Signal Pattern Analysis for Bengali Alphabets Audio Applying Machine Learning Approaches

Jubaida Quader Jerin*, MD. Akhtaruzzaman†

Department of Computer Science and Engineering, Military Institute of Science and Technology, Dhaka-1216, Bangladesh Email: *jqjerin@gmail.com,†akhter900@gmail.com

Abstract— This research focuses on signal pattern analysis for Bengali alphabets in audio, utilizing established machine learning approaches. It outlines the inherent linguistic challenges, provides insights derived from the collected data, and discusses benchmark results. The primary focus is on incorporating lan-guage processing for Bengali raw audio files, employing advanced signal pattern analysis for efficient feature extraction, and adapt-ing machine learning models for Bengali alphabet recognition. Specific tailored techniques, such as windowing and Overlap-add, were applied to address Bengali vowels. Feature extraction encompasses methodologies spanning Root Mean Square Energy, Zero Crossing Rate, and Mel-frequency cepstral coefficients (MFCCs). In experimental settings, MFCCs consistently demon-strate superior performance compared to alternative methods. Various machine learning models, including Linear Regression, MLP Classifier, SVM, and LSTM, were employed, with MFCCs consistently exhibiting enhanced performance in Bengali alphabet recognition. Future endeavors aim to advance automatic speech recognition for Bengali spoken alphabets, intending seamless integration into embedded systems like Arduino for practical applications focusing on raw audio data. Additionally, the study explores ensemble learning techniques for Bangla phoneme identification, aiming to enhance the robustness and accuracy of classification systems.

Index Terms—Signal pattern analysis, Speech Recognition, Machine Learning, Linear Regression, SVM, MLP Classifier, LSTM, Raw Bengali Phenoms

I. Introduction

Signal pattern analysis could refer to the examination of pat-terns within a signal, which could be any measurable quantity changing over time. This analysis often involves extracting meaningful information, detecting trends, or identifying spe-cific features within the signal[1]. Techniques such as Fourier analysis, wavelet analysis, and machine learning algorithms may be employed for signal pattern analysis. On a similar inspect, Speech recognition is the technology that enables a machine to interpret and understand spoken language[2]. It involves the conversion of spoken words into written text. This technology is crucial for various applications, including voice-activated systems, transcription services, and human-computer interaction.

Bengali described as an intonation language without tone, accent, or stress, exhibits a monotone speech with subtle pitch and loudness variations. Similar to written language sequences, spoken Bangla comprises elementary acoustic sounds called phonemes[3].In this research work, we studied with Bengali alphabet audio or phonemes. It is among the highly spoken languages globally, with over 300 million individuals using it as their primary language. It holds the position of being the second most spoken language in the Indian subcontinent[4].

In the course of our research, we have undertaken a com-prehensive exploration of signal pattern analysis for Bengali alphabets in audio using established machine learning method-ologies. The primary goals of our study encompassed imple-menting language processing techniques specific to Bengali raw audio files, employing advanced signal pattern analysis for effective feature extraction, adapting machine learning models to recognize Bengali alphabets, and making mean-ingful contributions to the evolving field of audio recognition research. Our methodological approach involved the utilization of techniques like windowing and Overlap-add, particularly tailored for Bengali vowels.

For the crucial task of feature extraction, we adopted a di-verse set of methodologies, spanning Time Domain, Frequency Domain, Spectrogram analysis, Amplitude Envelope, Root Mean Square Energy, Zero Crossing Rate, and Mel-frequency cepstral coefficients (MFCCs). Each of these techniques offers unique insights into the characteristics of the audio signal. Amplitude Envelope, for instance, focuses on the smoother amplitude variations of the signal, prioritizing overall ampli-tude characteristics over rapid changes. Root Mean Square Energy provides a measure of the signal's overall energy, while Zero Crossing Rate captures temporal characteristics, contributing to a holistic understanding of the audio patterns. However, for this research study, we only introduced Root Mean Square Energy, Zero Crossing Rate, and Mel-frequency cepstral coefficients (MFCCs).

Through rigorous experimentation, we found that Mel-frequency cepstral coefficients (MFCCs) stood out as par-ticularly effective, delivering more accurate results compared to other methods. In the realm of machine learning models, we employed four distinct models, namely Linear Regression, MLP Classifier, SVM, and LSTM. Notably, the MFCCs con-sistently demonstrated superior performance in these models, underscoring their efficacy in the context of Bengali alphabet recognition. In addition, in our experiment, we found that SVM and MLP Classifiers serve solid consistent results re-garding which feature extraction we implement.

Looking ahead, our future endeavors will be geared towards the development of automatic speech recognition tailored for Bengali spoken alphabets. The ultimate goal is to create a system that can seamlessly integrate into embedded systems, such as Arduino, opening up avenues for practical applications and real-world implementations of our findings, in addition, for this particular embedded system development, we worked with raw 132 audio files collected from native individuals.

The paper is organized as follows. In section II, the related work of this research is discussed briefly. In section III, a detailed study methodology of Study I is discussed. The findings of Study I are briefly presented in Section IV. The study methodology and outcomes of Study II are discussed in section V. The limitations, future work, and a brief concluding remark are presented in the final section.

II. Related Works

This literature review provides an overview of relevant studies and developments in audio signal analysis and machine learning applied to language-specific contexts. Speech recogni-tion and signal analysis have been fundamental research areas in the human-computer interaction domain. This technology holds significance in areas with limited English proficiency, emphasizing the importance of regional language recognition for seamless human-computer interaction[5].

Speech recognition technology has advanced significantly, with a notable focus on the Bangla alphabet. In one approach, a system is presented with a small database, utilizing MFCC for feature extraction. Word recognition is achieved through DTW and K-NN classification, resulting in a remarkable 10 percent performance boost. This methodology showcases the effectiveness of combining MFCC, DTW, and K-NN for Bangla alphabet recognition[3].

In contrast, the introduction of the Bengali Common Voice Speech Dataset[6] stands out, emphasizing size and diversity. This paper navigates through linguistic challenges, details dataset characteristics, analyzes feature diversity, and presents preliminary benchmarks. Its comprehensive exploration con-tributes valuable insights to Bangla speech recognition sys-tems, offering a foundational understanding for future devel-opments.

Another exploration addresses challenges in speech pro-duction, underscoring human language's unique features. It dives into Bengali alphabet recognition, assessing various methods, including Hidden Markov Models (HMM) and fuzzy logic[7]. Recognizing the limitations of traditional speech recognition in noisy environments, the paper proposes using surface electromyogram (sEMG) signals for improved com-munication. The proposed method, focusing on classifying voiceless Bangla vowels using sEMG signals, presents a novel approach that could benefit individuals who have undergone laryngectomy surgeries.

