

# Bangla song genre recognition using artificial neural network

**Mariam Akter, Nishat Sultana, Sheak Rashed Haider Noori, Md Zahid Hasan**

Health Informatics Research Laboratory (HIRL), Department of Computer Science and Engineering, Faculty of Science and Information Technology, Daffodil International University, Dhaka, Bangladesh

## Article Info

### Article history:

Received Apr 2, 2023

Revised Jul 20, 2023

Accepted Aug 2, 2023

### Keywords:

Artificial neural network  
Bangla song genre recognition  
Chroma frequency  
Deep learning  
Mel frequency cepstral coefficients  
Tempo

## ABSTRACT

Music has a control over human moods and it can make someone calm or excited. It allows us to feel all emotions we experience. Nowadays, people are often attached with their phones and computers listening to music on Spotify, SoundCloud, or any other internet platform. Music information retrieval plays an important role for music recommendation according to lyrics, pitch, pattern of choices, and genre. In this study, we have tried to recognize the music genre for a better music recommendation system. We have collected an amount of 1820 Bangla songs from six different genres including Adhunik, rock, hip hop, Nazrul, Rabindra, and folk music. We have started with some traditional machine learning algorithms having k-nearest neighbor, logistic regression, random forest, support vector machine, and decision tree but ended up with a deep learning algorithm named artificial neural network with an accuracy of 78% for recognizing music genres from six different genres. All mentioned algorithms are experimented with transformed mel-spectrograms and mean chroma frequency values of that raw amplitude data. But we found that music tempo having beats per minute value with two previous features present better accuracy.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Mariam Akter

Health Informatics Research Laboratory (HIRL), Department of Computer Science and Engineering

Faculty of Science and Information Technology, Daffodil International University

Dhaka, Bangladesh

Email: mariam15-12754@diu.edu.bd

## 1. INTRODUCTION

Speech is the most powerful tool for communication and it's the only way to express thoughts and emotions. A song is a piece of music created with the intention of being sung by a human voice. This is frequently done at specific, fixed pitches [1], such as melodies that employ sound and silence patterns. Songs come in a variety of formats, such as those with sections that are repeated and varied, which determine the genre of the song. Songs have been a remarkable form of entertainment for decades. Almost all songs are written with a focus on a particular genre, for example, rock, folk, pop, jazz, classical, hip hop, and country music. Day by day new genres are emerging daily for example nowadays people are listening to more electronic music and cover song [2]. Song genre classification and prediction plays a valuable role in various fields and industries such as recommending music based on user mood [3], emotion [4] is the important task and on the other hand applications for genre classification and prediction include music education and music therapy, where choosing music based on genre can aid in therapeutic purposes.

Some research has been done for music genre classification. Liang *et al.* [5] try to classify music

genres using different techniques such as hidden Markov model (HMM's) and using canonical correlation analysis (CCA). In another paper, try to classify potential hit song and music genres [6]. For classifying those, they extract one dimensional audio features and using those they apply on different classification algorithms but feed-forward neural network gives almost 81% accuracy. Sanden *et al.* [7] tries to predict genres using a clustering approach. For the clustering approach, they calculated structural distance from musical data and applied a k-means clustering algorithm and got high clustering accuracy. Zangerle *et al.* [8] propose to combine low- and high-level audio features of songs in a deep neural network that distinguishes low- and high-level features to account for their particularities. Bahuleyan [9] tried to extract different song features and using those features developed a deep learning model using artificial neural network (ANN), visual geometry group-16 (VGG-16), and convolutional neural network (CNN) model using YouTube song and VGG-16 CNN model gives almost 64% accuracy. Mamun *et al.* [10] presented an approach to classifying Bangla music song genre. They tried to classify 6 different Bangla music genre named 'Bangla Adhunik', 'Bangla hip-hop', 'Bangla band music', 'Nazrulgeeti', 'Palligeeti', and 'Rabindra Sangeet'. They proposed a neural network model which gives 74% accuracy.

