State of art Research in Bengali Speech Recognition

S.M. Saiful Islam Badhon
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
saiful15-7878@diu.edu.bd

Md. Habibur Rahaman
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
habibur15-7761@diu.edu.bd

Farea Rehnuma Rupon
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
farea15-7707@diu.edu.bd

Sheikh Abujar
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
sheikh.cse@diu.edu.bd

***Abstract*— Do we want AI as our future working partner or as our personal assistant? if the answer is yes than we must think about the communication between human and machines. Undoubtedly most comfortable communication media is the verbal communication. When it's about human speech it goes to Natural Language processing (NLP) and there are already many advanced works done in this sector specially in English. And some renown big companies already made some operating system-based assistant and robots those or really working near to human, but the concern is almost everything is very rich in English, English is international language but there are some popular languages as well where we need to focus, Bangla is one of them. Bangla is world's 8th most popular language almost 163 million people talk in this language which is not ignorable number at all [1]. There are very few quality works in this language even big companies are not that much advanced in Bangla like English or other languages. Here we muster some works in Bangla where researchers tried to solve Bangla speech recognition problems by their own methodologies. We tried to gather most accurate and recent work in this sector.**

Keywords—Speech to text, Bengali voice recognition, MFCCs, bangla speech dataset.

I. Introduction

There was a time when people communicate with computer only with 0 and 1 and then it improved in human understandable language and the upcoming era is looking for not punching the keyboard for communicate with machines. The future is communicating with machines through voice. It assumes, by 2020 there will be 50 voice search out of 100 in search engines [2]. Generally, if we try to represent sounds or voice of human in digital media it will just show some analogue signals, the challenge is converting those signals into text so that machine can work with them. NLP is not about only converting speeches into text but also analyzing the speeches in many ways such as: try to detect the emotion of the speakers, detecting hate words, even understanding sarcastic mood of speakers and many more things is part of natural language processing but this work strictly focusing on very basic thing which is converting speeches into texts. And obviously there will be some impact in Bangla as well. There are lots of applications of Automatic Speech Recognition System (ASR) but as we are specifically looking at Bangla, we tried to find applications and reasons why we need ASR system in Bangla. Number of smartphone users, number of smartphone users in Bangladesh and Kolkata are 8,921,000 and 3.5 million [3] [4] which are very big percentages of total population. So, it's now logical demand to work with Bangla speech recognition system. Detecting crime, if we look at number of crimes happened in Bangladesh in 2019, the number is very alarming which is 17484 [5]. In those crimes there must be some communications over phones or digital medias if use ASR system it must be easier to detect them and prevent them. In chat bots, in many places we can use chat bots as information providers. Even we can use ASR system in customer services or in call centers.

II. General Model of Speech Recognition:

This section tried to put some idea about how and which components we need to work with a speech recognition system. Very simple flow of below figure 1 will give us a better understanding of this.

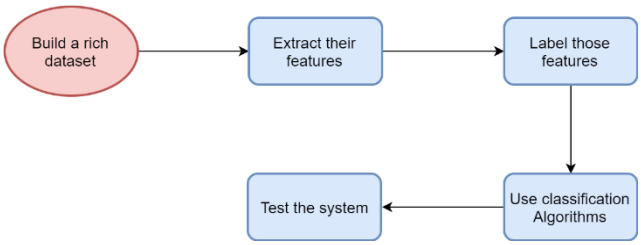

Fig 1. General model

III. Recent Advancements in Bengali Speech Recognition System

We tried to put some very recent and quality work in Bengali below:

In 2018, Syfullah, S.M. et al. [7] has worked on recognition of Bengali character from speech and they worked on vowel (e.g., অ, আ, ই, ঈ, উ, ঊ, ঋ, এ ,ঐ ,ও, ঔ) named Sorborno and consonant (e.g., ক ,খ ,গ ,ঘ, ঙ, চ, ছ, জ, ঝ, ঞ, ট, ঠ, ড, ঢ, ণ, ত, থ, দ, ধ, ন, প, ফ, ব, ভ, ম, য, র, ল, শ, ষ, স, হ, ড়, ঢ়, য়, ৎ, ং, ঃ, ঁ) named Byanjonborno. They used proposed revised algorithm named k-meansLBG (K-means + Linde Buzo Gray) for getting a well code book. For extracting features they used MFCCs. In their proposed method they propose two processing section training and simulation. In bellow figure 2 we tried to visualize their proposed method