The integration of modern technology into the education system for visually impaired individuals is a critical aspect. Existing works focus on text-to-Braille conversion, leaving a gap in voice-to-Braille conversion. The proposed system utilizes Convolutional Neural Network (CNN) models for voice recognition, achieving high accuracy and aiming to em-power visually impaired individuals in their basic education. This approach is a step towards a more inclusive educational environment[4] [8].

An innovative feature, Mel Frequency Wavelet Transform Coefficient (MFWTC), combining MFCC and DWT, proves effective in identifying Bengali letters from uttered sounds[9]. This approach demonstrates a practical solution for recog-nizing Bengali letters, particularly Swarabarna and Banjan-barna[10].

Moving to classification, a system accurately classifies 39 Bangla alphabets using audio features, showcasing the poten-tial of Multilayer Perceptron Classifier (MLPC) and support vector machine (SVM). This technology is a promising step towards efficient Bangla alphabet recognition[11].

The importance of regional language-based speech recog-nition is highlighted in a study focusing on Bengali isolated spoken numerals. The utilization of Gaussian Mixture Model (GMM) and Mel Frequency Cepstral Coefficients (MFCC) attains a high correct prediction rate, emphasizing the signif-icance of regional language recognition for effective human-computer interaction[5].

In a domain with limited progress compared to other lan-guages, a study adopts a long short-term memory (LSTM) approach for Bengali speech recognition[12]. The study's fo-cus on individual Bengali words, utilizing distinctive features such as mel-frequency cepstral coefficients (MFCC), attains promising results. The system achieves a word detection error rate of 13.2 percent and a phoneme detection error rate of 28.7 percent on the Bangla-Real-Number audio dataset, show-casing the potential of LSTM in Bengali speech recognition. This approach holds the potential to improve Bengali speech recognition systems.

Addressing real-life conditions, a study introduces the Bangla short speech commands dataset, collected in a real-life setting[13]. Three convolutional neural network (CNN) architectures, including one with raw audio files, present varying degrees of accuracy. The study emphasizes proficiency in single-syllable word recognition and identifies challenges in multi-syllable command recognition.

The research emphasizes the limited implementation of the Bengali language in speech recognition for Human-Computer Interaction. Employing convolutional neural network (CNN) and recurrent neural network (RNN) techniques, the study aims to contribute to the development of Bengali language processing in the field of speech recognition[14].

Finally, the need for simplifying user interfaces through speech-based interaction is highlighted. A Bangla Phoneme recognition system is proposed as a foundational step for a comprehensive Bangla Speech recognition system. The sys-tem utilizes Line Spectral Frequency-Grade (LSF-G) features and an ensemble learning-based approach, achieving high accuracy. Tested on a Bangla Swarabarna Phoneme dataset consisting of 3290 clips, the system achieves an accuracy of 94.01 percent, showcasing the effectiveness of an ensemble learning-based approach in Bangla Phoneme recognition[15].

III. Methodology

The methodology for this research work comprises the following steps:

- Data Acquisition:

  Data was sourced from open-access platforms such as GitHub and native participant contributions.
- Data Preparation:

  The collected data underwent a meticulous formatting process to ensure uniformity and noise-reduction tech-niques were applied to enhance the quality of the audio files.
- Audio Feature Extraction:

  A pivotal phase involved the extraction of audio features. Various methods were explored, including Mel-frequency cepstral coefficients (MFCCs), Zero Crossing Rate, and Root Mean Square Energy.
- Model Selection and Implementation:

  In this stage, four distinct categories of supervised learn-ing models were chosen for experimentation: Linear Regression, Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Multilayer Perceptron (MLP) Classifier.
- Performance Evaluation:

  The accuracy and predictive capabilities of the models were assessed using the Scikit-Learn library. This com-prehensive evaluation aimed to validate the efficiency of each model in the given context.

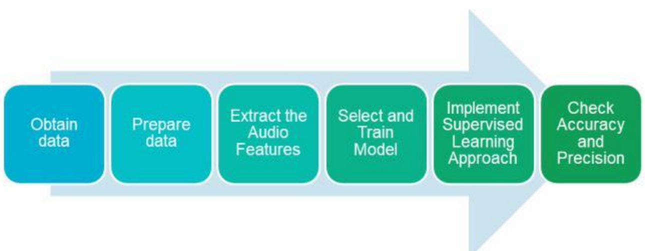

Fig. 1: Methodology of Signal Pattern Analysis Process

This methodological framework ensures a systematic ap-proach to data collection, preparation, feature extraction, model selection, implementation, and rigorous assessment of model performance. The utilization of diverse features and a range of models contribute to a thorough exploration of the research objectives.

A. Data Collection Approach

In our data collection approach, we primarily centered our efforts on open sources. Despite the uniformity of phonemes, there were marginal differences in the audio files. Subse-quently, we secured 132 raw audio files from four participants, each providing their informed consent for this research. The participant group comprised three males and one female, all within the age range of 26 to 29 years, with none reporting any medical conditions. In our dataset of 132 audio clips only 11 audio clips were the female participant's voice.

B. Data Description and Processing

The acquired audio dataset encompassed files in .aac, .mp3, .ogg and .wav formats. To maintain consistency for the experiment, all data underwent conversion to .wav format. However, it's noteworthy that this conversion process resulted in some audio quality loss, particularly noticeable in Bengali vowels 7th and 11th. In addition, the experiment was done by incorporating two types of file naming conventions and labeling. The first set of 132 audio files dataset was labeled in Bengali and the second set of 132 audio files dataset was labeled in English.

To mitigate noise, the Windowing and Overlap-Add (OLA) method was applied to the audio files. Despite these efforts, discernible improvements in the overall audio quality were limited. Importantly, it is worth mentioning that the audio files were recorded using simple Android phones, underscoring the practicality of the data collection setup, especially environ-mental noise.

C. Machine Learning Model Architecture

Algorithm 1 LSTM Model

Architecture Input:

Require: Sequence Length (L), Feature Dimension (D)

Output:

Ensure: LSTM model with the specified architecture

1: Initialize a Sequential model.
2: Add an LSTM layer:
   - Number of units: 256
   - Return sequences: False
   - Input shape: (L, D)
   - Activation function: Tanh
3: Add a Dropout layer with a dropout rate of 0.2.
4: Add a Dense layer (Hidden Layer 1):
   - Number of units: 128
   - Activation function: ReLU
5: Add another Dropout layer with a dropout rate of 0.2.
6: Add another Dense layer (Hidden Layer 2):
   - Number of units: 64
   - Activation function: ReLU
7: Add one more Dropout layer with a dropout rate of 0.2.
8: Add the Output Dense layer:
   - Number of units: 11
   - Activation function: Softmax

Algorithm 2 Linear Regression Model

Architecture Input Layer:

1: Neurons: Number of features (input variables)
2: Activation Function: None (or linear activation)

Output Layer:

