EmoTa: A Tamil Emotional Speech Dataset

Jubeerathan Thevakumar and Luxshan Thavarasa and Thanikan Sivatheepan
and Sajeev Kugarajah and Uthayasanker Thayasivam
Department of Computer Science and Engineering
University of Moratuwa
Colombo, Sri Lanka
{jubeerathan.20, luxshan.20, thanikan.20, kugarajah.21, rtuthaya}@cse.mrt.ac.lk

Abstract

This paper introduces EmoTa, the first emotional speech dataset in Tamil, designed to reflect the linguistic diversity of Sri Lankan Tamil speakers. EmoTa comprises 936 recorded utterances from 22 native Tamil speakers (11 male, 11 female), each articulating 19 semantically neutral sentences across five primary emotions: anger, happiness, sadness, fear, and neutrality. To ensure quality, inter-annotator agreement was assessed using Fleiss' Kappa, resulting in a substantial agreement score of 0.74. Initial evaluations using machine learning models, including XGBoost and Random Forest, yielded a high F1-score of 0.91 and 0.90 for emotion classification tasks. By releasing EmoTa as open-access, we aim to encourage further exploration of Tamil language processing and the development of innovative models for Tamil Speech Emotion Recognition.

1 Introduction

The development of emotional speech datasets has significantly advanced Speech Emotion Recognition (SER), enhancing human-computer interaction by enabling systems to interpret and respond to emotions in nuanced, human-like ways (Cowie et al., 2001). However, for low-resource languages like Tamil, particularly in the Sri Lankan context, high-quality emotional datasets remain scarce. This scarcity restricts the development of SER models tailored to Tamil-speaking communities, limiting applications in localized assistive technology, emotion-based mental health diagnostics, and customer service for Tamil-speaking users.

In this work, we introduce EmoTa[1] , a novel, acted Tamil emotional speech dataset that captures the linguistic and cultural diversity of Sri Lankan Tamil speakers. The dataset focuses on five fundamental emotionsanger, happiness, sadness, fear, and neutralitychosen for their wide recognition and relevance in SER and to reflect the range of emotional expressions commonly encountered in everyday Tamil communication. These core emotions also align with those used in established SER datasets, allowing comparative studies while maintaining a culturally relevant focus. To ensure clarity and isolate prosodic cues, we use semantically neutral sentences, a practice inspired by similar works in emotional speech corpora such as EMOVO (Costantini et al.), EMODB (Burkhardt et al., 2005), CaFE (Gournay et al., 2018). This approach reduces the risk of lexical bias, ensuring that models trained on this dataset can accurately recognize emotional tone independent of content.

The accessibility of this high-quality emotional speech dataset is a cornerstone of its contribution, enabling reproducibility, facilitating collaboration, and supporting the validation and benchmarking of SER models across varied Tamil dialects. By filling an essential gap in Tamil SER resources, this open-access dataset advances the SER field for low-resource languages, serving as a reference point for related work and creating pathways for future developments.

2 Related Work

Established SER datasets such as EMOVO (Italian) (Costantini et al.), EMODB (German) (Burkhardt et al., 2005), and CaFE (Canadian French) (Gournay et al., 2018) have set benchmarks by providing structured, acted datasets that ensure consistency, quality, and controlled variation in emotional portrayals. These datasets highlight the benefits of the acted approach, where carefully guided performances yield emotionally distinct, reliable data that enhances model training and generalization.

For Tamil, some progress has been made in developing emotional speech datasets. Rajan et al. (2019) introduced TaMaR-EmoDB, a multilingual emotional speech database covering Tamil, Malay-

[1] https://github.com/aaivu/EmoTa

Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL 2025), pages 193–201
January 19, 2025. ©2025 International Committee on Computational Linguistics

alam, and Ravula. However, this corpus remains inaccessible to the public, limiting reproducibility and broader applicability. Similarly, Ram and Ponnusamy (2014) created a custom Tamil emotional speech dataset using standard feature extraction techniques, but it focused on a specific subset of Tamil speakers. Vasuki et al. (2020) developed two Tamil emotional corpora for children and adults, drawn from Tamil films and plays to capture culturally resonant expressions. Furthermore, Fernandes and Mannepalli (2021) constructed a dataset to support deep learning model development for Tamil SER.

