45% for four different datasets.
3.	A Hierarchical Approach for Multilingual Speech Emotion Recognition [6].	kNN, YAMNet, and BiLSTM	GTCC, Delta GTCC, Delta-delta MFCC, Male Spectrum, spectral crest	94%
4.	Regional language Speech Emotion Detection using Deep Neural Network [7].	SVM	pitch, formant frequency, and MFCC	100%
5.	A Study of Cross-Linguistic Speech Emotion Recognition Based on 2D Feature Spaces [8].	CNN	cochleagrams, spectrograms, mel-cepstrograms, and fractal dimension-based	88 % to 96% for different datasets
6.	Speech Emotion Recognition System Using Machine Learning [9].	CNN	MFCC	78.8%
7.	A Study on a Speech Emotion Recognition System with Effective Acoustic Features Using Deep Learning Algorithms [10].	RNN	F0, Mel-frequency cepstrum coefficients, spectral features, harmonic features, and others.	83.81%
8.	Speech Emotion-Based Sentiment Recognition using Deep Neural Networks [11]	CNN, LSTM, and GRU	MFCC, Mel Spectrum	97.1%
9.	An Improved Hindi Speech Emotion Recognition System [12]	Neural Network	Formant frequency	97.14
10	A DCRNN-based ensemble classifier for speech emotion recognition in the Odia language [13].	ensembled classifier using a Deep Convolutional Recurrent Neural Network.	deep features which were represented as 3-D log Mel-spectrograms	85.31%
11	Sentiment Analysis on Speaker Specific Speech Data [14].	SVM, Naive Bayes	MFCC	96%

3. MOTIVATION AND OBJECTIVES

Sentiment analysis offers numerous benefits in different applications by providing information that can help improve human-computer interaction. From the literature review, it is observed that over the years diverse methodologies have been proposed, incorporating different features for sentiment classification. Different speech features like pitch, formant frequency, MFCC, Mel, etc have been used for sentiment analysis and emotion detection. Different machine learning methods, mainly kNN, SVM, Random Forest, and different neural network models such as RNN, and CNN have been as classifiers.

Most of the research in this area has focused on widely spoken languages. There is also notable research on widely spoken Indian languages such as Hindi, Gujarati, Marathi, Tamil, Bangla, Assamese, Telegu, Udiya, Kannada, Punjabi, etc. [12][15].

However, many low-resource languages in India with unique phonetic characteristics remain unexplored. Tai-Phake is one such language that has yet to be studied extensively. By conducting research in the Tai-Phake language, we aim to fill this gap in the literature. Also motivated by the growing significance of sentiment classification in various applications and the lack of research in this field for the Tai-Phake language, this paper aims to explore and enhance the understanding of sentiment analysis from speech data for the Tai-Phake language. The objectives of this study are as follows:

1. To develop a robust sentiment analysis tool using a neural network model from speech data based on speech features.
2. To compare the performance of the proposed model with other machine learning classifiers.

4. METHODOLOGY

Sentiment analysis aims to identify and classify the underlying sentiments expressed in speech signals, which carry a wealth of information about the speaker. This includes not only linguistic content but also non-linguistic elements such as sentiment, emotion, and intent. The challenge lies in accurately extracting and utilizing these features to build robust sentiment analysis models, especially for under-resourced languages like Tai-Phake. This section outlines the methodology adopted for developing our sentiment analysis tool, leveraging Convolutional Neural Networks (CNNs) and comparing its performance with Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) networks.

4.1 Feature Extraction

The initial step in sentiment analysis involves extracting relevant features from the speech signal. These features are critical for distinguishing between different sentiments. Commonly used features include:

Pitch: The perceived frequency of the sound, which can indicate the emotional state of the speaker.

Energy: The intensity of the speech signal, often correlating with the speaker's emotional intensity.

Mel-Frequency Cepstral Coefficients (MFCC): MFCCs are widely used in speech and audio processing as they effectively represent the short-term power spectrum of a sound.

Recent studies have shown that these features, when appropriately extracted and combined, significantly improve sentiment and emotion recognition accuracy [16][17][18][19]. However, it remains a challenge to determine the most critical features for sentiment analysis, as the importance of specific features can vary across different languages and datasets [18].