1: Neurons: 1 (for predicting a continuous value)
2: Activation Function: None (linear activation)

Architecture Summary:

1: Input (Features) → Output (Prediction)

Algorithm 3 MLP Classifier Model

Architecture Input Layer:

1: Neurons: Number of features (input variables)
2: Activation Function: Usually ReLU (Rectified Linear Unit)

Hidden Layers:

1: Layer 1:
   - Neurons: 128
   - Activation Function: ReLU
2: Layer 2:
   - Neurons: 64
   - Activation Function: ReLU

Output Layer:

1: Neurons: Number of classes (depends on the problem)
2: Activation Function: Usually Softmax for classification

Model Configuration:

1: Hidden Layer Sizes: (128, 64)
2: Maximum Iterations: 50
3: Random State: 42

Algorithm 4 SVM Model

Architecture Model Configuration:

1: Kernel: Linear
2: Decision Function Shape: One-vs-Rest (OvR)
3: Random State: 42

Training:

1: Fit the model:
   - Input Data: $X_{train\ scaled}$ (scaled training features)
   - Target Labels: $y_{train\ encoded}$ (encoded training labels)

IV. Result Analysis

In this section, we discuss the findings of this experiment, including their evaluation score in Fig 2, and all the experiment results are showcased in the respective subsection in the section VI for easier readability.

- LSTM Model:
  - Loss (MFCC): 1.4228
  - Accuracy (MFCC): 0.5143

  The LSTM model has a relatively low accuracy on both the training and validation sets. This indicates that the model may not be effectively capturing patterns of our Bengali audio data.
- Linear Regression Model::
  - Score (RMS & Chroma): -0.19447130010631564

  The linear regression model has a negative score. In the context of regression, a negative score might suggest that the model is not fitting the data well. It's important to consider that Linear Regression inappropriate approach for Natural Language Processing.
- MLP Classifier Model:
  - Train Accuracy (MFCC): 0.9428571428571428
  - Test Accuracy (MFCC): 0.48148148148148145

  The MLP model achieves high accuracy on the training set (approximately 94 percent), but the test accuracy is relatively low (48 percent). This performance gap suggests the possibility of overfitting, where the model performs well on training data, however struggles to generalize to new, unseen data.
- SVM Model::
  - Train Accuracy (MFCC): 1.0
  - Test Accuracy (MFCC): 0.4444444444444444

  The SVM model has perfect accuracy on the training set, but the test accuracy is moderate (44 percent). This indicates that the model may have learned the training data well but faces challenges in generalizing to new instances.

V. Discussion and Conclusion

SVM is powerful for classification tasks, MLP is a ver-satile neural network for complex patterns, Linear Regres-sion predicts continuous values, and LSTM is specialized for sequences and time-dependent patterns. In this research, each model brought unique advantages and faced specific challenges. The LSTM model, designed for sequential data and capturing temporal dependencies, encountered difficulties in identifying sequential patterns within audio signals in this study. Conversely, linear regression models assume a simple linear relationship between input features and output. The negative score obtained suggests a poor fit to the data. In contrast, the MLP Classifier Model exhibited the capability to learn complex relationships within the data. However, the model's high training accuracy and comparatively lower test accuracy signal potential overfitting. The SVM Model, effective in both linear and nonlinear classification tasks, displayed a higher training accuracy compared to the test accuracy, indicating a degree of overfitting. Nevertheless, it still demonstrated reasonable performance.

A. Limitations

The observed variations in performance could stem from several contributing factors, including data complexity, data size, feature extraction techniques, model hyperparameters, file naming conventions, raw audio data, environmental noise, and other inherent attributes. In this study, a dataset comprising 132 raw Bengali vowel phoneme audio files was utilized

Category	Extraction Feature	Evaluation Metrics			
		LSTM	MLP Classifier	SVM	Linear Regession
Bengali Font Label	MFCC with windowing	loss: 1.4228 - accuracy: 0.5143 - val_loss: 5.6218 - val_accuracy: 0.0370	Train Accuracy: 0.9238095238095239 Test Accuracy: 0.4074074074074074	Train Accuracy: 1.00 Test Accuracy: 0.37037037037037035	-
	MFCC without windowing	loss: 1.4571 - accuracy: 0.4381 - val_loss: 5.5575 - val_accuracy: 0.0370	Train Accuracy: 0.9238095238095239 Test Accuracy: 0.4074074074074074	Train Accuracy: 1.00 Test Accuracy: 0.37037037037037035	-
	Zero crossing rate	loss: 2.3588 - accuracy: 0.1810 - val_loss: 2.6367 - val_accuracy: 0.0000e+00	Train Accuracy: 0.14285714285714285 Test Accuracy: 0.037037037037037035	Train Accuracy: 0.3333333333333333 Test Accuracy: 0.037037037037037035	-
	root mean square energy	loss: 2.3491 - accuracy: 0.1619 - val_loss: 2.7200 - val_accuracy: 0.0000e+00	Train Accuracy: 0.1619047619047619 Test Accuracy: 0.037037037037037035	Train Accuracy: 0.11428571428571428 Test Accuracy: 0.037037037037037035	-
	RMS and Chroma with widowing	-	-	-	score: -0.03116173130980493
	RMS and Chroma without widowing	-	-	-	score: -0.08119102821270796
English Font Label	MFCC with windowing	loss: 1.3580 - accuracy: 0.4952 - val_loss: 6.0053 - val_accuracy: 0.0000e+00	Train Accuracy: 0.9428571428571428 Test Accuracy: 0.48148148148148145	Train Accuracy : 1.00 Test Accuracy: 0.4444444444444444	-
	MFCC without windowing	loss: 1.4744 - accuracy: 0.5143 - val_loss: 5.5183 - val_accuracy: 0.0000e+00	Train Accuracy: 0.9428571428571428 Test Accuracy: 0.48148148148148145	Train Accuracy History: 1.00 Test Accuracy History: 0.4444444444444444	-
	Zero crossing rate	loss: 2.3617 - accuracy: 0.2095 - val_loss: 2.6894 - val_accuracy: 0.0000e+00	Train Accuracy: 0.1619047619047619 Test Accuracy: 0.07407407407407407	Train Accuracy History:0.3333333333333333 Test Accuracy History: 0.07407407407407407	-
	root mean square energy	loss: 2.3634 - accuracy: 0.1714 - val_loss: 2.7087 - val_accuracy: 0.0000e+00	Train Accuracy: 0.18095238095238095 Test Accuracy: 0.07407407407407407	Train Accuracy: 0.13333333333333333 Test Accuracy: 0.037037037037037035	-
	RMS and Chroma with widowing	-	-	-	score: -0.19447130010631564
	RMS and Chroma without widowing	-	-	-	score: -0.027339880722507592

Fig. 2: Evaluation Metrics

including their environmental noise. It's noteworthy that the linguistic characteristics of Bengali, encompassing a variety of monophthongs, diphthongs, and triphthongs, might play distinct roles compared to other languages. The dataset, con-sisting of only 132 audio files, represents a limited sample size. A more extensive dataset could potentially enhance training results and provide a more comprehensive understanding of the Bengali phonetic nuances. The choice of model, its com-patibility with sequential data, the intricacy of the task, and the effectiveness of feature extraction methods all contribute to the observed performance differences. In conclusion, conducting experiments with various models and fine-tuning their parame-ters is crucial to achieving optimal results for the specific task of Bengali audio analysis. Exploring a broader spectrum of linguistic features and increasing the dataset size may further enrich the insights gained from the analysis.

