Bangla Speaker Accent Variation Classification from Audio Using Deep Neural Networks: A Distinct Approach

Khorshed Alam
Computer Science and Engineering
United International University
Dhaka, Bangladesh
mohdkhurshed120@gmail.com

Mahbubul Haq Bhuiyan
Department of CSE
Independent University, Bangladesh
Dhaka, Bangladesh
mh_bhuiyan.sets@iub.edu.bd

Md Fahad Monir
Department of CSE
Independent University, Bangladesh
Dhaka, Bangladesh
fahad.monir@iub.edu.bd

***Abstract*— Accent Variation Classification is the technique of detecting an accent or dialect of a human speech based on speech patterns and features from speech. This is useful in developing speech recognition systems, language learning systems, dialect preservation systems, sociolinguistic studies, voice assistance, improving speech synthesis and voiceover systems. It can be used in conducting forensic analysis on audio data to determine regional origin or specific accent traits. Furthermore, it is a useful tool in criminal investigations and judicial actions. Deep Neural Networks (DNNs) are utilized for speech recognition tasks because they can successfully learn complex variables of speech input such as patterns, intensity, rhythm, and temporal information. In this study, we propose Zero Crossing Rate (ZCR), Mel Frequency Cepstral Coefficients (MFCC), Root Mean Square (RMS), Mel-Spectrogram based feature extraction and DNN based Bangla Speaker Accent Variation Classification model to classify the speaker's variation from Bangla Speech data. We train our model with 7443 audios from 9303 audios (Formal, Dhaka, Khulna, Barisal, Rajshahi, Sylhet, Chittagong, Mymensingh and Noakhali) and our model achieves 94% accuracy from unseen or new data. We compare its accuracy and performance with other neural networks where LSTM, Stacked LSTM and DCNN achieve accuracy of 67%, 71% and 85% respectively.**

Keywords—Accent classification, Speech recognition, Filtering, Deep neural network, Human Voice, Accent variation classification.

I. Introduction

Accents may reveal a lot about a person's history, including their original language, location of origin, and ethnicity [1]. Recognizing different types of accents can help improve the quality of voice-to-text transcription by allowing for specialized preprocessing of recordings based on accent type. The capacity to distinguish accent variation categorization in speech is highly valued in the field of signal processing. This technology can be a key contribution in the context of developing speech recognition systems, language learning and forensic linguistics. It can be used to identify the speaker from his utterance which is useful in the context of a variety of criminal investigations and asylum hearings, thereby contributing to the advancement of Accent Variation Classification research and Multimedia Signal Processing and Analytics.

The task of identifying accent variation from human speech, as mentioned in [2], heavily depends on the fine acoustic features of human speech that can convey essential sociolinguistic and linguistic information about speakers during spoken communication, and hence these properties can operate as trustworthy identification markers of speakers' identity. Nonetheless, recognizing the most subtle underlying accent difference is difficult for automatic systems. Recently, Deep Learning techniques show superior results in Speech recognition and Accent variation classification tasks.

A significant amount of studies is conducted on Accent variation classification for languages like English, Spanish, Hindi etc. However, according to the utmost of our knowledge, very few studies or research activities in Bangla Speaker Accent Variation classification are currently accomplished. Bangla is rated seventh among the top 100 most spoken languages in the world [3], hence research with improved accuracy in Bangla Speaker Accent Variation categorization is in high demand now.

Bangla is different from other languages in terms of accent variation classification in deep learning due to several reasons. First, Bangla's phonetic inventory is more complicated, resulting in a greater range of sounds and different pronunciations. For instance, the vowel "a" can be pronounced differently depending on the region. Second, Bangla has more intricate stress patterns, leading to varying emphasis on different syllables. Such as, the word "আমি" (ami, meaning "I") can be pronounced with the stress on the first syllable or the second syllable, depending on the region. Third, Bangla's morphology is also more complex, causing variations in how words are formed and inflected, such as different conjugations for the verb "to go." These factors pose challenges in developing deep learning models for accent variation classification in Bangla compared to some other languages

A discreate Bengali accent variation classification model is proposed in this paper to recognize underlying information about accent variation by regions of Bangladesh including *Dhaka, Khulna, Barisal, Rajshahi, Sylhet, Chittagong, Mymensingh and Noakhali* from Bangla speech data. We use the Accent variation corpus dataset of this [4] paper to train our Deep Neural Network-based deep learning model and the model achieves 94% accuracy on unseen or new data. To validate our model's adaptability and usefulness, we compare its accuracy and performance to the findings of other neural networks as shown in Table IV.