4.2 Feature Extraction Process

In our approach, we use the following steps for feature extraction:

Preprocessing: The speech signals are first preprocessed to remove background noise and normalize the amplitude.

Segmentation: The continuous speech signal is segmented into smaller frames (typically 20-40 milliseconds) to capture the dynamic nature of speech.

MFCC Extraction: For each frame, MFCC features are computed. These coefficients are derived by taking the Fourier transform of the signal, mapping the powers of the spectrum to the Mel scale, taking the logarithm, and then applying the inverse Fourier transform.

Additional Features: Alongside MFCCs, pitch, and energy features are also extracted for each frame to capture additional emotional cues.

4.3 Model Architecture

The different types of neural network models used in the experiment are:

Convolutional Neural Network (CNN): The proposed model is based on a CNN architecture,

chosen for its ability to automatically learn spatial hierarchies of features through backpropagation by using multiple building blocks, such as convolution layers, pooling layers, and fully connected layers.

Convolutional Layers**:** These layers apply convolutional filters to the input data, capturing local patterns such as edges in images or, in our case, specific speech features.

Pooling Layers: Pooling layers reduce the dimensionality of the feature maps, which helps to make the computation more manageable and reduces the risk of overfitting.

Fully Connected Layers**:** These layers act as the classifier, taking the high-level features learned by the convolutional and pooling layers and producing the final sentiment classification.

Multi-Layer Perceptron (MLP): An MLP consists of multiple layers of perceptron (neurons) with activation functions. Each neuron in one layer is connected to every neuron in the next layer, making it a densely connected network.

Input Layer: Takes the extracted features as input.

Hidden Layers: Multiple layers with nonlinear activation functions to model complex relationships.

Output Layer: Uses a SoftMax activation function to classify the sentiment into predefined categories.

Long Short-Term Memory (LSTM): LSTMs are a type of Recurrent Neural Network (RNN) capable of learning long-term dependencies, making them suitable for sequential data like speech signals.

Input Layer: Takes the sequential features as input.

LSTM Layers: Use memory cells to store and update information over time, effectively capturing temporal dependencies in the speech signal.

Output Layer: Uses a SoftMax activation function for sentiment classification.

4.4 Training and Evaluation

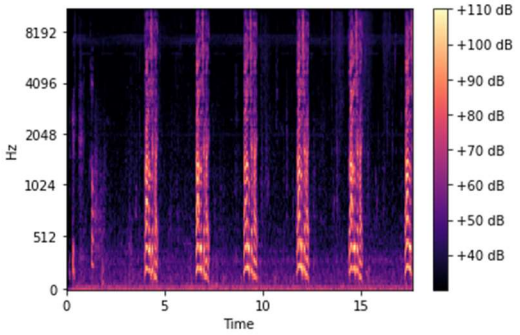
The models were trained on a labeled dataset of Tai-Phake speech signals. The dataset was divided into training, validation, and test sets to ensure a fair evaluation of the model's performance. The training process involved optimizing the model parameters to minimize the loss function, typically using gradient descent-based algorithms.

4.5 Features:

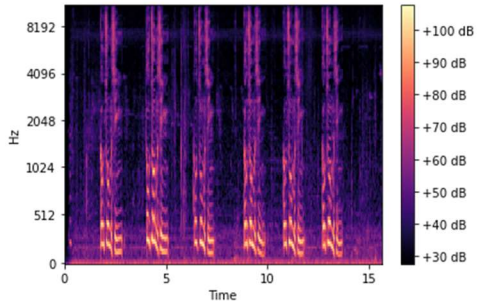
Over the years different methodologies have been proposed with different acoustic features for Sentiment Analysis and Emotion Detection. Speech features like MFCC, Mel, ZCR, RMS, etc have been used in this process. By integrating two or more features, it is possible to enhance the performance of some systems. Tai-Phake is a tonal language, therefore the selection of features is critical due to its tone. For this experiment, we have used Mel-Frequency Cepstral Coefficients (MFCC) features. MFCCs are derived from the Mel-frequency cepstrum, which is a representation of the short-term power spectrum of a sound. This technique is designed to mimic the human ear's nonlinear perception of sound frequencies, making it particularly useful for tasks like speech recognition and audio classification. For this experiment, we are using 40 numbers of MFCC coefficients.