B. Future Work

The planned future work intends to close the gap between experimental insights and real-world applications by expanding and improving the current study. The goal of this research is to make significant contributions to the field of audio recognition research, with a focus on Bengali spoken alphabets, by using a multifaceted approach that includes feature extraction, model refinement, dataset expansion, real-world integration, user interface considerations, cross-linguistic comparisons, and collaborative contributions.

References

[1] A. V. Oppenheim, A. S. Willsky, S. H. Nawab, and J.-J. Ding, Signals and systems. Prentice hall Upper Saddle River, NJ, 1997, vol. 2.

[2] L. R. Rabiner and B.-H. Juang, "Speech recognition: Statistical methods," Encyclopedia of language & linguistics, pp. 1–18, 2006.

[3] A. Sayem, "Speech analysis for alphabets in bangla language: automatic speech recognition," 2014.

[4] M. A. Hussain, R. I. Rifat, S. B. Iqbal, S. Biswas, M. G. R. Alam, and M. T. Reza, "Deep learning based bangla voice to braille character conversion system," in 2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE, 2022, pp. 0262–0267.

[5] B. Paul, S. Bera, R. Paul, and S. Phadikar, "Bengali spoken numerals recognition by mfcc and gmm technique," in Advances in Electronics, Communication and Computing: Select Proceedings of ETAEERE 2020. Springer, 2021, pp. 85–96.

[6] S. Alam, A. Sushmit, Z. Abdullah, S. Nakkhatra, M. Ansary, S. M. Hossen, S. M. Mehnaz, T. Reasat, and A. I. Humayun, "Bengali common voice speech dataset for automatic speech recognition," arXiv preprint arXiv:2206.14053, 2022.

[7] S. Mostafa, M. Awal, M. Ahmad, and M. Rashid, "Voiceless bangla vowel recognition using semg signal," SpringerPlus, vol. 5, no. 1, pp. 1–15, 2016.

[8] F. Faisal, M. Hasan, S. Sabrin, M. Z. Hasan, and A. H. Siddique, "Voice activated portable braille with audio feedback," in 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). IEEE, 2021, pp. 418–423.

[9] A. Gupta and K. Sarkar, "Recognition of spoken bengali numerals using mlp, svm, rf based models with pca based feature summarization," Int. Arab J. Inf. Technol., vol. 15, pp. 263–269, 2018.

[10] S. Phadikar, P. Das, I. Bhakta, A. Roy, S. Midya, and K. Majumder, "Bengali phonetics identification using wavelet based signal feature," in Computational Intelligence, Communications, and Business Analytics: First International Conference, CICBA 2017, Kolkata, India, March 24– 25, 2017, Revised Selected Papers, Part I. Springer, 2017, pp. 253–265.

[11] M. G. Hussain, M. Rahman, B. Sultana, A. Khatun, and S. Al Hasan, "Classification of bangla alphabets phoneme based on audio features using mlpc & svm," in 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI). IEEE, 2021, pp. 1–5.

[12] M. M. H. Nahid, B. Purkaystha, and M. S. Islam, "Bengali speech recog-nition: A double layered lstm-rnn approach," in 2017 20th international conference of computer and information technology (ICCIT). IEEE, 2017, pp. 1–6.

[13] S. A. Sumon, J. Chowdhury, S. Debnath, N. Mohammed, and S. Momen, "Bangla short speech commands recognition using convolutional neural networks," in 2018 international conference on bangla speech and language processing (ICBSLP). IEEE, 2018, pp. 1–6.

[14] J. Islam, M. Mubassira, M. R. Islam, and A. K. Das, "A speech recognition system for bengali language using recurrent neural network," in 2019 IEEE 4th international conference on computer and communi-cation systems (ICCCS). IEEE, 2019, pp. 73–76.

[15] A. Dutta and S. Chaudhury, "Bengali alpha-numeric character recognition using curvature features," Pattern Recognition, vol. 26, no. 12, pp. 1757–1770, 1993. [Online]. Available: https://www.sciencedirect.com/science/article/pii/003132039390174U

VI. Appendix

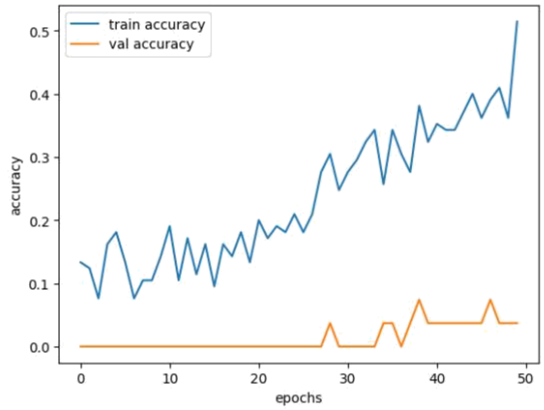

Fig. 3: Result Plot of Bengali Audio Files Labeled in Bengali LSTM Accuracy - MFCC Feature Extraction with Windowing Effect

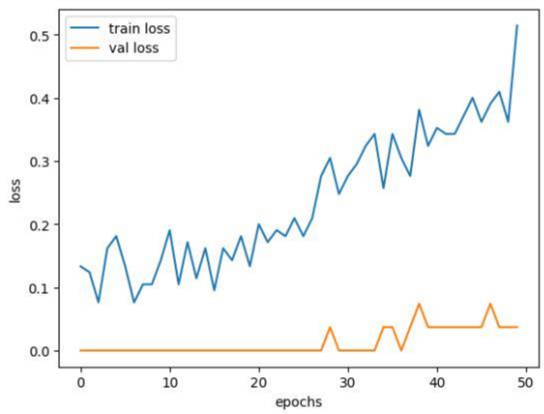

Fig. 4: Result Plot of Bengali Audio Files Labeled in Bengali LSTM Loss - MFCC Feature Extraction with Windowing Effect