This paper includes the following features and contributions:

The proposed solution targets Bangladesh's core regions, including *Dhaka, Khulna, Barisal, Rajshahi, Sylhet, Chittagong, Mymensingh, and Noakhali* and successfully labels the accent variation of each region with a higher degree of accuracy than previously released Bangla Accent Variation Classification approaches. Also, we compare the performance of our proposed model to other neural network models accessible in Table IV, and the one we propose achieves an accuracy rate of 94%, which is the most significant so far in the domain of Bangla Accent Variation Classification, according to the authors' findings. Furthermore, following a thorough domain analysis, we discover that most prior models frequently misclassify Mymensingh and Barishal classes as Noakhali. By that we find out that feature extraction should be improved to learn more complex variable of accent variation by model discussed in Section III (B). We use Zero Crossing Rate (ZCR) [5], Root Mean Square (RMS) [6], Mel-frequency Cepstrum Coefficient (MFCC) [7] and Mel-Spectrogram [8] based advanced feature extraction methods which is discussed in Section III (B). As a result, it enhances the machine's knowledge of the complex aspects associated with accent variation recognition from Bangla speech, raising the accuracy of each class to a mean of 94%, as shown in Table III. Lastly in Table III, we raise the precision score of the most misclassified region class, Mymensingh to 92%, Noakhali to 86%, and Khulna to 97%. As per the discussion in Section IV, many previous models struggle due to the misclassification of the classes.

II. Related Work

There is a scarcity of research conducted in the domain of Bangla Speaker Accent Variation Classification. This is a highly pertinent topic for linguistic research and documentation. In [9], the authors employed Mel Frequency Cepstral Coefficients (MFCC) as the feature extraction technique. For classification purposes, they utilized various algorithms including linear regression, decision tree, gradient boosting, random forest, and neural network. The dataset consisted of 9303 samples, and the highest achieved accuracy was 86%. However, a notable limitation of this research is the relatively high rate of misclassification observed for the Mymensingh accent, which is frequently classified as the Noakhali accent. Furthermore, the number of misclassifications is significantly higher for the Barishal accent, which is often erroneously identified as the Noakhali accent.

The paper [10] provides a comprehensive literature review on a study that used Stacked Convolutional Autoencoder (SCAE) for feature extraction and multi-label extreme learning machines (MLELMs) for classification. The authors have developed a customized dataset comprising regional data from Khulna, Bogra, Rangpur, Sylhet, Chittagong, Noakhali, and Mymensingh. However, it is worth noting that certain essential regional data, including Rajshahi, Dhaka, Barisal, and Formal Bangla, have been omitted, which can be considered as a limitation of this study. Furthermore, the researchers have solely utilized a single type of speaker and microphone for recording speech utterances in the dataset, which could potentially impact the generalizability of the results.

Authors in [11] worked on a paper that employed Mel Frequency Cepstral Coefficients (MFCC), along with Delta and Delta-delta features, for feature extraction, followed by classification using Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) algorithms. However, the study suffers from limitations, including the creation of small datasets specific to each dialect from Barishal, Noakhali, Sylhet, Chittagong, and Chapai Nawabganj regions, without including important regional data from Rajshahi, Dhaka, Khulna, Mymensingh, and Formal Bangla. Moreover, the paper reports a higher error rate of 37.92% in the Noakhali region using the GMM system, emphasizing the inadequacy of drawing conclusive results for an Accent Recognition system based on such a high error rate from a particular region in a smaller dataset.