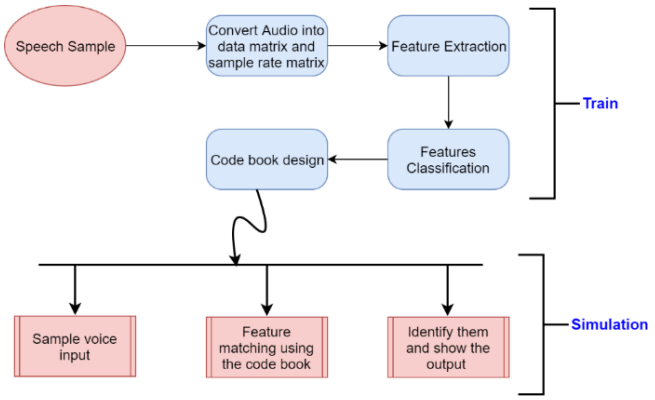

Fig 2. Syfullah, S.M. et al proposed model

They collected total 900 voices of 20 people. 45 characters from every speaker. For train the data this work used 12 speaker's data that's mean 60% of total 900 data and rest 40% for test the system. Finally, it has produced 81.61% accuracy with this data set. They work with Bengali characters and in future they want to work with words and sentences.

There is another work in 2018 which was done by Ahmed Sumon, S. Et al. [8] They worked on Bangla short speech command detection. The authors designed three different convolutional neural network (CNN) architectures and MFCCs for feature extraction. They used only 10 Bengali commanding words for the work. Those are given in table I.

TABLE I.

COMMANDING BENGALI WORDS

1	Agerta	আগেরটা(previous)
2	Aste	আস্তে(slowly)
3	At	আট(eight)
4	Baba	বাবা(father)
5	Bame jao	বামে যাও(go left)
6	Bari	বাড়ি (house)
7	Basa	বাসা (home)
8	Bon	বোন (sister)
9	Bondho koro	বন্ধকর (close)
10	Boro	বড় (big)

This paper works with a combined dataset. Their data set includes 1000 voice of 100 people 10 classes of every speaker. And additionally, they used Google's speech Commands dataset which is in English which contains 65,000 samples and duration was not more than 1 sec where they found 30 words with 1000 variation. They used three approaches for ultimate result, and They were just experiencing the Convolutional neural network in voice recognition with very small dataset. Their future target is increasing their dataset and fit the model with them for better outcome. table II will describe accuracy for different model:

TABLE II.

ACCURACY OF DIFFERENT MODEL

Model	*Training Accuracy*	*Testing Accuracy*
MFCC	85.44	74.01
Raw	69.08	71.44
Transfer	68.06	73.00

Amin, A. et al. [9] has done some research on continuous Bengali speech identification with the help of deep neural network in 2019. They used DNN-HMM and GMM-HMM based model on SRUTI open source dataset using kaldi toolkit and achieved better WER than previous approach which was CMU-SPHINIX based GMM-HMM for continuous Bangla speech recognition. They used SHRUTI Bengali speech corpus which open source. The duration of dataset is total 21.64 hours. The details of the dataset are given in table III, and this dataset was created by some researchers of IIT, Kharagpur.

TABLE III.

DATASET INFORMATION OF AMIN, A. ET AL

Unique Words	*Utterances*	*Speaker*	*Male*	*Female*
22012	13025	34	26	8

75% and 25 speaker's voice of whole dataset used as training data and rest 9 speaker's 25% used for testing. They tried mainly two mixed combination of algorithm for reaching their goal those are GMM-HMM and DNN-HMM and for feature extraction they used MFCCs, they used a better technique of feature extraction by using MFCC on top of LDA + MLLT + fMLLR. Their work didn't convert the Bengali voice into Bengali text directly its first converted in English written Bangla words then converted in Bengali characters. Below example will give a better understanding of their work.