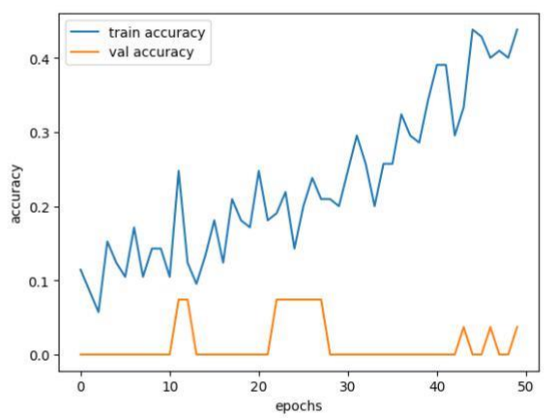

Fig. 5: Result Plot of Bengali Audio Files Labeled in Bengali LSTM Accuracy - MFCC Feature Extraction without Windowing Effect

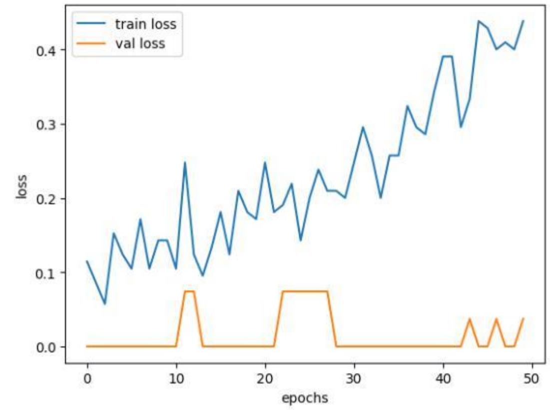

Fig. 6: Result Plot of Bengali Audio Files Labeled in Bengali LSTM Loss - MFCC Feature Extraction without Windowing Effect

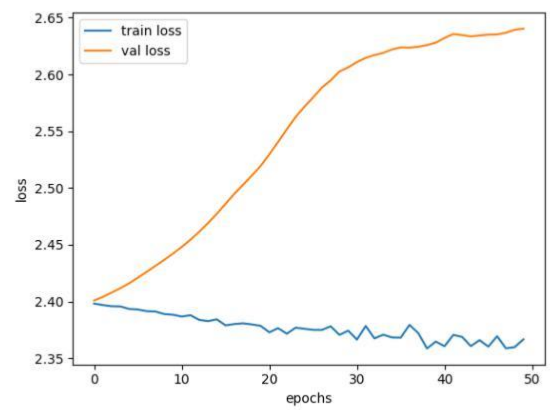

Fig. 7: Result Plot of Bengali Audio Files Labeled in Bengali LSTM Loss - Zero Crossing Rate Feature Extraction

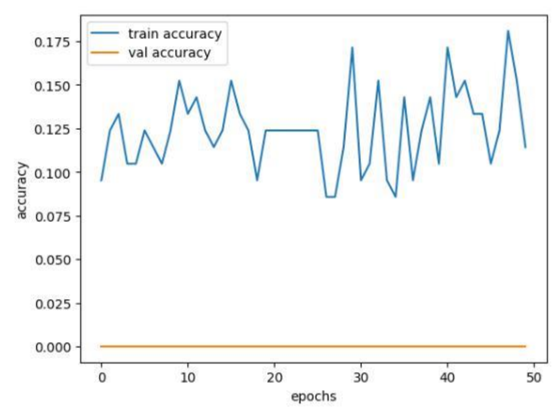

Fig. 8: Result Plot of Bengali Audio Files Labeled in Bengali LSTM Accuracy - Zero Crossing Rate Feature Extraction

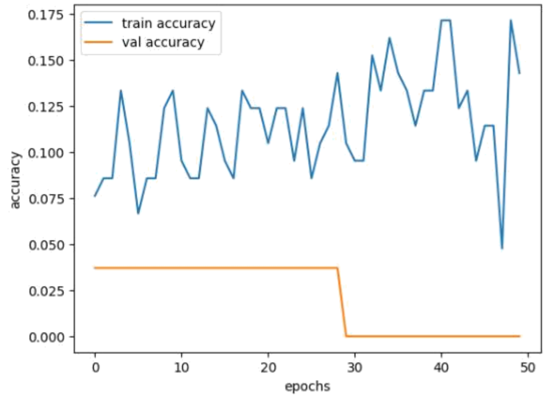

Fig. 9: Result Plot of Bengali Audio Files Labeled in Bengali LSTM Accuracy - Root Mean Square Energy Feature Extrac-tion

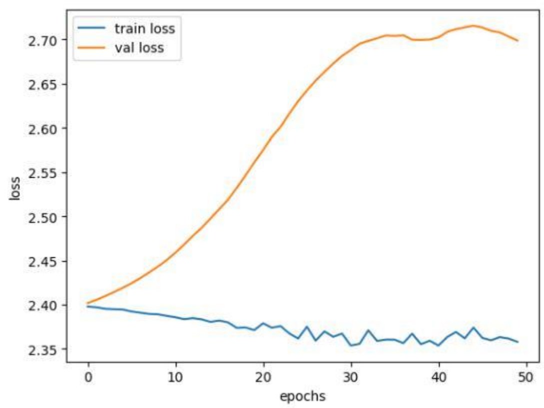

Fig. 10: Result Plot of Bengali Audio Files Labeled in Bengali LSTM Loss - Root Mean Square Energy Feature Extraction

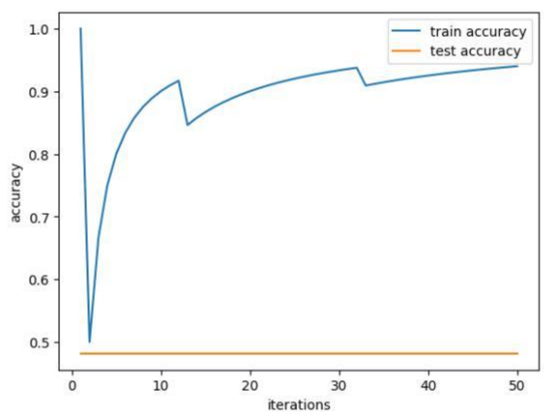

Fig. 11: Result Plot of Bengali Audio Files Labeled in Bengali MLP Classifier - MFCC Feature with Windowing Effect

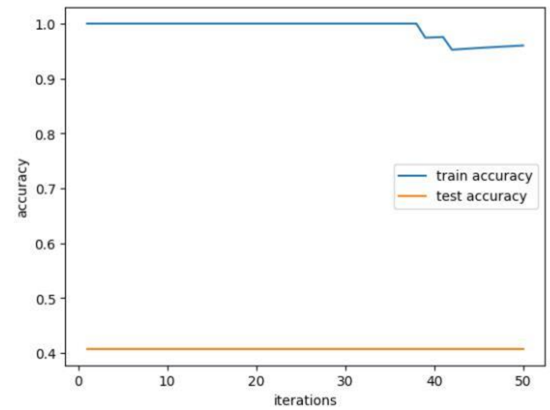

Fig. 12: Result Plot of Bengali Audio Files Labeled in Bengali MLP Classifier - MFCC Feature without Windowing Effect