The study [14] analyses seven monophthongal and four diphthongal vowels in Bangla. Acoustic features such as formant frequencies and vowel durations are examined, along with prosodic features like pitch and pitch slope. Instead of using the average method, the study employs linear regression to generate average contours of formant frequencies (F1, F2, and F3) for each vowel. This approach aims to provide a more precise analysis of the vocal tract shape in accented speech. The study investigates the role of acoustic and prosodic features in accent classification and recognizes the need for an accent-based Automatic Speech Recognition (ASR) system for robust speech recognition in Bangladeshi Bangla. The focus of the study is primarily on formant frequencies, pitch, pitch slope, and vowel durations. The analysis does not consider other potentially relevant linguistic features that could contribute to accent variation, such as consonant articulation or intonation patterns. The study's findings and conclusions are specific to the Bangladeshi Bangla language and the Neutral accent (NEU) and the deviant Sylheti accent (SYL) within that context. The applicability of the results to other languages or dialects may be limited.

The paper [15] proposes MFCC (Mel Frequency Cepstral Coefficients) features for classifying Bangla vowel phonemes. The paper focuses specifically on the classification of Bangla vowel phonemes, which may limit the generalizability of the findings to other languages or phonetic contexts. Since the system is not tested on a larger consonant phonemes database, its effectiveness and accuracy on a broader range of phonetic units remain unknown.

To extent of the authors' comprehension, no prior study in the field of Bangla Accent Variation Classification has attained over 94% accuracy. Furthermore, attaining an impressive accuracy of 94% in this type of robust dataset using DNN is unparalleled in previous research. In this paper, we show ZCR, MFCC, RMS and Mel Spectrogram based feature extraction techniques and a DNN-based model achieve noteworthy performance in classifying Bengali Accent Variation from 8 different region of Bangladesh with 94% accuracy.

III. Proposed Framework

Deep Neural Network (DNN) is used to develop our deep learning model for achieving this framework. DNN [12] is

proven as one of the most effective in terms of speech recognition task. It is a neural network with a high level of complexity, often at least two layers. It analyzes information in complex ways using advanced math modeling. Our framework is divided into three parts:

- Dataset Formation: In this part, we gather Bangla Accent Variation speech data from this paper [4] and create a data frame with audio paths and corresponding regions.
- Data Preprocessing: In this part, to make the best out of dataset, we use data augmentation and feature extraction before model development.
- Model Development: In this part, we develop a DNN-based deep learning model for Bangla Accent Variation classification shown in Fig 1.

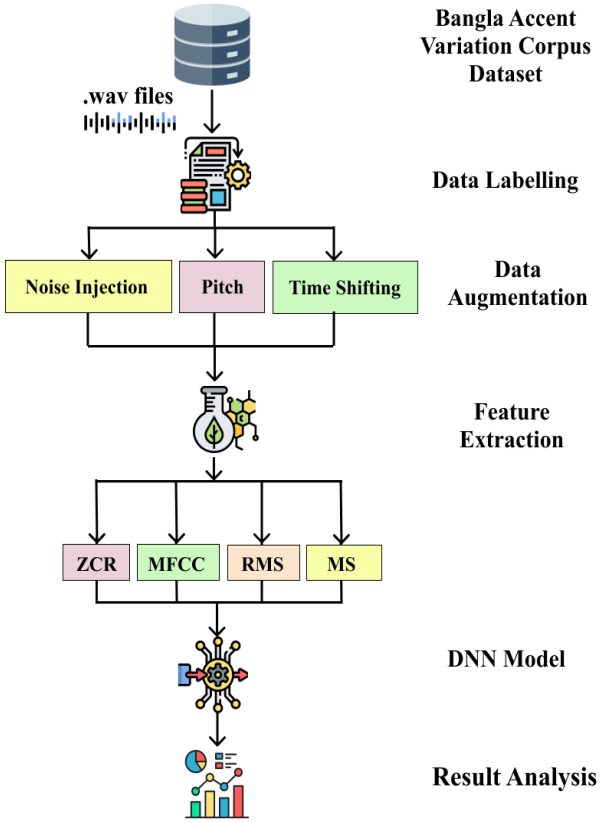

Fig. 1. Our Solution Diagram.