Bangla is the most spoken language. Almost 200 million people worldwide say Bangla as their first language which is the 4th among all over the world [11]. Bangla natural language processing (BNLP) [12] resources are very less compared to the English language but it is growing very rapidly in some field such as speech recognition [13]-[16], Bangla offensive word recognition [17], but there are more less work for Bangla music and statistics regarding the classification of musical genres have been provided in literature. Now we are aware of improvements made thus far in both this sector and Bangla music. We explored a number of related processes, such as data collection with various genres, different types of music feature extraction and feature selection, and performance comparison. As a result, the method calls for the opportunity to make a new contribution to the Bangla music community and to the field of music genre recognition. So, using the BNLP and deep learning in this work, we try to classify song genres named folk, Lalon Giti, Robindro Sangit, and Bangla pop music using Bangla songs.

## 2. METHOD

### 2.1. Process workflow

The overall process workflow outlines our approach to performing this work and the steps we take to accomplish them, and how we intend to carry out this task. We encounter numerous issues and review our workflow to produce the finest outcome. The project's overall working procedure is shown in Figure 1.

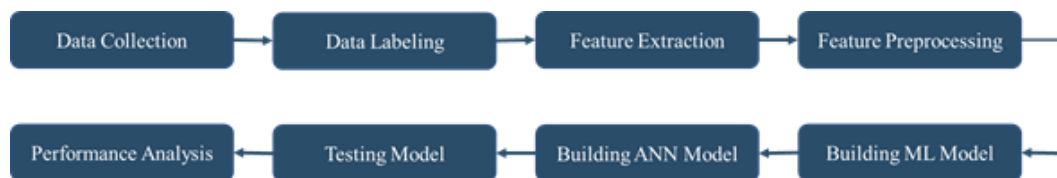


Figure 1. Overall working process

### 2.2. Data collection

Data collection for the method was one of the most challenging tasks over the entire process. We used our dataset. For collecting this dataset, we used YouTube as a base resource. We make a playlist for each genre and after making this genre playlist we use the python pytube library with the help of ffmpeg to download the song playlist. Pytube is the most popular python library for downloading YouTube videos and ffmpeg is responsible for converting different formats into desired formats. We download all songs in .mp3 formats. The overall data collection process of the project shown in Figure 2.



Figure 2. Overall data collection process

### 2.3. Data description

We already discuss the process of data collection for this study. We collect a total of 1820 songs of 6 different genres named Adunik, rock, hiphop, Nazrul, Rabindra, and folk. Genre-wise data ratio which we used to build our algorithm are shown in the Table 1.

Table 1. Genre-wise statistics

Genre	Collection
Adunik	297
Rock	300
Hiphop	301
Nazrul	312
Rabindra	298
Folk	312

### 2.4. Feature extraction

#### 2.4.1. Mel frequency cepstral coefficients

The first Mel frequency cepstral coefficients (MFCCs) were released in the late 1970s [18]. Over the years that followed, numerous developments were made. For different use cases, many implementations have been suggested. The team, led by Todor Ganchev, has analyzed the four most popular speaker recognition systems, starting with the MFCC FB-20 implementation [19]. But MFCCs are also very popular for music feature extraction. In our work, we extracted the 20 MFCC features using the MFCC FB-40 implementation. This process involves a number of steps:

- Framing: the audio signal is represented by 20 MFCC characteristics. These 20 sections required to be in the same place for this to be more practical. However, an audio signal is dynamic and evolves over time. In order to partition the entire signal into frames, it is assumed that each segment will behave as a stationary signal. The frame has a default length of 25 ms [20]. The majority of the music files in our dataset have a duration of 180 seconds on average and a sample rate of 44.1 kHz. Following framing, each audio file is divided into 400 segments, or frames, with  $(0.025 * 16000) = 400$  samples in each frame.
- Calculating power spectrum: power an estimation of the frequencies present in each frame is provided by the power spectrum. Each frame has been applied to the discrete fourier transform in order to calculate that. For this calculation, we use in (1):

$$P_i(k) = 1/N \left( \left| \sum_{n=1}^N (S_i(n)h(n)e^{-i2\pi kn/N}) \right|^2 \right) \quad (1)$$