4.6 Dataset

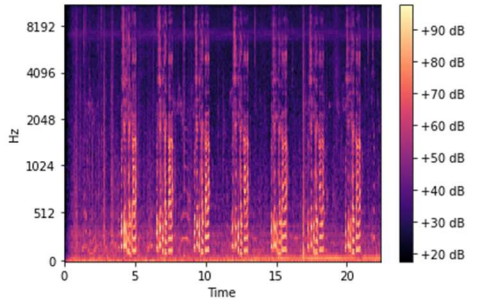
The dataset used for this experiment is a Tai Phake language dataset, which is a primary dataset. For developing the dataset, we have considered a total of 20 speakers or voice actors, out of which 10 are male and 10 female. The speakers are native Tai-Phake language speakers aged between 20 and 55. The gender distribution is balanced. Speech samples from each speaker are collected by speaking the sentences in 3 different sentiments. This dataset can be considered as a semi-natural dataset as the sentiments are artificially created sentiments and the speakers knew that they were recorded for analysis purposes. The sentiment class is divided into three classes, positive, negative, and neutral. Figure 1 represents the Mel spectrogram for the three classes of sentiments. There are a total of 1500 speech samples present in the dataset with an equal number of samples in each category as shown in Figure 2. The audio files are in .wav format with a sampling rate of 22050. The audio files are of good quality and were recorded in a noise-free environment.



A) Negative



B) Positive



C) Neutral

Figure 1: Mel Spectrogram Of The Three Sentiment Classes A) Negative B) Positive, And C) Neutral

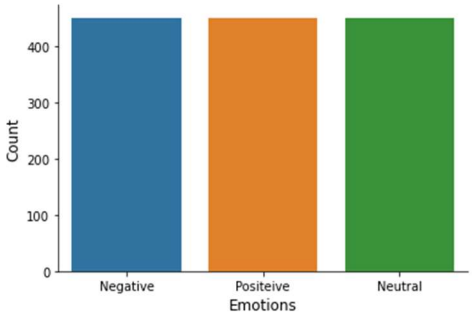


Figure 2: Dataset Sentiment Class Distribution

4.7 Sentiment Analysis Model

Classic machine learning algorithms such as SVM, HMM, GMM, etc. are some of the popular methods used in sentiment analysis. Traditionally these methods are used with different features such as pitch, formant frequency, MFCC, etc. Recently, with the growth of neural networks, RNN and CNN have become a good choice for sentiment analysis and emotion detection.

This study used the Convolutional Neural Network (CNN) model for sentiment analysis from speech. First, we extracted all the MFCC features from the sound files using the Python Librosa library. For the classification of the sentiments, a CNN is used with 1 input layer, 3 hidden layers, and 1 output layer. The architecture of the proposed CNN is shown in Table 2. The input layer has 40 input values which takes the 40 MFCC features extracted from the speech signal. Three hidden layers are used in the system and the activation functions used are RELU (Rectified linear unit) and softmax. The output layer has 3 output values representing the three types of sentiments. The network was tested with various hyperparameters and the best result was obtained with the hyperparameter given in Table 3.

Table 2: CNN Architecture

Layer Name	Input Shape
Input layer	40
Hidden layer	--
Output layer	3

Table 3: CNN Hyperparameters

Parameter	Value
Epochs	50
Batch size	64
Learning Rate	0.00001
Optimizer	Adam

For the experiments, the dataset is divided into three parts viz. training, testing, and validation. The training part is used to train the models used in the experiments and for training 80 percent of the data in the dataset are considered. The testing part of the data is used to test the model after it is trained with the training data. For this purpose, 15 percent of the data are used and the rest 5 percent of the data is used as validation data for evaluating the results of the experiments.

Another two classifiers, MLP classifier and LSTM were also used with the 80/20 percent data for training and testing purposes to classify the data into three categories positive, negative, and neutral to compare the result with the result given by the CNN model.