A. Dataset Formation

In this study, we gather data from the authors of the publication [4], which are gathered individually from various people, several YouTube videos, and a Google form. The speakers' ages range from 20 to 50, and both male and female voices are recorded. The average voice duration is 4-7 seconds. Dataset contains exactly 9303 different-region people's voices. Fig. 2. shows the amount of data of different regions. Furthermore, the audio file has been adjusted to 16,000 kHz In Table I, the count of male and female speakers and total samples are shown for different regions.

Table I. DATASET DESCRIPTION

Region	Number of speakers	Male	Female	Total Samples
Formal	89	49	40	1652
Barisal	67	29	38	1257
Noakhali	55	20	35	1213
Mymensingh	42	26	16	1053
Sylhet	27	14	13	958
Rajshahi	21	13	08	860
Chittagong	39	25	14	797
Dhaka	26	10	16	763
Khulna	52	32	20	750

B. Data Preprocessing

Following data collection, certain data augmentations are performed to boost the amount and diversity of training data available to train our model. We bring more randomness and non-linearity into training data by augmenting it, which allows our model to learn the complex acoustic variables to identifying the underlying accent variation context by region from given speech data rather than memorizing particular examples as we want to check the machine is intelligent or not by performing good in unseen data rather than checking weather learning is perfect. Checking weather learning perfections often refers to perform well in known data. But we want to develop an intelligent model which can shows its efficiency in real life unseen data. We supplement the data using noise injection, time shifting, and pitch.

Bangla is a tonal language, which means that the pitch or tone of a word can change its meaning. Accents in tonal languages often result from variations in pitch patterns, and speakers of different regions may use distinct tonal patterns, leading to accent variations. The high rate of misclassification in Bangla Accent Variation Classification studies can be due to numerous kinds of feature extraction approaches. Additionally, the feature extraction process may not capture the distinct phonetic or phonological characteristics of different accents, causing the classifier to struggle in distinguishing similar accents like Mymensingh, Noakhali, and Barishal. Furthermore, some accents may have complicated acoustic features that require the use of sophisticated feature extraction techniques for effective representation. To get the most out of our data, we use feature extraction algorithms such as Zero Crossing Rate (ZCR), Root Mean Square (RMS), Mel-frequency Cepstrum Coefficient (MFCC) and Mel-Spectrogram. ZCR is used for capturing the temporal information from Bangla speech data, RMS is used for capturing the amplitude of the audio signal. RMS is useful to capture overall energy, levels of intensity of Bangla speech. For capturing spectral envelope of the audio signal, we use MFCC. Mel-Spectrogram is for visualizing audio signals by regions shown in Fig. 2, Fig 3 and Fig.4. By that, it gets easier for machine to learn the complex variable of accents from speech.

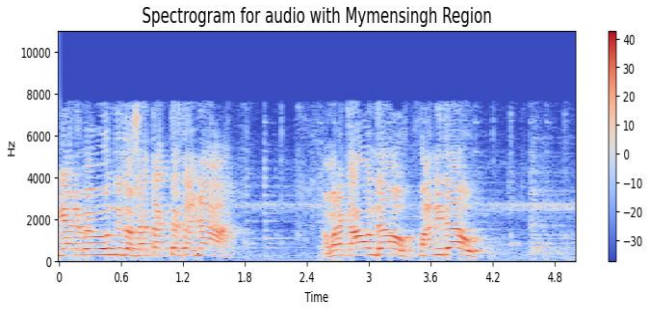

Fig. 2. Spectrogram for audio with Mymensingh Region.

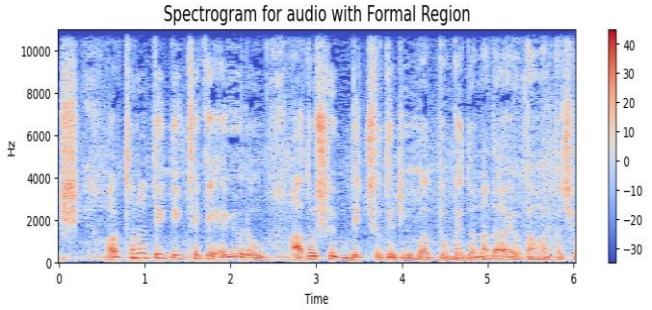

Fig. 3. Spectrogram for audio with Formal Bangla Accent Data.