Here,  $h(n)$  is a hamming window,  $S_i(n)$  is the signal representation of each sample of the  $i^{th}$  frame,  $P_i(k)$  is the power spectrum of the  $i^{th}$  frame, and  $K$  is the discrete fourier transform's length.

- Applying Mel filterbank: it is crucial to the implementation of the MFCC. A set of 40 equal area filters have been employed in this implementation. A mathematical representation of each filter is shown in (2):

$$\begin{cases} 0 & \text{for } k < \int_{b_{i-1}} \\ \frac{2(k - \int_{b_{i-1}})}{(\int_{b_{i-1}} - \int_{b_i})(\int_{b_{i+1}} - \int_{b_{i-1}})} & \text{for } \int_{b_{i-1}} \leq k \leq \int_{b_i} \\ \frac{2(\int_{b_{i+1}} - k)}{(\int_{b_{i+1}} - \int_{b_i})(\int_{b_{i+1}} - \int_{b_{i-1}})} & \text{for } \int_{b_i} \leq k \leq \int_{b_{i+1}} \\ 0 & \text{for } k > \int_{b_{i+1}} \end{cases} \quad (2)$$

Here,  $\int_{b_i}$  denotes the filter's boundary, and "k" denotes the  $k^{th}$  coefficient of N-point DFT. These 40 filters are just a collection of 40 vectors with a few non-zero values at various vectors. The remaining vector values are all zeros. The value of each vector, when multiplied by the power spectrum, corresponds to the energy of the corresponding power spectrum segment. After that it gives 40 energy representations for 40 filters.

- Taking the energy log: for each filterbank segment, the energy estimation logarithm is now calculated.
- Taking the discrete cosine transform (DCT) of log energies: all of the log energies are subjected to a DCT in the last step. This serves as a representation of the 40 audio features as an energy estimation.

We only used the first 20 features after computing the 40 energy features. Mainly the reason is, signal or music is represented by the first 20 linear characteristics.

### 2.4.2. Chroma frequency

The magnitude spectrum, which represents the spectral content of each frame, is then obtained by performing the short-time fourier transform (STFT). The energy within each chroma band is calculated by adding the magnitudes of the associated frequency bins after using a filterbank to map the frequencies to chroma bands. This measures how many musical notes are present in each frame. Figure 3 shows the mean chroma frequency values of each genre.