5. RESULTS AND DISCUSSION:

The experiments conducted aimed to evaluate the performance of three different models: Convolutional Neural Network (CNN), Multi-Layer Perceptron (MLP), and Long Short-Term Memory (LSTM) networks using Mel-Frequency Cepstral

Coefficients (MFCC) features for sentiment analysis of Tai-Phake speech data. The results are summarized in Table 4.

Table 4: Accuracy Rate Of Different Methods Applied

Method Name	Test Accuracy
MLP Classifier	93.33%
LSTM	35.53%
CNN	95.56%

5.1 Performance Comparison

The CNN model outperformed the other two models, achieving a test accuracy of 95.56%. This high accuracy underscores the effectiveness of the convolutional architecture in capturing relevant patterns in speech features, which are critical for sentiment classification. The performance of the CNN model is graphically represented in Figure 3, highlighting its robustness and precision in sentiment analysis tasks.

The MLP classifier, with a test accuracy of 93.33%, was the second-best performing model. While it performed admirably, it lagged behind CNN in terms of capturing the intricate nuances of the speech data, which likely contributed to the slight drop in accuracy.

The LSTM model, however, exhibited significantly lower performance, with an accuracy of 35.53%. This poor performance may be attributed to several factors:

Data Complexity: The Tai-Phake language's phonetic and prosodic complexities may not have been effectively captured by the LSTM model, leading to inadequate sentiment classification.

Overfitting: The LSTM's tendency to overfit, especially with limited data, could have resulted in poor generalization to the test set.

Temporal Dependencies: While LSTMs are designed to capture temporal dependencies, the specific nature of the MFCC features and their representation might not have aligned well with the LSTM's strengths.

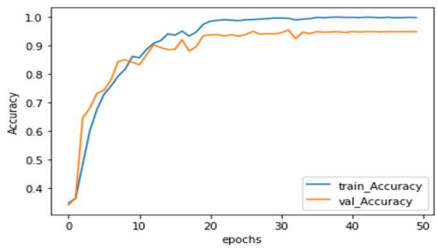


Figure. 3 Training And Validation Performance Of The CNN Model

The superior performance of the CNN model can be attributed to its ability to learn spatial hierarchies of features through convolutional layers. CNNs are particularly adept at identifying local patterns and structures within the data, which is crucial for tasks involving speech signals. The pooling layers further aid in reducing the dimensionality, making the model more computationally efficient while preventing overfitting.

The MLP classifier's relatively high accuracy demonstrates that even simpler neural network architectures can perform well when fed with carefully extracted features like MFCCs. However, its lack of capability to capture local dependencies and hierarchical patterns as effectively as CNNs explains the slight performance gap.

The LSTM model's underperformance was unexpected given its theoretical advantages in handling sequential data. This suggests that for the specific task of sentiment analysis using MFCC features, the temporal aspect captured by LSTMs may not be as critical or that the LSTM model requires further tuning or a different representation of the input features.

The results highlight the importance of choosing the right model architecture for specific tasks in speech processing. For sentiment analysis of Tai-Phake speech data, CNNs have proven to be highly effective. This finding is consistent with other studies that have demonstrated the superiority of CNNs in various speech and audio-processing tasks [16][20].

Moreover, the results underscore the need for further research into optimizing LSTM models for speech-related tasks or exploring hybrid models that combine the strengths of CNNs and RNNs.

The model was used to classify three classes of sentiments, namely positive, negative, and neutral, and the performance of the model for each class is shown in the confusion matrix in Figure. 4. The model recognized negative sentiment at a higher rate compared to positive and neutral sentiments. The recognition rate of negative sentiment is 98% while the recognition rate for positive and neutral is 92% and 93% respectively.