Despite these contributions, existing datasets primarily focus on Indian Tamil and lack representation of the linguistic and cultural diversity found in the Sri Lankan Tamil community. Unique prosodic features, dialects, and emotional expressions specific to Sri Lankan Tamil speakers require tailored datasets for reliable SER applications. Inspired by the methodologies of EMOVO (Costantini et al.), EMODB (Burkhardt et al., 2005), and CaFE (Gournay et al., 2018), our dataset adopts an acted approach, ensuring distinct and reproducible emotional expressions while capturing the dialectal richness of Sri Lankan Tamil.

3 Dataset Development

3.1 Selection of Emotions and Actors

For the Tamil speech emotion dataset, we focused on five core emotions commonly recognized in speech emotion recognition: Anger, Happiness, Sadness, Fear, and Neutral. These emotions were selected based on their relevance in affective computing and their consistent presence in other emotional speech datasets. To ensure clarity and consistency in emotional expressions, detailed descriptions of each emotion were provided to the actors.

We adopted a discrete theory of emotions, specifically drawing from Ekmans classification (Ekman, 1992), which effectively captures the range of emotional expressions applicable to spoken language. Although some researchers (James, 1922; Lazarus, 1994) suggest additional emotions, such as love or hope, we opted for the more universally understood basic emotions that can be reliably conveyed through speech. This approach minimizes ambiguity in emotional expression, which is crucial for the effective training of machine learning models in speech emotion recognition.

The actors were selected through a recruitment process that included local advertisements and a pre-selection session. Approximately 40 native Tamil-speaking candidates, aged between 23 and 25 from diverse regions of Sri Lanka performed sample utterances in a controlled environment. A panel of three experienced drama teachers evaluated these recordings for emotional delivery and naturalness. From this pool, 22 actors were chosen, 11 males and 11 females, ensuring a balanced representation of the genders. All selected actors demonstrated exceptional ability to convey emotions and have at least a ’B’ grade in drama at the G.C.E (O/L) examination[2]. Figure 1 illustrates the geographic distribution of the actors, highlighting the richness of the data set in regional variation.

By carefully selecting emotions and actors from diverse regions and backgrounds, this dataset provides a robust resource for research on speech emotion recognition, offering valuable insights into emotional expression in different Tamil dialects.

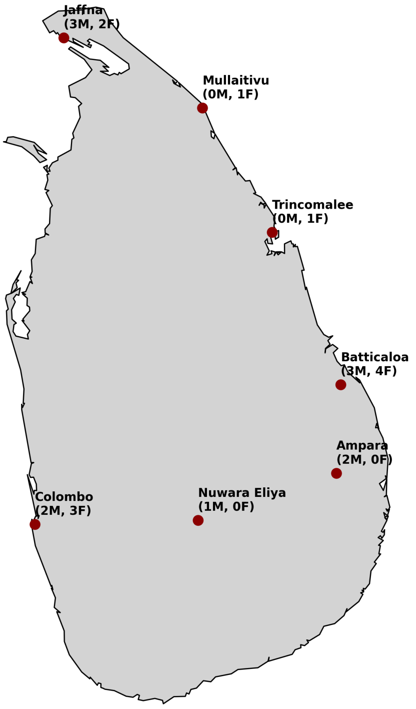


Figure 1: Regional breakdown of Actors.

3.2 Linguistic Material and Emotional Contexts

To develop the Tamil Speech Emotion Recognition (SER) dataset, EmoTa, we selected 19 semantically neutral sentences, allowing for a wide range of

[2] https://en.wikipedia.org/wiki/GCE_Ordinary_Level_in_Sri_Lanka

emotional expressions without introducing word-based biases. Inspired by datasets such as EMOVO (Costantini et al.) and EMODB (Burkhardt et al., 2005), we focused on everyday Tamil phrases that actors can easily recall and connect with emotionally. This approach encourages emotional delivery through vocal tone and prosody rather than specific words, helping SER models capture authentic vocal expressions.

Figure 2 provides sample sentences with their pronunciation and English translation. These sentences, used commonly in Tamil interactions, contain balanced phonetic elements to support detailed acoustic analysis. By keeping the language neutral, the dataset encourages actors to rely on vocal delivery for emotion expression.

Example of Emotional Adaptability: One versatile sentence in the dataset, "Nan unnai cantikka ventum." ("I need to meet you"), can represent multiple emotions depending on the context:

- **Sadness:** Used when conveying longing for a loved one who has been absent for a long time, this sentence is spoken with vulnerability and yearning.
- **Neutral:** In a professional setting, it simply requests a meeting with a colleague, delivered in a straightforward tone.
- **Happiness:** Spoken with excitement to share good news, such as a new job or engagement, this phrase becomes an expression of joy.
- **Fear:** When responding to an urgent situation concerning a friend, its spoken with anxiety and concern.
- **Anger:** In a tense setting, this phrase becomes a demand to meet in person to resolve a disagreement, with a tone of frustration and assertiveness.