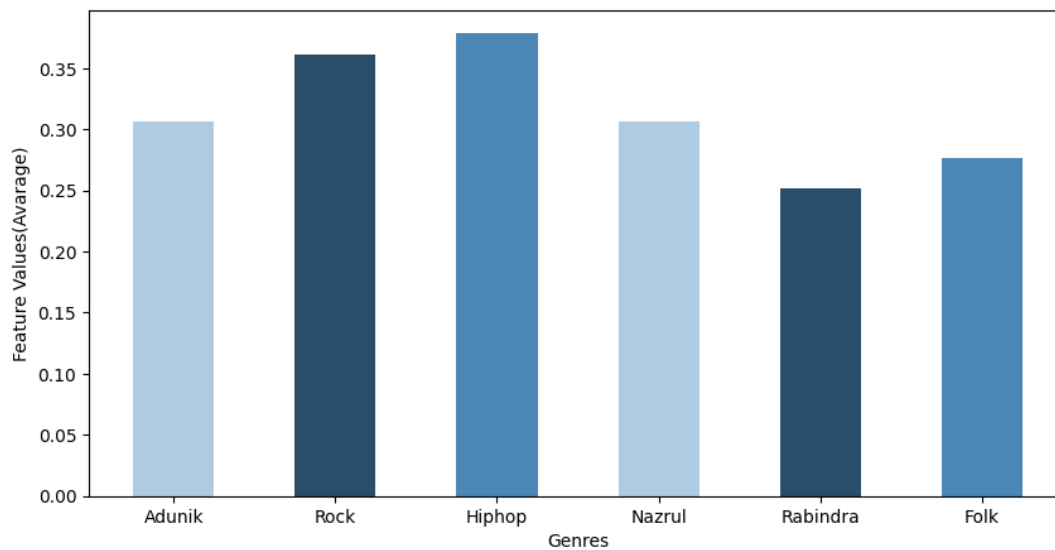


Figure 3. Mean chroma frequency values

### 2.4.3. Tempo

Tempo denotes the speed or slowness of music for musical signaling. Beats per minute (BPM) is used to measure it. When the BPM is high, the music is considered to be fast, and when it is low, the music is considered to be slow. The mean tempo values of each genre shown in Figure 4.

### 2.4.4. Root mean square error

An audio signals overall energy or amplitude can be estimated using the root mean square error (RMSE) feature. It divulges details regarding the average signal power level. In (3) is used to calculate the signal energy.

$$\sum_{n=1}^N |x(n)|^2 \quad (3)$$

Here,  $x(n)$  is a discrete-time signal. Energy indicates the volume of any audio signal. As each genre differs in terms of how it sounds and we observe that each has a unique energy value pattern. So, to extract the musical genre, we used the root mean square feature. Calculating RMSE involves using (4), where N is the sample size.

$$\sqrt{\frac{1}{N} |x(n)|^2} \quad (4)$$

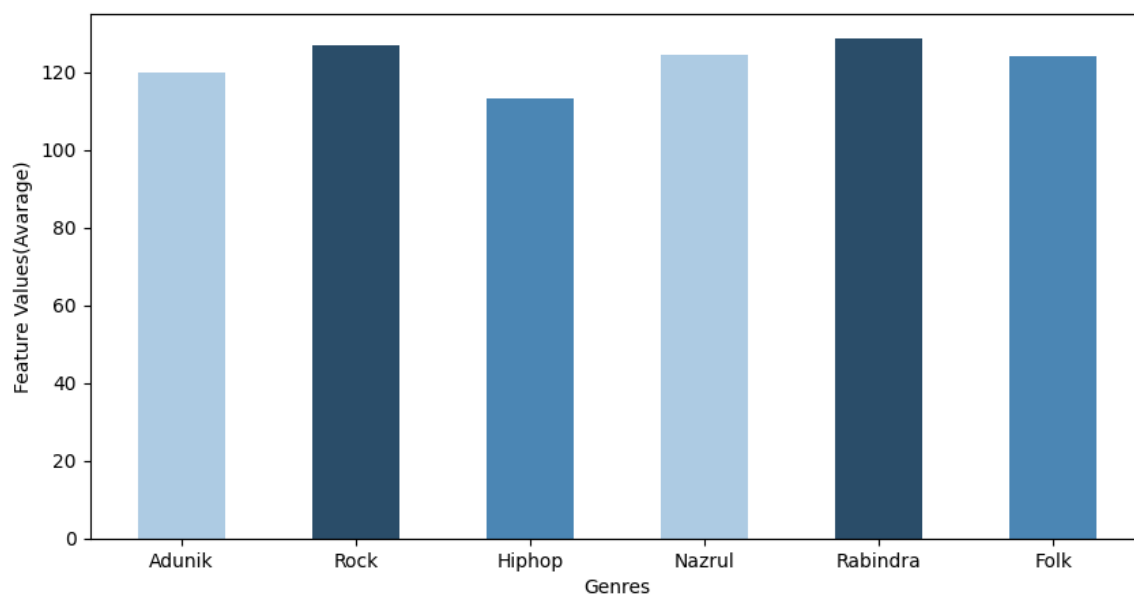


Figure 4. Mean tempo values

## 2.5. Data preprocessing

For making any machine learning or deep learning model, the voice feature needs to be preprocessed before using it as the input. Performance of the model is negatively impacted by distortion features. To lessen the distortion of features, feature scaling is required. Here, we apply a few straightforward transformations and condense the characteristics into a small range. We calculate this using (5):