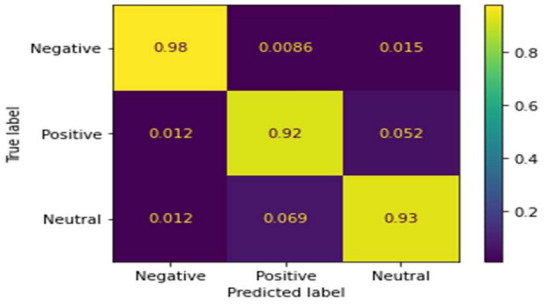


Figure. 4 Confusion Matrix

To further evaluate the performance and practical applicability of our sentiment analysis model, we conducted a perception test involving 20 naïve listeners. The participants were native Tai-Phake speakers, ranging in age from 12 to 75 years, ensuring a diverse representation of the community. The objective was to assess how well humans could classify the sentiments in the audio samples and to compare this with the performance of our automated model.

The listeners were presented with a set of audio samples and asked to classify each one into one of three sentiment categories: positive, negative, or neutral. The task was designed to be straightforward, with each participant listening to the same set of samples under consistent conditions. The results of this perception test are detailed in Table 5.

Table 5: Result Of The Perception Test

Listeners	Accuracy of classification
1	100
2	98.33
3	96.67
4	98.33
5	90.00
6	86.67
7	98.33
8	100
9	93.33
10	93.33
11	90.00
12	90.00
13	90.00
14	90.00
15	93.00
16	93.00
17	66.67
18	66.67
19	98.33
20	98.33
Average	**91.55**

The perception test results indicate that human listeners achieved an average classification accuracy of 91.55%. Notably, two listeners were able to identify all the sentiments correctly, achieving a perfect recognition rate of 100%. This high level of accuracy among the majority of participants suggests that the Tai-Phake speakers are quite adept at discerning sentiments in their native language. However, there was some variability in the results. Two listeners had notably lower accuracy rates of 66.67%. This discrepancy might be due to individual differences in auditory perception, familiarity with the test procedure, or even minor variations in the audio playback environment. Despite these outliers, the overall high average accuracy demonstrates the reliability of human perception for sentiment classification in this language context. The high average accuracy observed in the perception test aligns well with the performance of our CNN model, which achieved an accuracy of 95.56%. This comparison highlights the effectiveness of the model in replicating human-like sentiment classification performance. The slight differences can be attributed to the model's consistency and lack of subjective biases that might affect human listeners. Furthermore, these results underscore the potential of integrating human insights into the development and refinement of automated sentiment analysis tools. By understanding the strengths and limitations of human perception, we can better tune our models to capture the nuances of speech sentiment more accurately.

6. CONCLUSION:

This paper presents the development of a Convolutional Neural Network (CNN) model for sentiment classification from speech data, using Mel-Frequency Cepstral Coefficients (MFCC) features, specifically tailored for the Tai-Phake language, which is both tonal and low-resourced. In addition to the CNN model, two other classification methods were evaluated for comparison: Support Vector Machine (SVM) and Recurrent Neural Network (RNN). A primary dataset of Tai-Phake speech samples was utilized for the experiments. The experimental results demonstrated that the proposed CNN model significantly outperformed the other two methods, achieving a test accuracy of 95.56%. This high performance can be attributed to CNN's ability to effectively capture and leverage the intricate patterns in the MFCC features, which are crucial for sentiment analysis in speech. Given the lack of existing models for sentiment analysis in the Tai-Phake language, we conducted a perception test involving 20 native Tai-Phake speakers. The

perception test aimed to assess human accuracy in classifying the sentiments of the same speech samples. The results showed an average classification accuracy of 91.55% among the participants. This close alignment between human performance and the CNN model's accuracy underscores the model's effectiveness and reliability. Overall, the study highlights the potential of using CNNs for speech-based sentiment analysis in under-resourced languages. The findings not only demonstrate the CNN model's superior performance compared to SVM and RNN but also validate its accuracy through a human perception test. This research paves the way for further advancements in sentiment analysis for other low-resourced languages, emphasizing the importance of tailored feature extraction and model selection.

REFERENCES

[1] Bing Liu, "Sentiment Analysis and Opinion Mining", *Synthesis Lectures on Human Language Technologies, vol. 5, no. 1*, 2012, pp. 1-167.