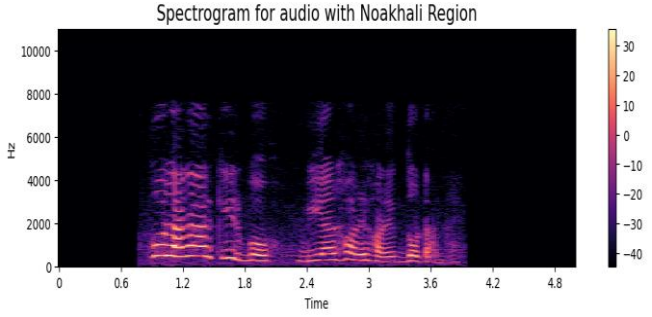

Fig. 4. Spectrogram for audio with Noakhali Region.

Finally, we use One-Hot Encoder to represent our data labels numerically so that our deep learning model can handle categorical data during training and prediction. We split the data into 80:20 ratio for training and testing.

C. Model Development

We propose a DNN-based model for accent variation classification and that ensures 94% accuracy on unseen or new data shown in Table III. In Fig. 5., we see the model consists of five dense layers [13], three ReLU Activation layers and a Sigmoid Activation layer. Dropout Layers are used to handle overfit issues as Dense layer comes with overfitting issues.

Let us begin by discussing the Model Architecture. Dense layers are often utilized for transforming input train data to high dimension feature space. It facilitates effective learning of perplexing characteristics of underlying accent variation context from audio. Dense layer additionally performs a linear transformation on the input data by multiplying it by a weight and adding a bias, as given in equation (1), where x is the input data, w is the weight, and b is the bias.

$$y = x_n \times w_n + b_n \tag{1}$$

As dense layer provides linearity, we use nonlinear activation functions to add nonlinearity to model. By that, it can learn complex variable of audio signal. When the input (x) is negative, such as -1.23, the ReLU function returns 0.0, as specified in equation (2). It simply disregards the negative values and incorporates nonlinearity onto the model.

$$(\max\ 0.0, x) \tag{2}$$

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 100)	16300
activation (Activation)	(None, 100)	0
dropout (Dropout)	(None, 100)	0
dense_1 (Dense)	(None, 200)	20200
activation_1 (Activation)	(None, 200)	0
dropout_1 (Dropout)	(None, 200)	0
dense_2 (Dense)	(None, 300)	60300
activation_2 (Activation)	(None, 300)	0
dropout_2 (Dropout)	(None, 300)	0
dense_3 (Dense)	(None, 100)	30100
dropout_3 (Dropout)	(None, 100)	0
flatten (Flatten)	(None, 100)	0
dense_4 (Dense)	(None, 9)	909
activation_3 (Activation)	(None, 9)	0

Total params: 127,809
Trainable params: 127,809
Non-trainable params: 0

Fig. 5. Model Architecture

For output layer, we use SoftMax activation function layer shown in equation (3), it is widely used logistic function which accepts value 0-1. The value closer to 1 tends towards to final output. The z stands for the value of neurons obtained by output layer.

$$softmax(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \tag{3}$$

The main purpose of using the Flatten layer here is to prepare the data for the subsequent Dense layers. The Flatten layer takes care of this conversion by "flattening" the output, resulting in a 1-dimensional representation.

The model is trained using the Adam optimizer, a notable technique of stochastic gradient descent that dynamically adjusts the learning rate during training phase. The categorical cross-entropy loss function is used for this work. It measures the difference between predicted probabilities and true class labels of accent regions. The ReduceLROnPlateau callback is used to enhance training efficiency of the model. This callback monitors the training loss and reduces the learning rate by a factor of 0.5 if the loss plateaus for a certain number of epochs (patience=2 in this case). The minimum learning rate is set to

0.001 to prevent it from becoming too small. The Batch size is 64 and epochs are 100 for training phase.

After training the model we use inverse transform method from One-hot encoder to convert numerical values of labels into actual labels and use prediction method to generate prediction on test data results shown in Table II where we see predicted output and actual output from test data.