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (5)$$

## 2.6. Machine learning model

After the feature extraction and feature preprocessing, initially we perform some machine learning algorithm to identify the song genre. We have 1820 song data and get 1820 features with 6 different genres. As we know MFCC gives 20 features, chroma frequency gives 1 one feature, tempo feature gives one feature, and RMSE gives 1 feature, so we got a total 23 features for each song. Using those features we tried to apply some traditional machine learning algorithm name k-nearest neighbors (KNN), support vector machine (SVM), decision tree, and random forest.

For applying those algorithms, we tried KNN with 'n neighbors = 5', decision tree classifier with 'random state=2', random forest classifier with 'n estimators = 30', SVM classifier with kernel= 'rbf'. But we are having complexity with traditional machine learning approaches more specifically that didn't get satisfactory accuracy. This is why we move into deep learning based algorithm named ANN.

## 2.7. Artificial neural network model

According to the quantity of data, label, the type of feature, the system's requirements, and the performance, we choose an ANN as our final algorithm. Here, the raw model is discussed together with the necessary amount of layers and optimization. To save time and resources, we take the optimal performance for our data together with model complexity. The ANN model for our song genre classification is shown in Table 2.

Table 2. Song genre recognition model

Surface layer	Layer details
1	: ADAM (learning rate)
2	: For 15 iterations in all batches do:
3	: Dense (Node, Activation)
4	: Dense (Node, Activation)
5	: Dense (Node, Activation)
6	: Dropout (Rate)
7	: Dense (Node, Activation)
8	: Dense (Node, Activation)
9	: Dropout (Rate)
10	: Dense (Node, Activation)
11	: Dense (Node, Activation)
12	: Dropout (Rate)
13	: Dense (Node, Activation)
13	: end for

### 2.7.1. Song genre recognition model structure

Song genre classification model uses a fully connected dense layer. A completely linked dense layer and several regularization techniques, such as batch normalization [21] and dropout [22], were applied in this model. Layer 1 has 128 dense nodes that have rectified linear unit (ReLU) activation as shown in (6). Layer 2 is dense with 256 nodes that have ReLU activation and layer 3 also dense with 256 nodes that have ReLU activation. Later, layer 5 with a 50% dropout was coupled to the output of these layers.

$$ReLU(Y) = \max(0, Y) \quad (6)$$

Next, layer 6 receives the output of layer 5. With 64 nodes, layer 6 is a dense layer. The output from layer 6 then moves to layer 7. With 64 nodes, layer 7 is a dense layer. Later, this layer's output was coupled to layer 8 with a 40% dropout. Layer 8's output is subsequently transferred to layer 9. With 64 nodes, layer 9 is a dense layer. The output from layer 9 enters layer 10. 32 nodes make up the dense layer 10. Later, this layer's output was connected to layer 11 with a 40% dropout. Layer 12 is the output layer for the mode which is coupled to layer 11's output by a dense layer 12 with 6 nodes and SoftMax (7) activation and the model architecture summary will help us in more clearly identifying the model. Table 3 provides a summary of the entire model for clear visualization.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \text{ for } j = 1, \dots, k \quad (7)$$

Table 3. Song genre recognition model structure summary

Layer no (type)	Output shape	Parameter	Connected to
(Input layer)	128	3072	-
1 (Dense)	256	33024	Input layer
2 (Dense)	256	65792	1
3 (Dropout)	256	0	2
4 (Dense)	64	16448	3
5 (Dense)	64	4160	4
6 (Dropout)	64	0	5
7 (Dense)	64	4160	6
8 (Dense)	32	2080	7
9 (Dropout)	6	0	8
10 (Dense)	6	198	9
Total params: 128,934			
Trainable params:128,934			
Non-trainable params: 0			