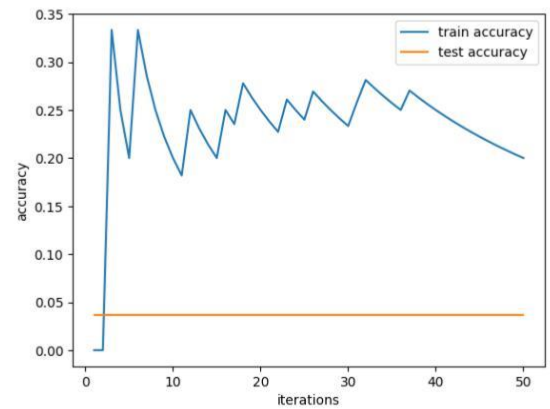

Fig. 13: Result Plot of Bengali Audio Files Labeled in Bengali MLP Classifier - Zero Crossing Rate Extraction Feture

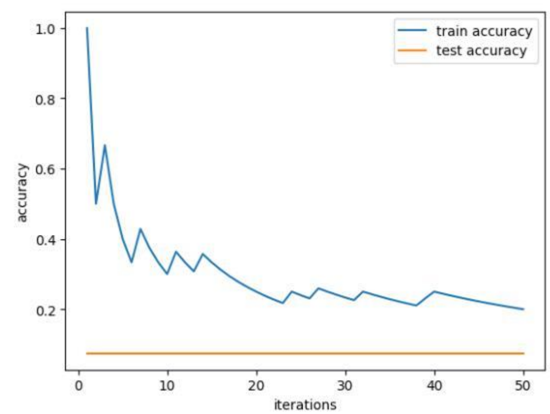

Fig. 14: Result Plot of Bengali Audio Files Labeled in Bengali MLP Classifier - Root Mean Square Energy Extraction Feature

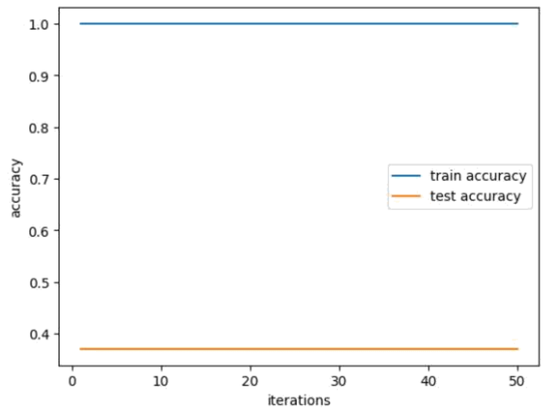

Fig. 15: Result Plot of Bengali Audio Files Labeled in Bengali SVM - MFCC Feature Extraction with Windowing Effects

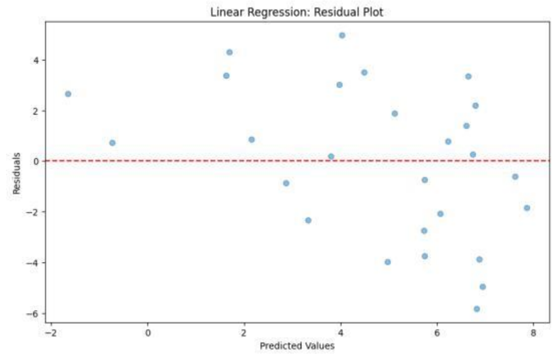

Fig. 18: Result Plot of Bengali Audio Files Labeled in Bengali Linear Regression - Chroma Feature Extraction with Window-ing Feature

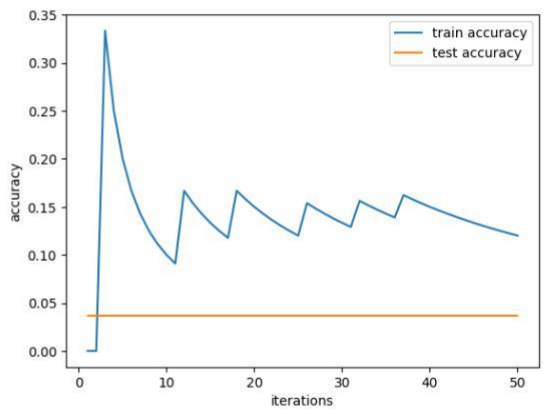

Fig. 16: Result Plot of Bengali Audio Files Labeled in Bengali SVM - Root Mean Square Energy Feature Extraction

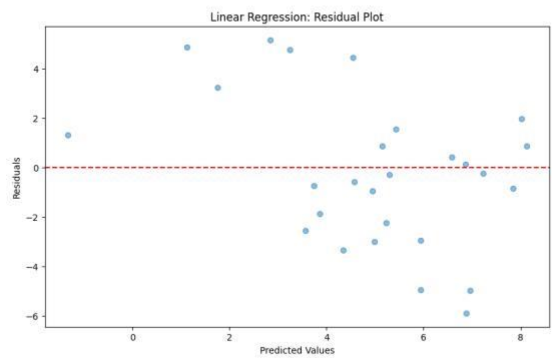

Fig. 19: Result Plot of Bengali Audio Files Labeled in Ben-gali Linear Regression - Chroma Feature Extraction without Windowing Feature

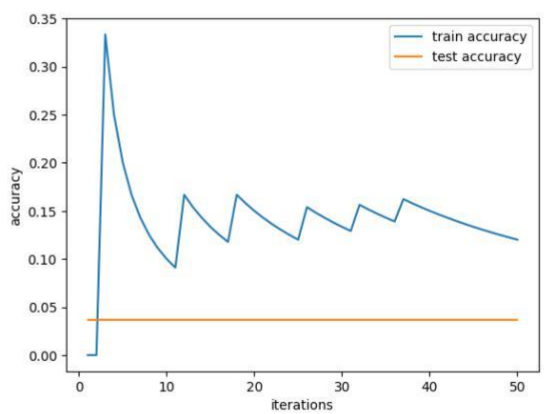

Fig. 17: Result Plot of Bengali Audio Files Labeled in Bengali SVM - Zero Crossing Rate Feature Extraction

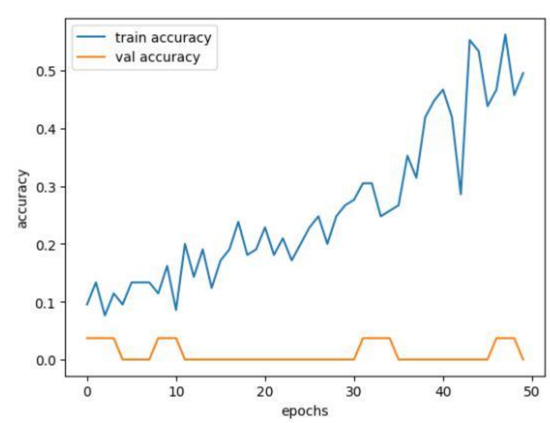

Fig. 20: Result Plot of Bengali Audio Files Labeled in English LSTM Accuracy - MFCC Feature Extraction with Windowing Effects