IV. Performance Analysis

Based on our analysis, the Deep Neural Network (DNN) model demonstrates a 94% test accuracy and 96% training Accuracy in effectively predicting precise accent variation labels from Bangla audio regional data. The test loss is 0.2064 and the train loss is 0.0769. To validate this performance, we test our trained model with test data as it is important to check whether the model is intelligent enough to predict correct labels from unseen data. We evaluated the model's performance by the following performance metrics: Precision, Recall, F1-Score and Accuracy.

Table II. PREDICTED VS ACTUAL (FINAL 20 ROWS OF TEST DATA)

SAMPLE NO	PREDICTED LABALS	ACTUAL LABELS
5562	Noakhali	Noakhali
5563	Noakhali	Noakhali
5564	Rajshahi	Rajshahi
5565	Mymensingh	Mymensingh
5566	Noakhali	Noakhali
5567	Barisal	Barisal
5568	Sylhet	Sylhet
5569	Sylhet	Sylhet
…..		
5581	Chittagong	Chittagong

Accuracy is the number of correct predictions produced by the model. In this case, equation (4) is utilized to calculate the accuracy of our model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision is the number of correct positive predictions provided by the model. Equation (5) is used below to calculate the accuracy of our model.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall is determined by dividing the number of right positive outcomes by the total number of positive outcomes. Our model's recall is calculated using Equation (6).

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

F1-Score estimates the harmonic mean of accuracy and recall and gives a balance of the two. The following equation (7) is utilized to calculate our model's f1-score.

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

Table III. PERFORMANCE ANALYSIS OF DNN MODEL

CLASS	PRECISION	RECALL	F1-SCORE
Barishal	0.92	0.91	0.91
Chittagong	0.94	0.95	0.94
Dhaka	0.94	0.92	0.93
Formal	0.98	0.96	0.97
Khulna	0.97	0.98	0.97
Mymensingh	0.92	0.89	0.90
Noakhali	0.86	0.91	0.89
Rajshahi	0.95	0.95	0.95
Sylhet	0.97	0.97	0.97
accuracy			0.94
macro avg	0.94	0.94	0.94
weighted avg	0.94	0.94	0.94

To ensure the robustness of our model, we compare our model with other neural networks shown in Table IV.

Table IV. PERFORMANCE COMPARISON AMONG OTHER MODELS

MODEL NAME	PRECISION	RECALL	F1-SCORE	ACCURACY
LSTM	0.66	0.66	0.66	0.67
STACKED LSTM	0.70	0.71	0.70	0.71
DCNN	0.85	0.85	0.85	0.85
OUR PROPOSED DNN MODEL	0.94	0.94	0.94	0.94

V. Conclusion & Future Work

This study provides a Bangla Accent Variation Classification model based on deep neural networks that achieves highest accuracy of 94% in identifying accent variation by regions including *Dhaka, Formal, Khulna, Barisal, Rajshahi, Sylhet, Chittagong, Mymensingh and Noakhali* on Bangla speech data and when we compare with other neural network models available in Table IV. Our model achieves 94% accuracy which is so far the highest in this domain according to the authors' knowledge. The suggested model outperforms earlier research efforts in this field of Bangla Accent Variation Classification, showing superior performance. Overall, this research advances Signal processing technology and offers an important tool for voice data Bangla Accent Variation Identification.

In future, we are planning to release our own custom Bangla Accent Speech Corpus Dataset by adding more regions data like Bogra, Rangpur, Chapai Nawabganj, Cumilla and

including the regions described in this paper. Then, we are planning to test our model with this dataset. According to the authors' knowledge, this kind of dynamic dataset is not yet available, and we have started initial works to develop such dynamic dataset. We are developing an Android app based on this model for building secured and robust Bangla forensic analysis system.

REFERENCES

[1] M. Bryant, A. Chow, and S. Li, Classification of Accents of English Speakers by Native Language, http://cs229.stanford.edu/proj2014/Morgan%20Bryant,%20Amanda%20Chow,%20Sydney%20Li,%20Classification%20of%20Accents%20of%20English%20Speakers%20by%20Native%20Language.pdf (accessed Jun. 12, 2023).