In 2018 there is another work done by Mukharjee, H. et al. [10] where they researched on Bangla phoneme recognition which directly not speech recognition but can help the Bengali speech recognition. They used new Linear Predictive Cepstral Coefficient (LPCC-2) for feature extraction and for classification they use Ensemble Learning based classifier. Their dataset was enriched with 3710 phonemes where all the voices are Bangla swarabarna (vowel) and they used,

অ	আ	ই ঈ	উ ঊ	এ	ও	অ্যা

The data are recorded with Frontec JL-344 in stereo mode and those are in .wav format. They proposed a methodology which is given in figure 3.

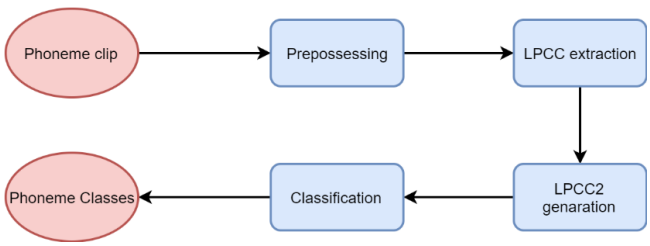

Fig 3. Mukharjee, H. et al proposed model

For testing they used fivefold cross validation procedure at 100,200,300,400 and 500 iteration where they used 53 speaker's voices. And finally, they got highest 99.06% accuracy.

Islam, J. et al. [11] worked on speech recognition using Recurrent Neural Network. Their proposed model can work

in offline; the work is open source and integrable. They proposed two models the first model is Convolutional Neural Network (CNN) and second one is Bengali speech system using recurrent neural networks as the technique of deep learning. They worked with a dataset where they got 33000 audio files. They used 80% as train data, 10% as test data and 90% as validation data. Total processing period of 33000 data was 33 hours. And the sampling rate was 16000. And they got an accuracy of 86.058%.

In 2012, Sultana, Akhand Et al. [12] has worked on Speech Application Program Interface (SAPI) for Bengali speech recognition. They collect an article from a newspaper and experimented on that lines using SAPI to enquire about speech to text conversation. They tried to fit SAPI utterance from continuous Bangla voice in SAPI's precompiled etymology file and if matches occur then SAPI turned Bengali terms in English character. Then the words were used to retrieve Bengali terms from the database, then return terms to real Bengali characters and finalize the phrases. Many English terms for a specific Bengali term in SAPI's grammar file have been identified to resolve tone variance of individuals as well as pronunciation variability in language communities and have been demonstrated to boost overall efficiency of the system. This study used Speech Application Programming Interface or SAPI 5.4, Microsoft SQL Server 2008, Microsoft Visual Studio 2010, object (.DLL) resource, Avro (Bangla writing Software) for experiments and got the accuracy is 78%. They used both one-to-one and one-to-many relationship of Bangla utterance with English letters. For one-to-one relationship from overall 396 words they can recognize 264 words by the system and the recognition rate is 66.67%. On the other hand, for one-to-many relationship from overall 396 words 311 words can be recognized by the system and the recognition rate is 78.54%. They used this process for distinct words too to find proper Bangla words from speech. Finally, they concluded that the problem is it's sluggish and its operations are sequential. A quicker voice recognition tool such as the Java Speech API could also be showing improved results. Parallel processing on SAPI will boost efficiency and continue as a test of the future. In 2016, another paper written by Ghosh, A.H.M Et al. [13] researched with continuous Bangla sentences. They use Bangla vowels to get the formant frequencies and its corresponding bandwidth. They contemplated Bangla language corpus named "SHRUTI" to get the formant characteristics of Bangla vowels. At first, vowels are tracked manually for formant estimation. Then the vowel is resampled to 8KHz signal because the maximum frequency of human voice is pursued as 4 KHz. A lowpass filter is planned for band limiting with a cutoff frequency which is of 3600 Hz for limiting given vowel signal considering with human voice signal frequency. The level of energy is gradually decreased because the formant can be referred. So, it become very complex to find the higher formant frequencies. Then a pre-emphasis filter is implemented to that band limited signal for defeating this problem which rises the higher formant frequencies energy label and appeases DC component of a given speech signal. Cepstrum reliant formant estimation and Linear Predictive Coding (LPC) methods have been used for the experimental studies. 5 male and 5 female speakers (5 utterances each) voice was used for the experiment. After that they find that for LPC the accuracy for Male is 75% and Female is 71%. On the other hand, for Cepstrum the accuracy for Male 77% is and Female is 72%.