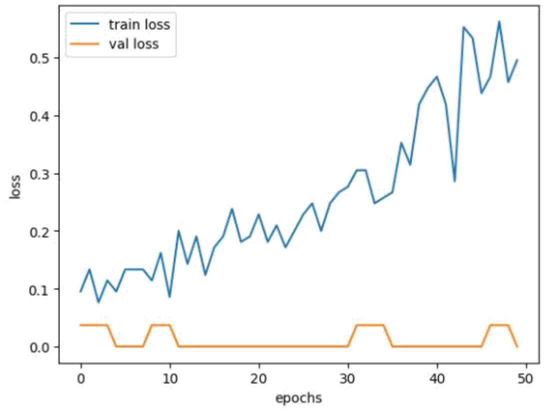

Fig. 21: Result Plot of Bengali Audio Files Labeled in English LSTM Loss - MFCC Feature Extraction with Windowing Effects

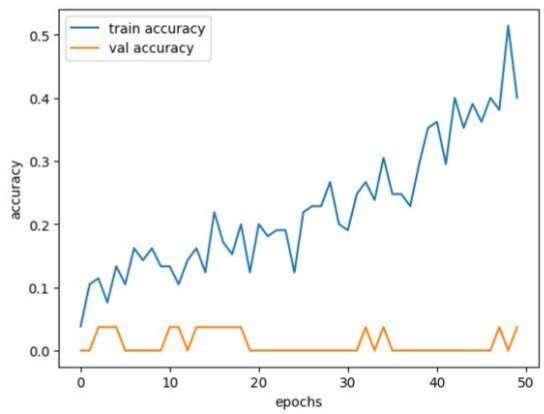

Fig. 22: Result Plot of Bengali Audio Files Labeled in English LSTM Accuracy - MFCC Feature Extraction without Window-ing Effects

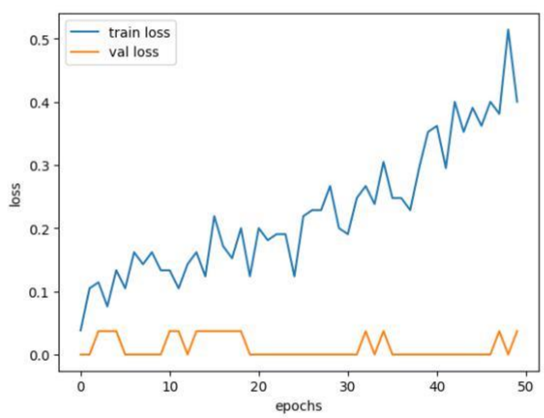

Fig. 23: Result Plot of Bengali Audio Files Labeled in English LSTM Loss - MFCC Feature Extraction without Windowing Effects

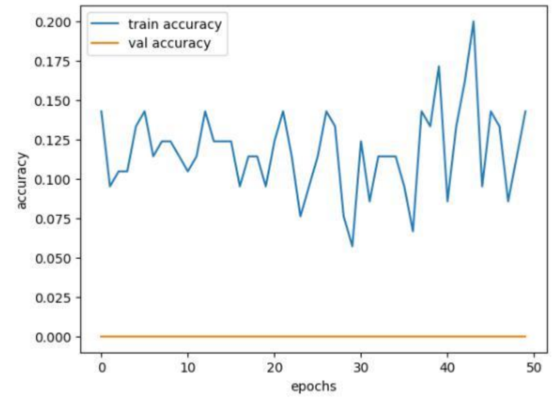

Fig. 24: Result Plot of Bengali Audio Files Labeled in English LSTM Accuracy - Zero Crossing Rate Feature Extraction

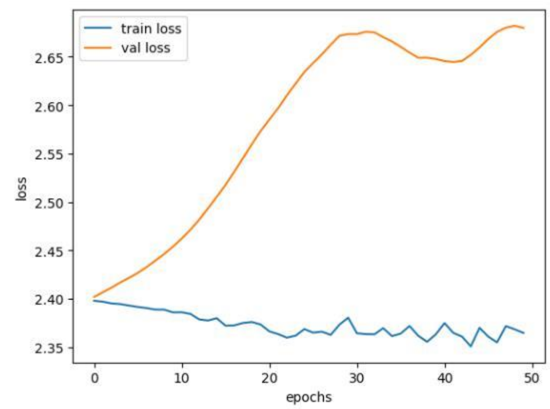

Fig. 25: Result Plot of Bengali Audio Files Labeled in English LSTM Loss - Zero Crossing Rate Feature Extraction

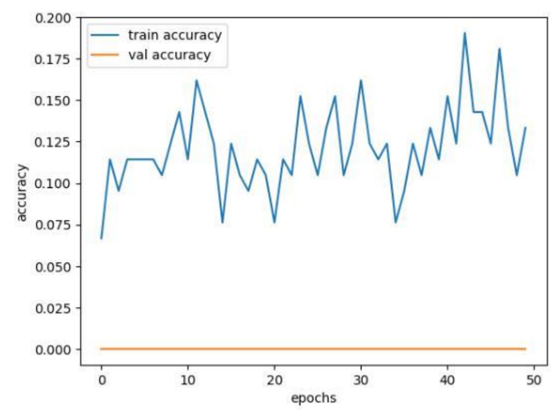

Fig. 26: Result Plot of Bengali Audio Files Labeled in English LSTM Accuracy - Root Mean Square Energy Feature Extrac-tion

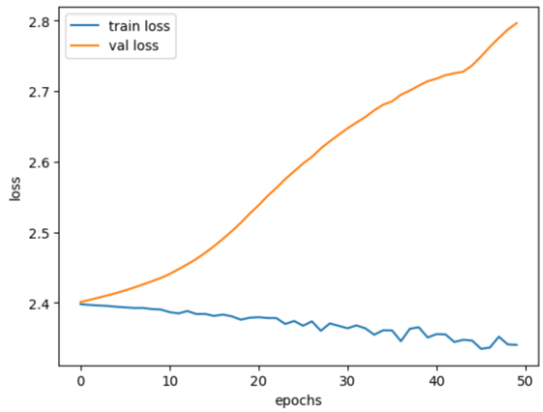

Fig. 27: Result Plot of Bengali Audio Files Labeled in English LSTM Loss - Root Mean Square Energy Feature Extraction