### 2.7.2. Optimizer and learning rate

With the support of the optimizer algorithm, researchers will reduce neural network algorithm error. Adam optimizer was utilized in the proposed classification model [23]. The majority of researchers use it to

boost performance. This optimizer has been modified with the help of a stochastic gradient descent algorithm. A crucial component of this optimizer's ability to tune hyper-parameters is updating network weight. The Adam optimizer (8) with a learning rate of 0.001 was employed for the proposed classification.

$$\theta_{t+1} = \theta_t - \frac{\eta}{1 + \sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (8)$$

In order to optimize the process, we used a function called categorical cross entropy (9) to calculate the error of our model. According to a study, mean squared error, classification error, and other loss-calculating functions have all been outperformed by cross entropy [24].

$$L_i = - \sum_j t_{i,j} \log(P_{i,j}) \quad (9)$$

CNN hyper-parameter tuning is significantly influenced by learning rate. Low learning rates are beneficial for accuracy because they move toward global optimums more gradually, reducing the possibility of overshooting. However, it takes a long time to get the global minima. Larger steps can be taken to resolve the time problem with a high learning rate, but there is a danger of overshooting the global minima. Additionally, accuracy and precision might be hampered. To solve this issue, the automatic learning rate reduction method [25] was employed. We initially specified a larger learning rate of 0.001 and modified it in response to the validation accuracy.

### 3. RESULTS AND DISCUSSION

#### 3.1. Performance of machine learning algorithm

When we prepared to implement the concept of recognizing the Bengali music song genre we tried some traditional machine learning algorithms. As we know we have 1800+ songs and we try to train the traditional machine learning model using 70% data and test the model using the rest of the 30% data. Table 4 shows the performance of different machine learning algorithms.

Table 4. Performance of different machine learning algorithms

Algorithm	Accuracy (%)
KNN	68
Logistic regression	69
Random forest	71
SVM	76
Decision tree	58

For song genre recognition the ANN model serves as our ultimate model. In section 2, we've already covered the model's structure and a description of the essential inputs we utilized to put it into practice. On a train set, test set, and validation set, the model produced encouraging results after being trained and validated on our dataset. We have a total data of 1800+ audio songs and we used 80% of our data to train the model, 10% to test it, and 10% to validate it. By enhancing several model layers and looking at the feature type of the audio data, we may also enhance the outcome. Finally, we can state that our effort was largely effective. The accuracy and loss of the song genre recognition model for the training and validation set are shown in Figure 5.

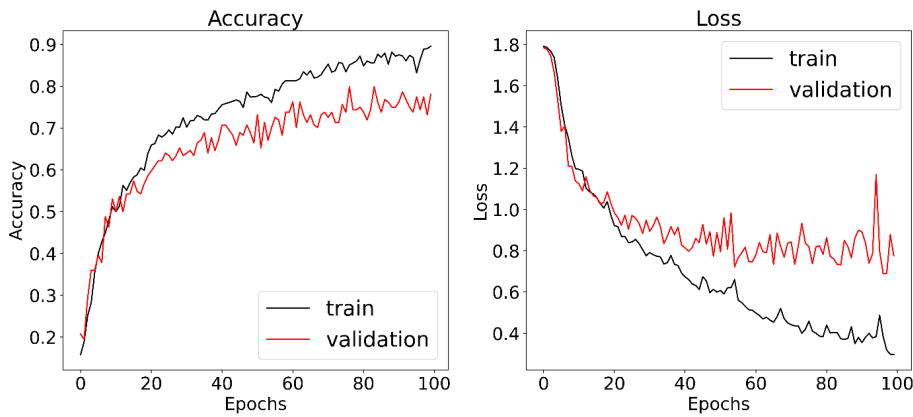


Figure 5. Training and validation accuracy and loss of song genre recognition model

After 100 epochs, the proposed model got 89.43% accuracy on the training set and 78.49% accuracy on a validation set of our song datasets and the confusion matrix for this model is shown in Figure 6. So, that everyone may better understand how the model performs. According to our result we compare our result with the previous best result and Table 5 shows the comparison between previous best accuracy and our proposed ANN accuracy and it clearly visible that out model gives better accuracy.