In another paper, which was done by Ahmed, Wahid Et al. [14] in 2018, developed a VISO calculator application applying CMU Sphinx and android TTS API capable of understanding and deriving mathematical term from both diverse and consecutive speech in the Bengali language. The term is then measured and in Bangla the auditory output is pronounced. They completed the functions of Bengali VISO calculator into three separate steps those are, Bengali voice Identification, determining and calculating mathematical equation, pronouncing calculated result. For getting the speech recognizer output at first, they generate ARPA transcriptions of Bangla words following some steps. They then construct a corpus of Bangla texts which used 25 Bangla vocabulary terms to write several sentences. Those were stated in a vocabulary chart. The sentences listed in the text-corpus were recorded as voice within an audio file which is then transmuted to WAV format with a sampling rate of 8KHz using Sphinx. To form an audio database, they remove noise and normalize the audio file and then separate that by the termination of every sentence. They use a transcription file for mapping the database to the text-corpus. For deriving mathematical expression, they get the known series, replace terms with Bangla digits and symbols, delete illicit inputs and display expression, then eliminate Bengali digits with English, measure results and finally reinstate English digits with Bengali and display output. They use an algorithm for the application of VISO calculator. After that voice recognizer executes its performance, the algorithm is used to record the prior determined outcome to update the result each time. If the measured outcome and the initial value matches, the speaker is regarded not to provide utterance. A timer with a clock frequency of 1Hz is then started. When the result stays the same for more than two seconds, the TTS API, which is programmed for the Bengali language, is brought up. This voice recognition system has achieved almost 86.7% accuracy while examined with 20 speakers in various environments with specific accents. In 2016 Bhattacharjee, Khan Et al. [15] proposed a design and implement a prototype of a robotic motion which is controlled by Bengali voice based on an effective algorithm for identifying Bangla voice commands accurately. For feature mapping methodology, they used MFCC (Mel-Frequency Cepstrum Coefficients) dependent speech feature extraction method and Vector Quantization. For software development they followed some steps. At first, for voice command recognition they set login option which will have password system and then give the voice commands which will be used in the system stored by Vector Quantized (VQ) codebooks. While codebook generation, the continuous speech signal varies to 10 MS with 5 MS overlap frames. On each frame hamming window is used. There are different frequencies tones in the speech signal. A subjective rate for the actual frequency f (Hz) of each tone is measured on the 'Mel' scale. For computing the mels for a given frequency f (Hz), the formula is given in equation (1):

$$M(f) = 2595 \times log10(1 + f/700) \qquad \ldots\ldots\ldots (1)$$

The spectrum of log mel is transformed back to the period deriving in the mel frequency cepstrum coefficients

(MFCC)s. After these steps the command generation happens in the system. Their database contains total 100 commands for Bangla voice command identification. 8 female and 12 male participants of 15-45 years, total 20 speakers provided 5 utterances each. Voice commands were collected for testing both in noiseless and noisy environment. Their proposed algorithm for Bangla voice recognition commands shows approximately 98% accuracy in the noise free condition and 93% accuracy in noisy environment.