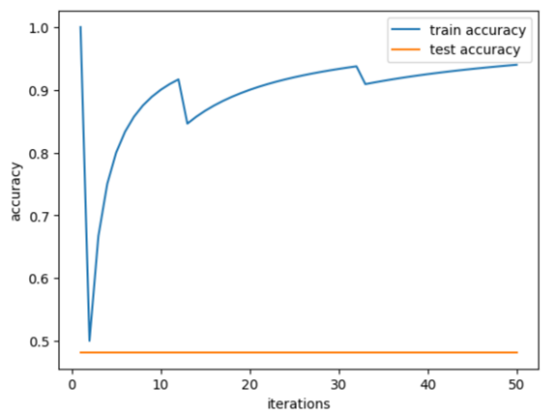

Fig. 28: Result Plot of Bengali Audio Files Labeled in English English
MLP Classifier - MFCC Feature with Windowing Effect

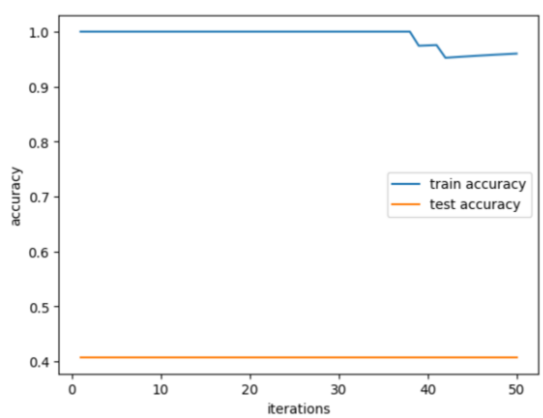

Fig. 29: Result Plot of Bengali Audio Files Labeled in English
MLP Classifier - MFCC Feature without Windowing Effect

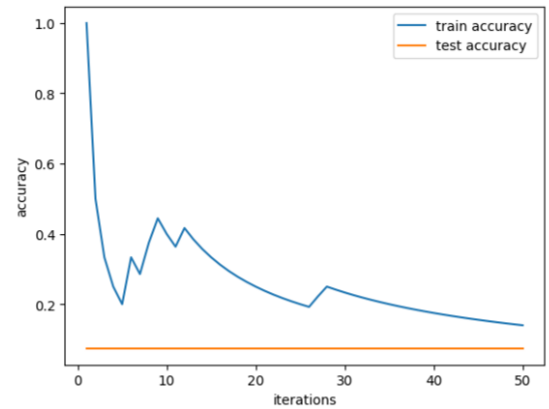

Fig. 30: Result Plot of Bengali Audio Files Labeled in English MLP Classifier - Zero Crossing Rate Extraction Feture

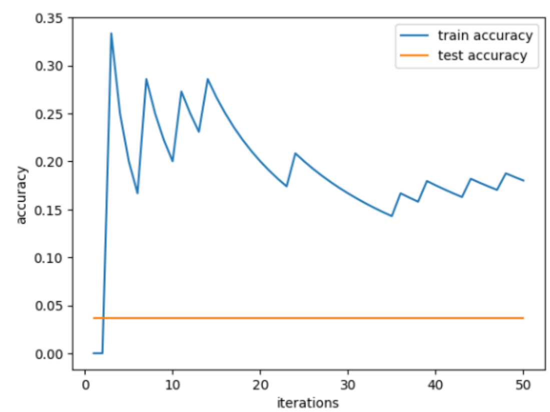

Fig. 31: Result Plot of Bengali Audio Files Labeled in
MLP Classifier - Root Mean Square Energy Extraction Feature

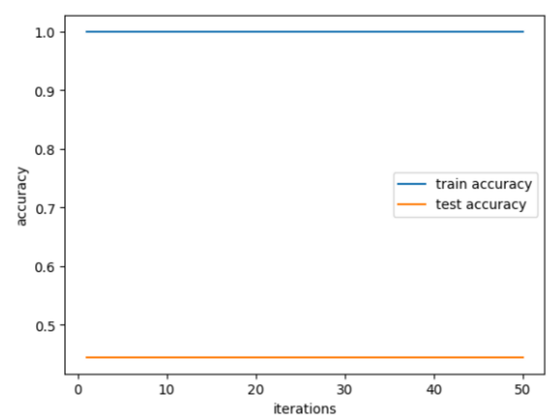

Fig. 32: Result Plot of Bengali Audio Files Labeled in English
SVM - MFCC Feature Extraction with Windowing Effects

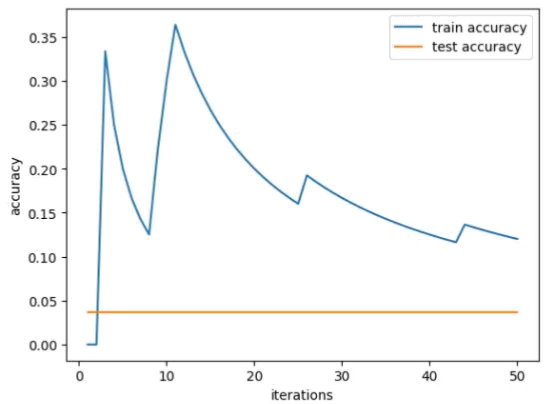

Fig. 33: Result Plot of Bengali Audio Files Labeled in English SVM - Root Mean Square Energy Feature Extraction

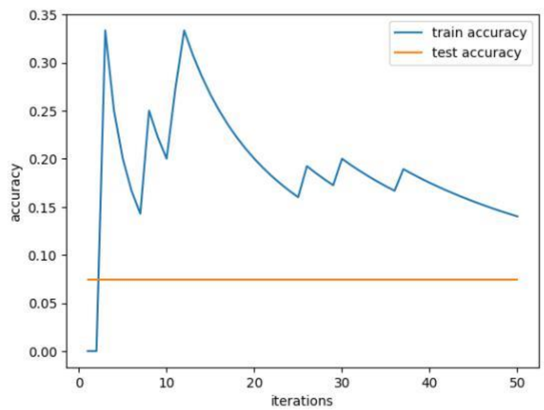

Fig. 34: Result Plot of Bengali Audio Files Labeled in English SVM - Zero Crossing Rate Feature Extraction

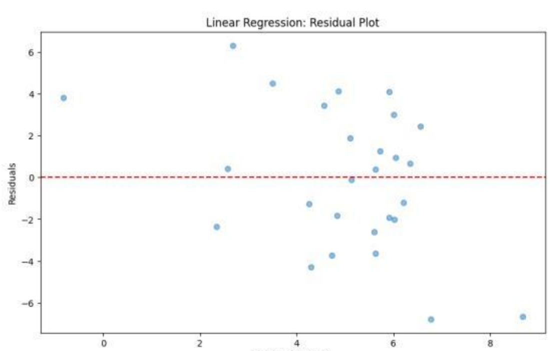

Fig. 35: Result Plot of Bengali Audio Files Labeled in English Linear Regression - Chroma Feature Extraction with Window-ing Feature