These context-based scenarios help capture the nuances of vocal emotion, allowing SER models to focus on emotional prosody over lexical content. This method enables a deeper understanding of how emotions are expressed through voice alone, enhancing the effectiveness of SER models in real-world applications.

3.3 Recording Protocol and Data Curation

The recordings for the Tamil speech emotion dataset were conducted in a soundproof studio using professional equipment to ensure high-quality audio. Recordings were captured at a 48 kHz sampling rate and later downsampled to 16 kHz for compatibility across applications. Actors were given flexibility in their movements and expression, while being mindful of microphone placement to ensure consistent sound quality. Each one-hour recording session was supervised by a team of phoneticians, with two providing instructions and feedback and one managing equipment. Before each utterance, actors received prompts to avoid reading intonations and were provided emotional context to encourage authentic expression.

Actors recorded each sentence four to five times to capture subtle variations, with the best take selected based on acting quality and clarity of recording. Special emphasis was placed on avoiding exaggerated expressions, maintaining a natural and conversational style. However, challenges such as proximity variations and fluctuating intonation contours were addressed by adjusting recording levels and providing additional guidance.

To ensure efficient organization and retrieval, we adopted a systematic file naming convention:

<spkID>_<senID>_<emo>.wav

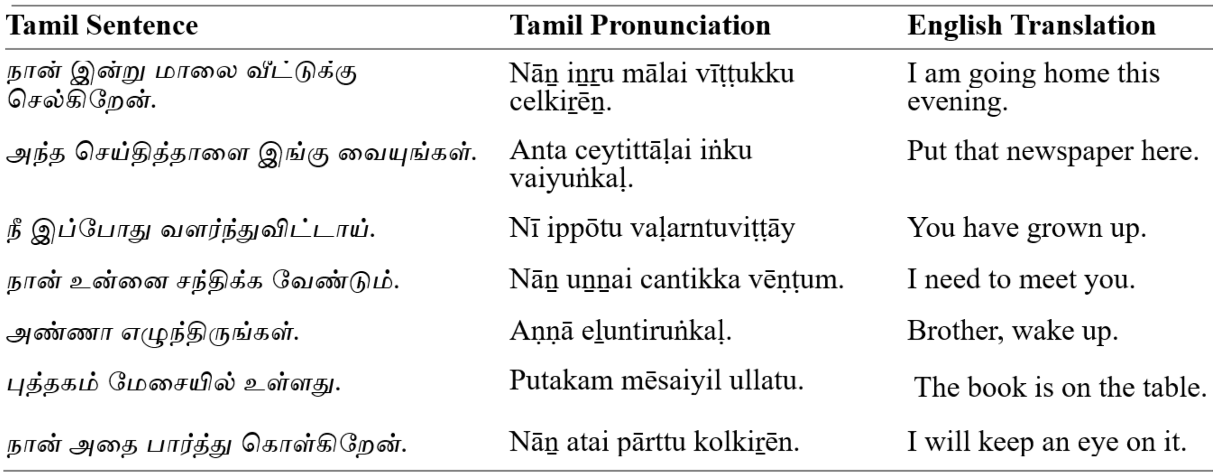
For example, the file name *01_02_ang.wav* indicates that this file corresponds to speaker ID 01, sentence ID 02, and an angry emotion.

The final dataset consists of 936 utterances from 22 actors, represents five different emotions happiness, sadness, anger, fear, and neutralitywith a total recording duration of approximately 48 minutes. Figure 3 provides a visual breakdown of the distribution of utterances across the various emotions. This distribution is key for ensuring that the dataset offers a balanced representation of emotional expressions, which is crucial for the development of emotion recognition models.

