

PAPERS TITLE

MULTI-LABEL EXTREME LEARNING MACHINE (MLELMS) FOR BANGLA REGIONAL SPEECH RECOGNITION

Authors

Prommy Sultana Hossain, Amitabha Chakrabarty, Kyuheon Kim, Md. Jalil Piran

Published in

Applied Science , 2022

Presented By
Naznin Jahan Noor

OVERVIEW

Introduction

Objective

Initial Hypothesis

Major Contribution

Literature Review

Dataset

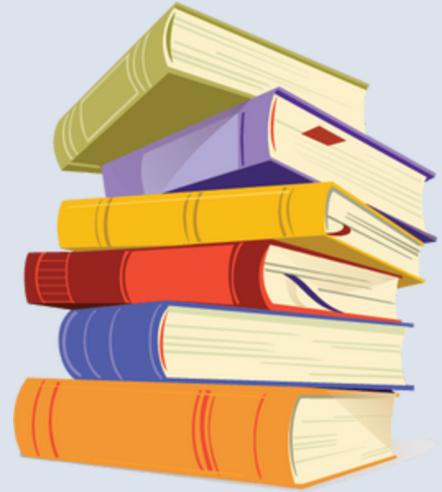
Methodology

Result

Future Scope

Conclusion

INTRODUCTION



Bangla Language Importance

- Bangla is the **fifth most spoken language** globally, with a rich variety of dialects across Bangladesh.

Research Gap

- Most current Bangla speech recognition research focuses on **content, age, and gender**, with little attention to regional dialect classification or synthetic speech detection.

Purpose of Study

- This study fills these gaps by enabling the classification of Bangla dialects and the detection of synthetic (computer-generated) speech.

OBJECTIVE

Develop a model to recognize and classify regional dialects within Bangla speech.

Distinguish between synthesized and authentic Bangla speech.

Improve classification accuracy by considering age as a factor in dialect recognition.

INITIAL HYPOTHESIS

Synthetic Speech
Detection

Regional Dialect
Recognition

Impact of Age



MAJOR CONTRIBUTION

New Dataset for Bangla Dialects

30 hours of Bangla speech covering seven regional dialects.

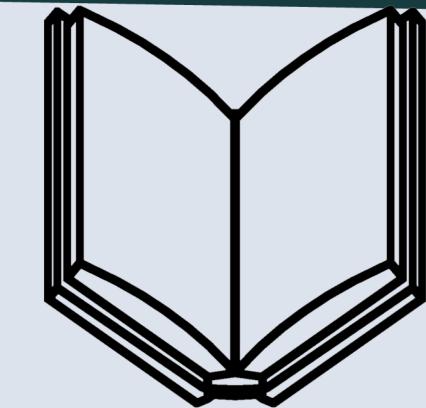
Innovative Model Architecture

Combines Stacked Convolutional Autoencoder (SCAE) and Multi Level Extreme Learning Machine (MLELM) for dialect and authenticity detection.

Broad Applicability

Effective on English speech datasets, showing adaptability beyond Bangla

LITERATURE REVIEW



Existing Bangla Speech Recognition

- Previous studies mainly focused on identifying age, gender, and content in Bangla speech.
- Limited work on recognizing regional dialects or synthetic speech detection.

Previous Models

- CNN and MLP Approaches: Deep CNN and multi-layer perceptron (MLP) models have shown promise in general audio classification but lack multi-label capabilities for dialect and authenticity classification.
- Limitations: These models do not incorporate age-related features, reducing accuracy in dialect recognition.

| District | Language |
|------------|-----------------------------------|
| Khulna | অ্যাক জন মানশির দুটো ছাওয়াল ছিল। |
| Bogra | য্যাক ঝনের দুটা ব্যাটা ছিল |
| Rangpur | একজন ম্যানশের দুইকনা ব্যাটা আছিল |
| Sylhet | এক মানুশৰ দুই পুয়া আছিল |
| Chittagong | এগুয়া মানশের দুয়া পোয়া আছিল |
| Noyakhali | একজনের দুই ছত আছিল |
| Mymensingh | য্যাক জনের দুই পুৎ আছিল |

Table 1: Representation of English sentence “A man has two sons” in the seven regional Bangla languages

DATASET

- **Total duration:** 30 hours of Bangla speech
- **Unique words:** 85,500
- **Total Sentences:** 100,057
- **Recording Quality:** 22 KHz



METHODOLOGY

1
Canva

Preprocessing the audio data by removing noise, trimming silent sections, and padding to a fixed length of 10 seconds.