In 2016 Mukherjee, Rakshit Et al. [16] designed a system which is known as REARC (Record Extract Approximate Reduce Classify) for Bengali character identification. For the scheme, Mel Scale Cepstral Coefficient (MFCC) features are considered. Its database consists of 3150 Bengali vowel phonemes. They build a database of 45 participants consisting 18 females and 27 males of 20-75 years old. The 7 Bengali vowel characters were represented in the database with their symbols for International Phonetic Alphabet (IPA), alphabetical mapping and an identical pronunciation in English. Every person uttered the 7 Phonemes from beginning to the end repeatedly 10 times each. The data was collected in audio mode in the.wav format with 1411 kbps of bit rate. Frames were been split by the signals with 256 sample points of width and an inter-frame overlap of 100 points. Each frame was multiplied by a hamming window to ensure consistency and remove the jitters as well. The equation of the hamming window is given in equation (2):

$$w(n) = 0.54 - 0.46\,cos(2\pi n/N - 1) \quad \ldots\ldots\ldots\ldots (2)$$

MFCC has been considered here for the feature extractions of the input data. Since each audio clips frame had 19 MFCC values, feature sets of different sizes had been attained that presented a challenge in the classification project. Each of these features was found and discarded, having the same value for all the Phonemes. It reduced the size to 175 of the feature set, representing a 56.14% retrenchment. The Principal Component Analysis (PCA) considering the Ranker Search Technique had been accomplished with the aid of the well-known open source software WEKA on the data to introduce the second type of dimensional diminution. This method reduced the feature dimension from 399 to just 94, showing a reduction of 76.44 percent. In the final phase, both sets were used separately. 5-Fold Cross Validation has been applied to their acquired 450 X 7 (3150) data to obtain the result. For the present work, a Multi-Layer Perceptron (MLP) classifier was conducted. To train the MLP, Back propagation method was applied. The method was tested with 4 Models. After that they find 98.22% of accuracy in this system.

In 2018, Jillur Rahman Saurav et al. [17] from SUST, Bangladesh have worked on Bengali Speech Identification system for a search browser named Pipilika. In Bengali voice recognition researchers remain confined to small-scale voice identification. In this paper, they used several methods like GMM-HMM and DNN-HMM with some advanced feature extraction techniques like LDA, MLLT and finally the noticed, The model based on DNN misinterprets less terms than other models and DNN has sensitivity issue for noise because sometime it converts noises into words. Instead of introducing new methodology for Automatic Speech Recognition System (ASR) they made a voice identification system used a toolkit named Kaldi. Which is considered as an open source voice recognition having almost all essential voice recognition algorithms. Though there are many more speech recognition toolkits named HTK, Sphinx, Julius but the reason of using Kaldi is it has a great support for Deep Neural Network, and it shows effectiveness on benchmark datasets. For making datasets they choose 50 different speakers and 500 unique words. They recorded all the words in studio environment. The rate of train and test data is 80 and 20. Finally they have got the lowest Word Error Rate (WER) for Bengali Speech Recognition. For GMM-HMM they got 3.96% WER and for DNN-HMM they got 5.30% WER.

Another Paper [18], tries to research about feature extraction and tried to recognize Bangla phoneme, commands, isolated words. It discussed short period energy estimation, silence dismissal process, rate of framing the date, Bangla voice recognition, phonemes, commands and words sample. Also, it discussed data pre-processing, extraction of features for power spectral analysis. For pattern recognition and training purposes NN is used. The paper also discussed the female and male speech signals. FFT is used for feature extraction. Finally, this paper provides good results for the signal of male and female speech. This paper consists of 640 samples from 8 different Bangla phonemes, commands and words for feature extraction and speech identification. Bengali phoneme gives 95% accuracy, but it gives 55% when the speech samples increase into 240 and participants up to 6 persons. For isolated words, it provides 75% accuracy and the same way it falls down into 51.5% when samples and participants number are increased. And for the Bangla command, it provides 70% accuracy for males and 32.5% for females. The result slightly influenced when the window length is slightly changed. For future work, LPCA and MFCC will be taken place. In other paper Rahman M.et al. [19] from KUET, Bangladesh discussed how the speech recognition process works and the steps of ASR system. They mentioned 3 main steps of ASR and for each step, they mentioned many algorithms. LPC, PLP, MFCC, RASTA-PLP algorithm is used for extracting a feature from data. Additionally, DTW, HMM, RNN is used for feature matching. K-Nearest neighbor, SVM, GMM used for the data classification. For Bangla word recognition they've used SVM with DTW. With this process, an Automatic Speech Recognition System is also proposed in this study. MFCC is considered for static feature extraction and Delta and Delta-Delta which is the first and second derivative of MFCC is used for dynamic feature extraction. For training purposes, they collected 5 unique words from 40 different speakers in a noise-proof room. They tested with 12 speakers through the words “পানি”, “বাবা”, “বই”, “মাটি’”, “দেশ” and got different results for each word. Average they have got 86.08% accuracy. They are hoping to extend this research in the future. In other research papers, Md. Abul Hasnat et al. [20] from BRAC University worked on voice recognition from application and performance, implementation aspects. In their research work, they have talked about isolated words and continuous speech recognition. They applied HMM for the pattern classification and incorporate stochastic language model in this system. In this signal processing level, they execute adaptive noise reduction and endpoint identification. The paper mentioned that in real-world implementation we’ll be able to use HMM if we can solve three problems and it includes the algorithm, they use to solve these problems. For evaluation problems,