3.4 Inter-Annotator Agreement

To evaluate the reliability of our dataset, we assessed the inter-annotator agreement using Fleiss' kappa (Randolph, 2005) coefficient, a metric suitable for measuring agreement among multiple annotators. Inter-annotator agreement quantifies how well annotators consistently make the same annotation decisions for a particular category. This is essential to ensure the annotation process is consistent and that different annotators are assigning the same emotion label to a given sample. The kappa score is calculated as follows:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{1}$$



Tamil Sentence	Tamil Pronunciation	English Translation
நான் இன்று மாலை வீட்டுக்கு செல்கிறேன்.	Nāṉ iṉṟu mālai vīṭṭukku celkiṟēṉ.	I am going home this evening.
அந்த செய்தித்தாளை இங்கு வையுங்கள்.	Anta ceytittāḷai iṅku vaiyuṅkaḷ.	Put that newspaper here.
நீ இப்போது வளர்ந்துவிட்டாய்.	Nī ippōtu vaḷarntuviṭṭāy	You have grown up.
நான் உன்னை சந்திக்க வேண்டும்.	Nāṉ uṉṉai cantikka vēṇṭum.	I need to meet you.
அண்ணா எழுந்திருங்கள்.	Aṇṇā eḻuntiruṅkaḷ.	Brother, wake up.
புத்தகம் மேசையில் உள்ளது.	Putakam mēsaiyil ullatu.	The book is on the table.
நான் அதை பார்த்து கொள்கிறேன்.	Nāṉ atai pārttu kolkiṟēn.	I will keep an eye on it.

Figure 2: Sample Selected Sentences.

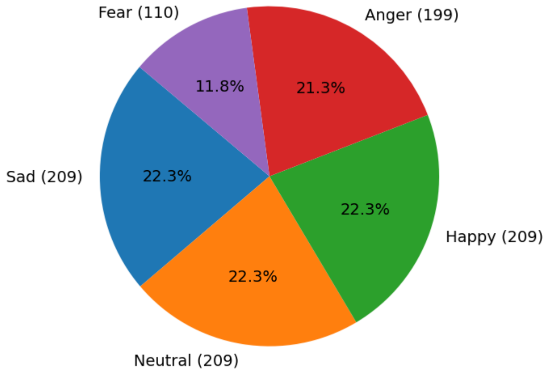


Figure 3: Distribution of utterances across emotions.

where p_0 represents the observed agreement, and p_e is the expected agreement by chance. The results of the inter-annotator agreement between our annotators yielded a substantial kappa score of 0.74, indicating high agreement across the five annotators. Notably, sadness and fear showed higher disagreements among annotators, contributing to the overall results.

4 Feature Extraction Techniques

In Speech Emotion Recognition (SER), selecting the right features is crucial, as speech signals contain various parameters that convey emotional information. Zero-Crossing Rate (ZCR), Chroma Features, Mel-Frequency Cepstral Coefficients (MFCC), Root Mean Square (RMS), and Mel Spectrogram are highly used for emotion classification (Ahmed et al., 2023). Each method offers unique insights into the emotional content of speech, which we will review briefly below.

4.1 Zero-Crossing Rate (ZCR)

Zero-Crossing Rate (ZCR) measures how often the signal changes sign, indicating the frequency of waveform zero crossings. It serves as an important indicator of speech dynamics and can help differentiate between voiced and unvoiced segments. Higher ZCR values typically indicate more dynamic speech, which may correspond to higher emotional intensity (Aouani and Ayed, 2020). Sample ZCR plots for each of the five emotions are shown in Figure 4.

4.2 Chroma Features

Chroma features capture the energy distribution across the 12 pitch classes of the chromatic scale. This technique is valuable for analyzing harmonic content in speech, as it can provide insights into the emotional expression related to musicality and intonation (Garg et al., 2020). Sample Chroma Features plots for each of the five emotions are shown in Figure 5.

4.3 Mel-Frequency Cepstral Coefficients (MFCC)

Mel-Frequency Cepstral Coefficients (MFCCs) are widely recognized as effective features for speech and audio analysis (Likitha et al., 2017). They are derived from the power spectrum of sound and are designed to reflect human auditory perception. MFCCs encapsulate the timbral properties of audio signals, making them essential for emotion recognition tasks, as they capture both spectral and temporal information in speech. Sample MFCC plots for each of the five emotions are shown in Figure 6.

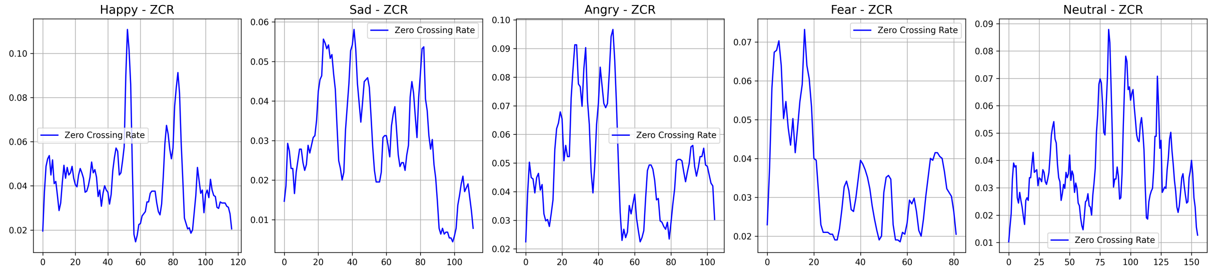


Figure 4: Zero Crossing Rate (ZCR) for various emotions: Happy, Sad, Angry, Fear, and Neutral. The ZCR measures the rate at which the signal changes from positive to negative.