[2] Bagus Tris Atmaja, Akira Sasou, "Sentiment Analysis and Emotion Recognition from Speech Using Universal Speech Representations", *Sensors*., 2022

[3] Sanjay Barman, "A Study on Culture of Tai Phake Community of Assam", International *Journal of Social Science and Humanities Research* , Vol. 8, Issue 2, 2020, pp: 290-292.

[4] Maganti Syamala, Nalini N.J., "An Efficient Aspect-based Sentiment Analysis Model by the Hybrid Fusion of Speech and Text Aspects", *(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 9*, 2021, pp.160-169.

[5] Sivakumar Murugaiyan, U. Srinivasulu Reddy, "Aspect-Based Sentiment Analysis of Customer Speech Data Using Deep Convolutional Neural Network and BiLSTM", *Cognitive Computation*, 2023, pp.914–931.

[6] Nicolini, M. and Ntalampiras, S., "A Hierarchical Approach for Multilingual Speech Emotion Recognition", *In Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, 2023, pp. 679-685

[7] Sweta Padman, Dhiraj Magare, "Regional language Speech Emotion Detection using Deep Neural Network", *ITM Web of Conferences* , 2022.

[8] Gintautas Tamulevi˘cius, Gražina Korvel, Anil Bora Yayak, Povilas Treigys, Jolita Bernatavi˘cien˙e, and Bo˙zena Kostek, "A Study of Cross-Linguistic Speech EmotionRecognition Based on 2D Feature Spaces", *Electronics,* 2020.

[9] Husbaan I. Attar, Nilesh K. Kadole, Omkar G. Karanjekar, Devang R. Nagarkar, Prof. Sujeet More, "Speech Emotion Recognition System Using Machine Learning", *International Journal of Research Publication and Reviews*, Vol 3, no 5, May 2022, pp 2869-2880.

[10] Byun, Sung-Woo & Lee, Seok-Pil. "A Study on a Speech Emotion Recognition System with Effective Acoustic Features Using Deep Learning Algorithms", *Applied Sciences*.,2021.

[11] Ravi Raj Choudhary, Gaurav Meena, and Krishna Kumar Mohbey, "Speech Emotion Based Sentiment Recognition using Deep Neural Networks", *2nd International Conference on Computational Intelligence & IoT (ICCIIoT), Journal of Physics*, 2021.

[12] Agness Jacob, P Mythili, "An Improved Hindi Speech Emotion Recognition System", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Volume-3 Issue-6, November, 2013.

[13] Swain, Monorama & Maji, Bubai & Kabisatpathy, Prithviraj & Routray, Aurobinda. "A DCRNN-based ensemble classifier for speech emotion recognition in Odia language", *Complex & Intelligent Systems*. 2022. Pp. 1-13.

[14] Maghilnan S, Rajesh Kumar M, "Sentiment Analysis on Speaker-Specific Speech Data", *International Conference on Intelligent Computing and Control (I2C2),* 2017.

[15] Pramod Mehra, Parag Jain, "ERIL: An Algorithm for Emotion Recognition from Indian Languages Using Machine Learning", *Wireless Personal Communications*, 2022, pp.126:2557–2577.

[16] Vaibhav K. P., Parth J. M., Bhavana H. K., Akanksha S. S., "Speech-Based Emotion Recognition Using Machine Learning", *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* 2021.

[17] Girija Deshmukh, Apurva Gaonkar, Gauri Golwalkar, Sukanya Anil Kulkarni, "Speech-based Emotion Recognition using Machine", *International Journal of Research and Analytical Reviews (IJRAR),* 2019.

[18] Linhui Sun, Yiqing Huang, Pingan Li, "Multi-classification speech emotion recognition based on two-stage bottleneck features selection and MCJD algorithm", *Signal, Image and Video Processing,* 2022.

[19] Samhith, T. Sai, G.Nishika, M.Prayuktha, Manik Chandra and G.Prasadu Dr.Sunil Bhutada, "Speech Emotion Recognition using Machine Learning Algorithms", *International Journal of Creative Research Thoughts*, 2021.

[20] Yogesh Kumar, Manish Mahajon, "Machine Learning Based Speech Emotions Recognition System", *International Journal of Scientific & Technology Research*, Volume 8, Issue 07, 2019, pp. 722-729.