they have used forward algorithm, for the decoding problem they have used the Viterbi algorithm, for proficiency problem they just adjusted all the parameters. MFCC with first and second-order coefficients have been extracted from every voice wave signal. Their system was implemented using Cambridge HMM Toolkit (HTK). For the dataset, they collected 100 unique words from 5 different speakers. For isolated speech recognition they used a normal office environment. The overall performance decreases by 20% when different speaker performs the word. Another paper in 2018 by khan, F. et al. [21] mainly constructed isolated words for speech corpus in Bangla Language. The paper mainly described the model and recording procedure of Isolated word speech corpus. they have mentioned that this is the first speech corpora in Bengali language in its type, coverage and size. To construct the database there were 100 speakers and each of them spoke 1081 words. The speakers were from a different location and they mentioned the number of speakers came from which location. They made another database for test purpose and there were 50 additional speakers who spoke the same 1081 words. They have collected a total of 375 hours of speech recording.

IV. Findings and Suggestions

This section of work, tried to find out some attributes by comparing the above methodologies which given in table iv

TABLE IV

IMPORTANT INFORMATION ABOUT THE STUDY

Attributes	***Values***
Largest Dataset	"SHRUTI" open source dataset (22012 unique words)
Best accuracy	99.06% on only 1000 voice [10]
Best effective accuracy	75% on largest dataset SHRUTI [9]
Most used Classification algorithm	Hidden Markov Model and Gussian Mixture Model
Most used feature extraction techniques	Mel-frequency cepstral coefficients and linear prediction coefficients
Tools used	HTK, Sphinx, Julius, kaldi
Latest technique	Extracting Mel spectrogram of voice and classify with CNN

Effective accuracy in above table means, accuracy on better datasets like more unique words and more verity of utterance. From all above study we got some important points like, for working with speech we must need healthy dataset. HMM and GMM is the most used and successful algorithm for speech recognition but with larger dataset we can try deep neural network (DNN) and convolutional neural network (CNN) which will give us better accuracy.

V. Conclusion

This work tried to show the importance of Bangla in ASR systems and presented some paper's methodology which are recently done. We reviewed here 15 papers by mentioning their data amount, accuracy, tools, feature extraction algorithms, classification algorithms, which type of problem they solved etc. All paper's ultimate goal was enriching the work in Bangla speech recognition. Hopefully, our work will help the new researchers in this sector for getting idea about previous works and what should they do in future.

.

References

[1] Wikipedia (2019) Bangladesh. https://en.wikipedia.org/wiki/Bangladesh Accessed 26 Nov 2019

[2] Sentence R (2018) The future of voice search: 2020 and beyond https://econsultancy.com/the-future-of-voice-search-2020-and-beyond/ Accessed 26 Nov 2019

[3] BankMyCell (2019) How Many People Have Smartphones in the World? https://www.bankmycell.com/blog/how-many-phones-are-in-the-world Accessed 26 Nov 2019

[4] Sharma P (2017) What is the number of smartphone users in Kolkata? https://www.quora.com/What-is-the-number-of-smartphone-users-in-Kolkata Accessed 26 Nov 2019