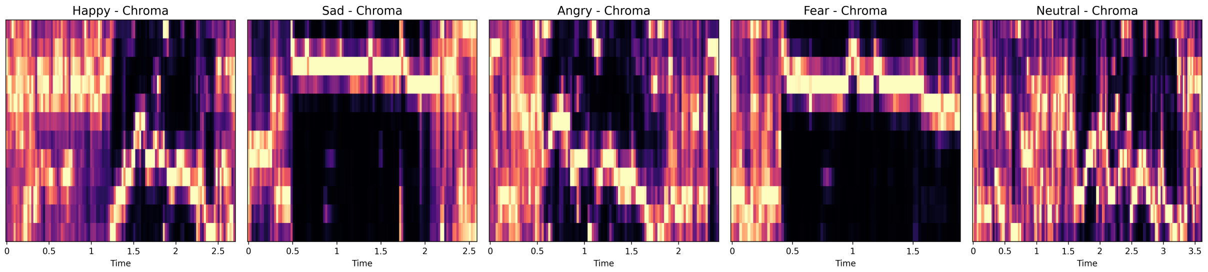


Figure 5: Chroma features for different emotions: Happy, Sad, Angry, Fear, and Neutral. Chroma features capture the energy distribution across 12 different pitch classes, providing insights into the harmonic content of the audio signals.

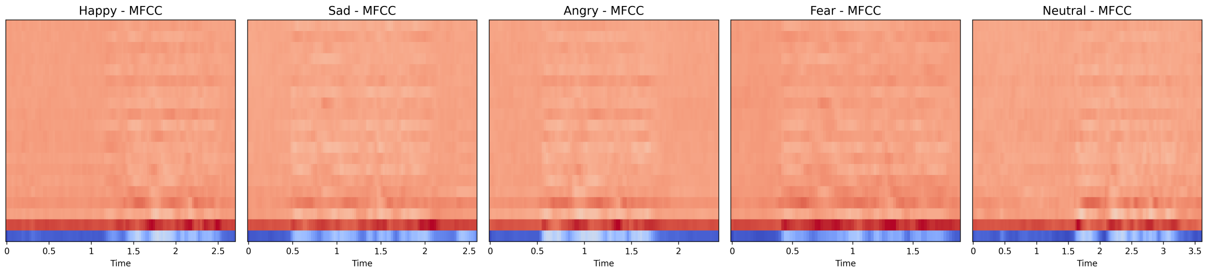


Figure 6: MFCCs for various emotions: Happy, Sad, Angry, Fear, and Neutral. MFCCs represent the short-term power spectrum of sound, widely used in speech and audio processing to capture the characteristics of human speech and music.

4.4 Root Mean Square (RMS)

The Root Mean Square (RMS) value quantifies the magnitude of a varying quantity in audio signals. In the context of speech, it measures loudness and energy. Higher RMS values are often associated with more intense emotions, while lower values may indicate calmer emotions. Thus, RMS is a crucial feature for differentiating emotional states based on speech amplitude. Sample RMS plots for each of the five emotions are shown in Figure 7.

4.5 Mel-Spectrogram

The Mel-Spectrogram combines the advantages of the Mel scale with spectrogram analysis, offering a visual representation of the frequency content of audio signals over time. This technique emphasizes perceptually relevant features, aligning with human auditory perception. By capturing the dynamic changes in sound, the Mel spectrogram enables effective modeling of emotional expression in speech, especially useful in deep learning applications (Venkataramanan and Rajamohan, 2019). Sample Mel-spectrogram plots for each of the five emotions are shown in Figure 8.

5 Experimental Design

5.1 Model Training and Evaluation

In this study, we evaluated various models for speech emotion classification. The dataset was divided into training (80%) and testing (20%) subsets.