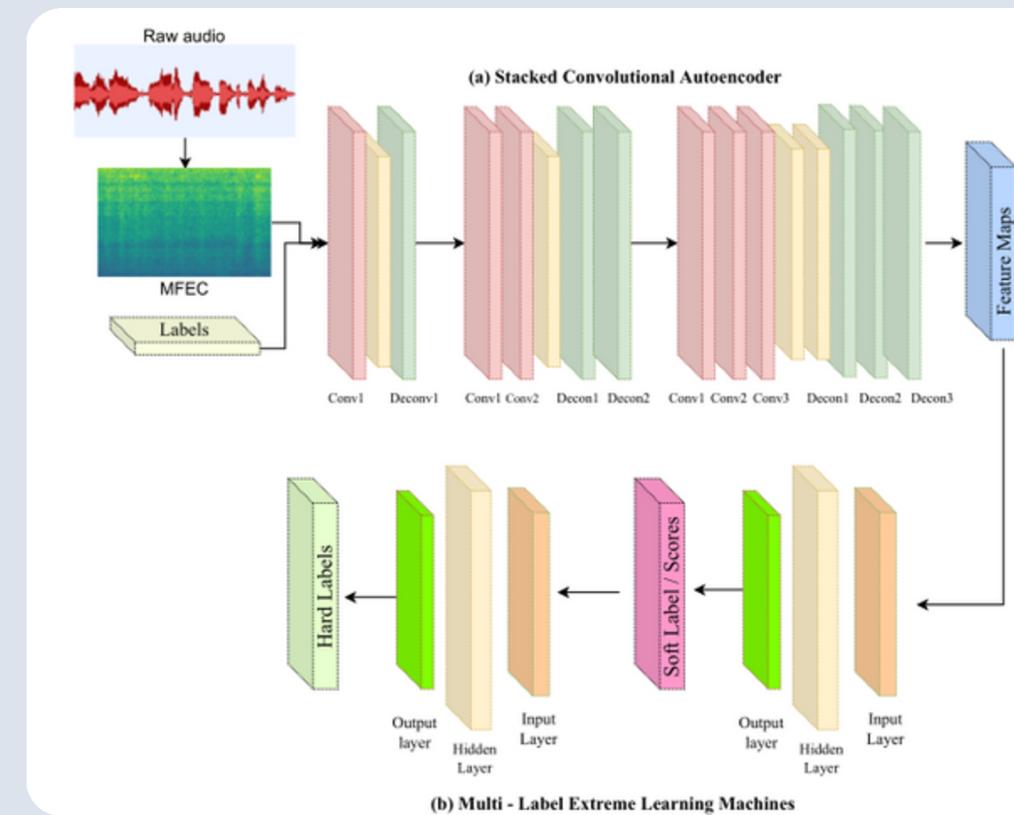
2
Canva

Extracting **Mel-Frequency Energy Coefficients (MFECs)** from the audio data using 65ms frames with 50% overlap, and segmenting the mel-spectrogram into 35 columns with a hop size of 100ms

3
Canva

3
Canva

Using a **Multilabel Extreme Learning Machine (MLELM)** framework for soft classification and score approximation.



Using a **Stacked Convolutional Autoencoder (SCAE)** to extract features from the MFEC input data.

RESULT

| Speech Type | Precision(P) | Recall (R) | F1-Score(FS) | Accuracy |
|-------------|--------------|------------|--------------|----------|
| Bangla | | | | 93% |
| Original | 90 | 94 | 93 | |
| Synthesized | 89 | 91 | 90 | |
| English | | | | 96% |
| Original | 94 | 95 | 94 | |
| Synthesized | 93 | 97 | 93 | |

Table 1: Classification results of Dialect and Age correlation for Bangla and English speech

Continuous

| Speech type | Accuracy |
|-------------|----------|
| Bangla | 0.91 |
| English | 0.82 |

Table 2: Results of Age classification

| Speech type | Accuracy |
|-------------|----------|
| Bangla | 0.92 |
| English | 0.96 |

Table 3: Results of Gender classification

FUTURE SCOPE

Dataset Expansion

The need to expand the Bangla speech dataset is highlighted to address limitations in regional dialect representation and phonetic diversity.

Cross-Language Adaptation

The model's potential to work with other languages, such as English, is mentioned, along with its demonstrated success in these domains.

Application in ASR and TTS

The paper discusses the relevance of the proposed method to **Automatic Speech Recognition (ASR)** and **Text-to-Speech (TTS)** systems.

Model Refinements

Efficiency and robust feature extraction using SCAE-MLELM for complex speech classification tasks are highlighted.

CONCLUSION

- *The model successfully improves dialect classification and synthetic speech detection for Bangla.*
- *Versatile model adaptable to multiple languages, including English.*
- *Opens doors to a multilingual, inclusive future of speech tech.*



THANK YOU