[5] Bangladesh Police (2019) Crime Statistics 2019 https://www.police.gov.bd/en/crime_statistic/year/2019 Accessed 26 Nov 2019

[6] Ajibola Alim, S., & Khair Alang Rashid, N. (2018). Some Commonly Used Speech Feature Extraction Algorithms. In from Natural to Artificial Intelligence - Algorithms and Applications. https://doi.org/10.5772/intechopen.80419

[7] Syfullah, Sadi & Zakaria, Zareen & Uddin, Md. Palash & Rabbi, Md. Fazle & Afjal, Masud Ibn & Nitu, Adiba. (2018). Efficient Vector Code-book Generation using K-means and Linde-Buzo-Gray (LBG) Algorithm for Bengali Voice Recognition. 10.1109/ICAEEE.2018.8642994.

[8] S. Ahmed Sumon, J. Chowdhury, S. Debnath, N. Mohammed and S. Momen, "Bangla Short Speech Commands Recognition Using Convolutional Neural Networks," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, 2018, pp. 1-6. 10.1109/ICBSLP.2018.8554395

[9] M. A. A. Amin, M. T. Islam, S. Kibria and M. S. Rahman, "Continuous Bengali Speech Recognition Based on Deep Neural Network," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6.

[10] Mukherjee H., Phadikar S., Roy K. (2018) An Ensemble Learning-Based Bangla Phoneme Recognition System Using LPCC-2 Features. In: Bhateja V., Coello Coello C., Satapathy S., Pattnaik P. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol 695. Springer, Singapore

[11] J. Islam, M. Mubassira, M. R. Islam and A. K. Das, "A Speech Recognition System for Bengali Language using Recurrent Neural Network," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 73-76.

[12] S. Sultana, M. A. H. Akhand, P. K. Das and M. M. Hafizur Rahman, "Bangla Speech-to-Text conversion using SAPI," 2012 International Conference on Computer and Communication Engineering (ICCCE), Kuala Lumpur, 2012, pp. 385-390.

[13] Ghosh, Tonmoy & Saha, Subir & Ferdous, A.H.M. (2016). Formant Analysis of Bangla Vowel for Automatic Speech Recognition. Signal & Image Processing: An International Journal (SIPIJ). 7. 10.5121/sipij.2016.7501.

[14] T. Ahmed, M. Ferdous Wahid and M. Ahsan Habib, "Implementation of Bangla Speech Recognition in Voice Input Speech Output (VISO) Calculator," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, 2018, pp. 1-5.

[15] A. Bhattacharjee et al., "Bangla voice controlled robot for rescue operation in noisy environment," 2016 IEEE Region 10 Conference (TENCON), Singapore, 2016, pp. 3284-3288.

[16] H. Mukherjee, S. Phadikar, P. Rakshit and K. Roy, "REARC-a Bangla Phoneme recognizer," 2016 International Conference on Accessibility to Digital World (ICADW), Guwahati, 2016, pp. 177-180.

[17] Saurav, Jillur & Amin, Shakhawat & Kibria, Shafkat & Rahman, M. (2018). Bangla Speech Recognition for Voice Search. 1-4. 10.1109/ICBSLP.2018.8554944.

[18] Chowdhury, Md & Khan, Md. (2019). Power spectral analysis and Neural network for feature extraction and recognition of speech. 1-7. 10.1109/IconDSC.2019.8816913.

[19] Rahman, Md & Roy, Debopriya & Hasan, Md. (2018). Dynamic Time Warping Assisted SVM Classifier for Bangla Speech Recognition. 1-6. 10.1109/IC4ME2.2018.8465640.

[20] Hasnat, M.A., Mowla, J., & Khan, Md. (2007). Isolated and continuous bangla speech recognition: implementation, performance and application perspective

[21] Farukuzzaman Khan, M. Abdus Sobhan, M. (2018). Construction of Large Scale Isolated Word Speech Corpus in Bangla. Global Journal Of Computer Science And Technology, https://computerresearch.org/index.php/computer/article/view/1690