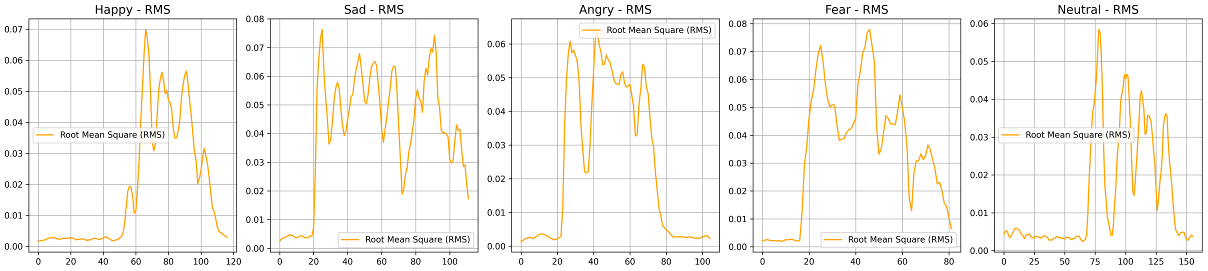


Figure 7: Root Mean Square (RMS) values for different emotions: Happy, Sad, Angry, Fear, and Neutral. The RMS value is a measure of the average power of the audio signal, indicating the loudness of the sound.

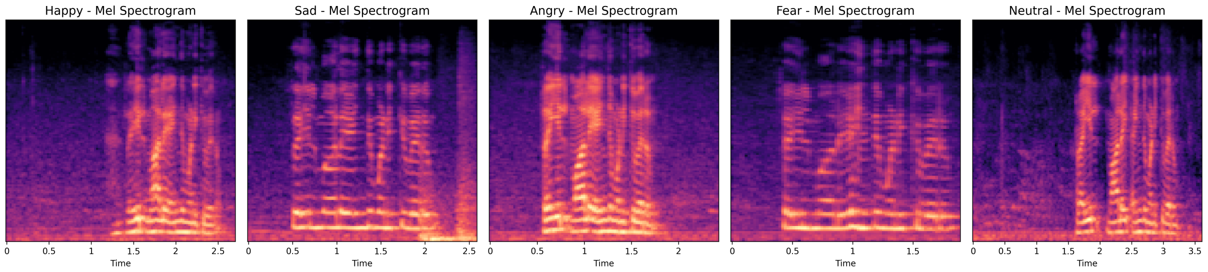


Figure 8: Mel Spectrograms for various emotions: Happy, Sad, Angry, Fear, and Neutral. The Mel spectrogram provides a time-frequency representation of the audio signal, where the frequency scale is non-linear and mimics human hearing.

Emotion	**P**	**R**	**F1**
angry	0.74	0.72	0.73
fear	0.88	0.85	0.86
happy	0.65	0.68	0.67
neutral	0.78	0.67	0.72
sad	0.60	0.69	0.64
macro avg	0.73	0.72	0.73

Table 1: Results of XGBoost without Data Augmentation.

Emotion	**P**	**R**	**F1**
angry	0.92	0.90	0.91
fear	0.93	0.94	0.93
happy	0.87	0.90	0.88
neutral	0.92	0.92	0.92
sad	0.91	0.90	0.90
macro avg	0.91	0.91	0.91

Table 2: Results of XGBoost with Data Augmentation.

The models were assessed using several evaluation metrics, including Macro average precision, recall, and F1-score. These metrics provide a comprehensive understanding of each model's performance on individual emotion classes as well as an overall classification assessment.

We selected several traditional models, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost. Additionally, we implemented a 1D CNN architecture, which captures important features from audio signals through the use of convolutional layers, making it well-suited for emotion classification tasks. To further enhance the model's capabilities, we employed a 1D CNN with an attention mechanism, which allows the model to focus on significant features within the audio input, thereby improving the identification of emotional cues.

5.2 Data Augmentation Techniques

To enhance the model's performance and increase the robustness of the speech emotion classification system, several data augmentation techniques were applied to the audio dataset. These techniques are designed to artificially expand the training data by introducing variations that simulate real-world conditions (Ahmed et al., 2023).