[2] C. Themistocleous, "Dialect classification from a single sonorant sound using Deep Neural Networks," Frontiers, https://www.frontiersin.org/articles/10.3389/fcomm.2019.00064/full (accessed Jun. 13, 2023).

[3] M. Szmigiera, "Most spoken languages in the world," Statista, Mar. 30, 2021. https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/

[4] S. M. S. I. Badhon, H. Rahaman, F. R. Rupon, and S. Abujar, "Bengali accent classification from speech using different machine learning and Deep Learning Techniques," SpringerLink, https://link.springer.com/chapter/10.1007/978-981-15-7394-1_46 (accessed Jun. 13, 2023).

[5] S. K. Park, R. M. Kil, Y.-G. Jung, and M.-S. Han, "Zero-crossing-based feature extraction for voice command systems using neck-microphones," SpringerLink, https://link.springer.com/chapter/10.1007/978-3-540-72383-7_154 (accessed Jul. 4, 2023).

[6] R. E. Deakin, "A note on standard deviation and RMS" in Taylor & Francis Online, doi: https://doi.org/10.1080/00050351.1999.10558776 (accessed Jul. 4, 2023).

[7] Z. K. Abdul and A. K. Al-Talabani, "Mel Frequency Cepstral Coefficient and its Applications: A Review," in IEEE Access, vol. 10, pp. 122136-122158, 2022, doi: 10.1109/ACCESS.2022.3223444.

[8] R. Shah, P. Shah, C. Joshi, R. Jain and R. Nikam, "Heartbeat Prediction using Mel Spectrogram and MFCC Value," 2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET), London, United Kingdom, 2023, pp. 1-5, doi: 10.1109/GlobConET56651.2023.10150129.

[9] R. K. Mamun, S. Abujar, R. Islam, K. B. Md. Badruzzaman, and M. Hasan, "Bangla speaker accent variation detection by MFCC using recurrent neural network algorithm: A distinct approach," SpringerLink, https://link.springer.com/chapter/10.1007/978-981-15-2043-3_59 (accessed Jun. 13, 2023).

[10] P. S. Hossain, A. Chakrabarty, K. Kim, and Md. J. Piran, "Multi-label Extreme Learning Machine (mlelms) for Bangla Regional speech recognition," MDPI, https://doi.org/10.3390/app12115463 (accessed Jun. 13, 2023).

[11] P. P. Das, S. M. Allayear, R. Amin and Z. Rahman, "Bangladeshi dialect recognition using Mel Frequency Cepstral Coefficient, Delta, Delta-delta and Gaussian Mixture Model," 2016 Eighth International Conference on Advanced Computational Intelligence (ICACI), Chiang Mai, Thailand, 2016, pp. 359-364, doi: 10.1109/ICACI.2016.7449852.

[12] S. K. Sarvepalli, Deep Learning in Neural Networks: The science behind an Artificial Brain, https://www.researchgate.net/publication/331400258_Deep_Learning_in_Neural_Networks_The_science_behind_an_Artificial_Brain (accessed Jul. 4, 2023).

[13] P. Sperl, C. -Y. Kao, P. Chen, X. Lei and K. Böttinger, "DLA: Dense-Layer-Analysis for Adversarial Example Detection," 2020 IEEE European Symposium on Security and Privacy (EuroS&P), Genoa, Italy, 2020, pp. 198-215, doi: 10.1109/EuroSP48549.2020.00021.

[14] S. Kibria, M. S. Rahman, M. R. Selim and M. Z. Iqbal, "Acoustic Analysis of the Speakers' Variability for Regional Accent-Affected Pronunciation in Bangladeshi Bangla: A Study on Sylheti Accent," in IEEE Access, vol. 8, pp. 35200-35221, 2020, doi: 10.1109/ACCESS.2020.2974799.

[15] H. Mukherjee, C. Halder, S. Phadikar, and K. Roy, "Read-a Bangla phoneme recognition system," SpringerLink, https://link.springer.com/chapter/10.1007/978-981-10-3153-3_59 (accessed Jun. 13, 2023)