One of the methods employed was noise injec-

Model	**Non-Aug**			**Aug**		
	P	**R**	**F1**	**P**	**R**	**F1**
Logistic Regression	0.48	0.48	0.47	0.40	0.41	0.40
Decision Tree	0.39	0.37	0.38	0.60	0.60	0.60
Random Forest	0.70	0.71	0.70	0.90	0.90	0.90
Support Vector Machine	0.32	0.22	0.11	0.46	0.27	0.23
XGBoost	**0.73**	**0.72**	**0.73**	**0.91**	**0.91**	**0.91**
K-Nearest Neighbors	0.59	0.59	0.59	0.58	0.58	0.58
1D CNN	0.52	0.53	0.49	0.60	0.59	0.59
1D CNN (with Attention)	0.60	0.59	0.59	0.88	0.87	0.87

Table 3: Macro Average Precision, Recall and F1-Score on the test set for all the Models with and without augmentation.

tion, which adds random noise to the audio signals, helping the model differentiate emotional content in noisy environments. Time stretching altered the speed of the audio without changing its pitch, simulating variations in speech delivery. Random shifting involved slightly shifting the audio in time, reducing sensitivity to minor timing variations while pitch-shifting modified the pitch to expose the model to a broader range of vocal expressions.

Together, these methods increased the diversity and variability of the training dataset, enhancing the model's ability to recognize and classify emotions. These techniques significantly improved the model's performance, making it more resilient to variations in speech delivery, timing, and audio quality.

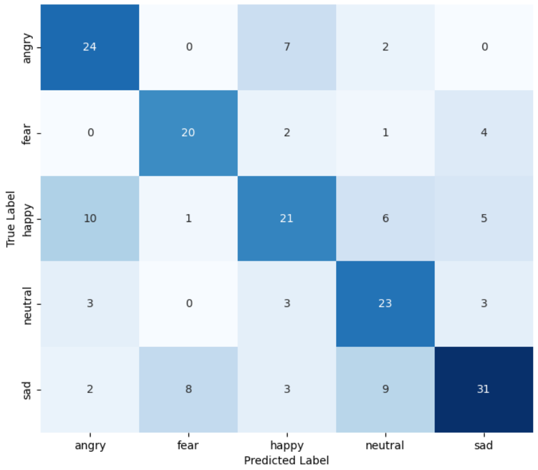


Figure 9: Confusion Matrix for XGBoost without Data Augmentation.

6 Results and Discussion

6.1 Model Performance Comparison

Table 3 summarizes the performance of various models on both the original (Non-Aug) and augmented (Aug) data. For Non-Augmented data, XGBoost achieves the highest F1-score of 0.73, followed closely by Random Forest with an F1-score of 0.70. On the augmented data, both XGBoost and Random Forest reach the high F1-score of 0.91 and 0.90 respectively, demonstrating their effectiveness after data augmentation. Among deep learning models, the 1D CNN with Attention performs notably well, achieving an F1-score of 0.87 on augmented data, suggesting that attention mechanisms enhance model accuracy. However, the Support Vector Machine (SVM) model performs the lowest on both Non-Augmented and Augmented data, with an F1-score of 0.11 and 0.23, respectively, indicating challenges in handling this task. The detailed results of XGBoost are presented in Table 1 for non-augmented data and Table 2 for augmented data.

6.2 Error Analysis and Model Limitations

The confusion matrix analysis (Figures 9 and 10) for XGBoost shows that data augmentation significantly enhances model performance. Without augmentation, notable misclassifications are observed, particularly happy being confused with angry and vice versa, as well as sad with fear and neutral. After augmentation, the model improves, with fewer misclassifications and an overall increase in prediction accuracy. However, confusion between an-

Province	Non-Aug			Aug		
	P	R	F1	P	R	F1
Nothern	0.55	0.55	0.54	0.98	0.98	0.98
Eastern	0.59	0.62	0.60	0.88	0.87	0.88
Western	0.65	0.66	0.65	0.93	0.89	0.90
Central	0.67	0.50	0.54	0.93	0.94	0.93

Table 4: Macro Average Precision, Recall and F1-Score on the test set for different dialects with and without augmentation using XGBoost.

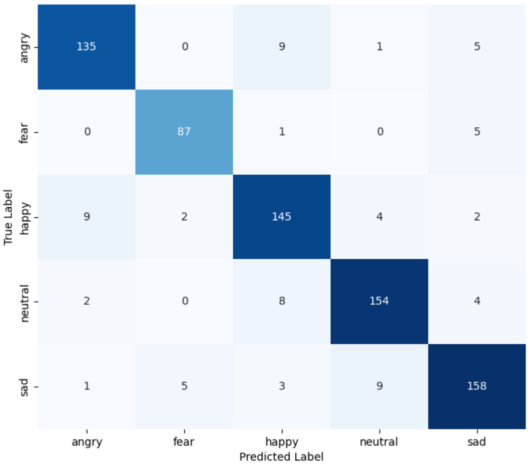


Figure 10: Confusion Matrix for XGBoost with Data Augmentation.

gry and happy persists, suggesting challenges in distinguishing emotions with overlapping acoustic features. This indicates that while augmentation is beneficial, further advancements in model architecture, feature extraction techniques, and possibly the inclusion of additional contextual or prosodic cues are needed to better distinguish emotions with similar acoustic characteristics.

7 Conclusion

This paper introduces the EmoTa dataset, specifically designed for Tamil speech emotion recognition, comprising 936 audio samples recorded from 22 native Tamil speakers. Each speaker conveys five distinct emotionsanger, happiness, sadness, fear, and neutralusing 19 semantically neutral sentences to eliminate semantic bias and focus purely on emotional delivery. The dataset provides a robust foundation for evaluating emotion recognition models, with results comprehensively reported in terms of precision, recall, and F1-Score to highlight various aspects of model performance. Furthermore, the dataset's reliability is assessed through inter-annotator agreement, quantified using Fleiss' kappa, ensuring consistency in emotional labeling. This resource aims to advance research in Tamil speech emotion recognition, addressing the scarcity of datasets in this domain.

8 Limitations of the work

The EmoTa dataset, created from emotional speech samples by actors aged 23 to 25, could benefit from an expanded age range to increase its generalizability across diverse age groups. This would make the dataset more suitable for broader applications in Tamil speech emotion recognition. Currently, EmoTa includes 48 minutes of recorded speech, which serves as a foundation for preliminary research. However, a larger volume of samples would enhance its robustness, supporting more in-depth analysis. The dataset presently covers five emotions: happy, sad, angry, neutral, and fear. Adding more emotions would improve the datasets utility and enable greater accuracy in recognizing emotional nuances in Tamil speech.

References

Md Rayhan Ahmed, Salekul Islam, AKM Muzahidul Islam, and Swakkhar Shatabda. 2023. An ensemble 1d-cnn-lstm-gru model with data augmentation for speech emotion recognition. *Expert Systems with Applications*, 218:119633.

Hadhami Aouani and Yassine Ben Ayed. 2020. Speech emotion recognition with deep learning. *Procedia Computer Science*, 176:251–260.

Felix Burkhardt, A. Paeschke, M. Rolfes, Walter F. Sendlmeier, and Benjamin Weiss. 2005. A database of German emotional speech. In *Interspeech 2005*, pages 1517–1520. ISCA.

Giovanni Costantini, Iacopo Iadarola, Andrea Paoloni, and Massimiliano Todisco. EMOVO Corpus: an Italian Emotional Speech Database.

Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. 2001. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80.

Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200.

Bennilo Fernandes and Kasiprasad Mannepalli. 2021. An analysis of emotional speech recognition for tamil language using deep learning gate recurrent unit. *Pertanika Journal of Science & Technology*, 29(3).

Utkarsh Garg, Sachin Agarwal, Shubham Gupta, Ravi Dutt, and Dinesh Singh. 2020. Prediction of emotions from the audio speech signals using mfcc, mel and chroma. In *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, pages 87–91.

Philippe Gournay, Olivier Lahaie, and Roch Lefebvre. 2018. A canadian french emotional speech dataset. In *Proceedings of the 9th ACM Multimedia Systems Conference*, pages 399–402, Amsterdam Netherlands. ACM.

William James. 1922. The emotions.

Richard S Lazarus. 1994. *Passion and reason: Making sense of our emotions*. Oxford University Press.

M. S. Likitha, Sri Raksha R. Gupta, K. Hasitha, and A. Upendra Raju. 2017. Speech based human emotion recognition using mfcc. In *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pages 2257–2260.

Rajeev Rajan, Haritha U.G., Sujitha A.C., and Rejisha T. M. 2019. Design and Development of a Multi-Lingual Speech Corpora (TaMaR-EmoDB) for Emotion Analysis. In *Interspeech 2019*, pages 3267–3271. ISCA.

C Sunitha Ram and R Ponnusamy. 2014. An effective automatic speech emotion recognition for tamil language using support vector machine. In *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, pages 19–23. IEEE.

Justus J Randolph. 2005. Free-marginal multirater kappa (multirater k [free]): An alternative to fleiss' fixed-marginal multirater kappa. *Online submission*.

P Vasuki, B Sambavi, and Vijesh Joe. 2020. Construction and evaluation of tamil speech emotion corpus. *National Academy Science Letters*, 43(6):533–536.

Kannan Venkataramanan and Haresh Rengaraj Rajamohan. 2019. Emotion recognition from speech. *arXiv preprint arXiv:1912.10458*.