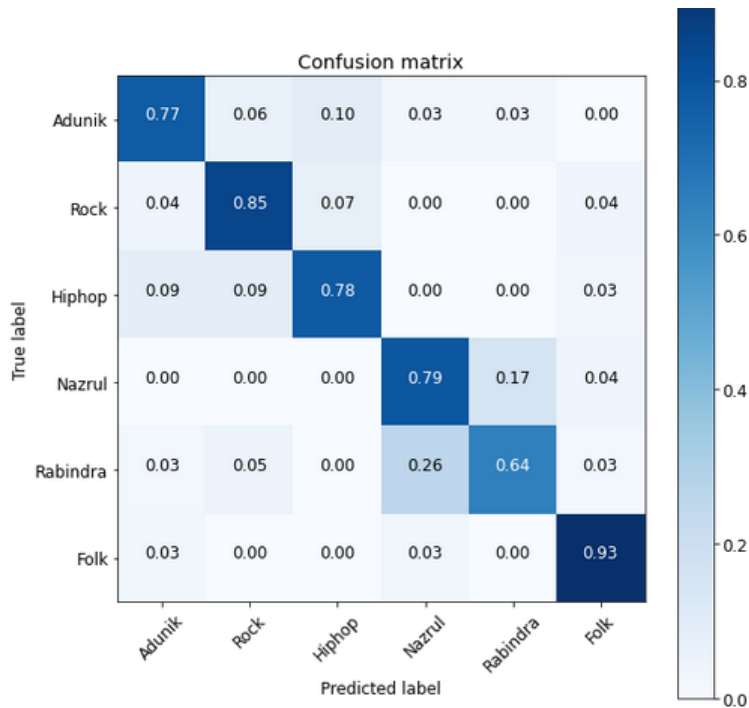


Figure 6. Confusion matrix of song genre recognition model

Table 5. Comparison between previous best accuracy and our proposed ANN accuracy

Previous best accuracy	Our proposed model accuracy
74%	78%



#### 4. CONCLUSION

For the Bengali people, Bengali is their most prized and valued language. This language generates a lot of memories and we are a very emotional nation. Since music enables us to experience all of our emotions, even a small contribution in this area makes us feel quite satisfied. We tried many approaches to extract music features and tried many approaches to train those features for making a better recognition model. We figured out that if we extracted the tempo feature then our music recognition model tends to be more accurate and we get better predictions with the help of the deep learning based algorithm. In our case, a deep learning based model called an ANN gives us 78% accuracy.




#### REFERENCES

- [1] K. C. Brata, M. H. Nugraha, and A. H. Brata, "An alternative M-learning scoring approach to automatically evaluate Iqro's pronunciation based on pitch, volume, and rhythm features," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 5, pp. 5313-5320, 2022, doi: 10.11591/ijece.v12i5.pp5313-5320.
- [2] R. Sarno, D. R. Wijaya, and M. N. Mahardika, "Music fingerprinting based on Bhattacharya distance for song and cover song recognition," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 2, pp. 1036-1044, 2019, doi: 10.11591/ijece.v9i2.pp1036-1044.
- [3] P. P. Rokade and A. K. D., "Business intelligence analytics using sentiment analysis-a survey," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 1, pp. 613-620, 2019, doi: 10.11591/ijece.v9i1.pp613-620.
- [4] A. Agrima, I. Mounir, A. Farchi, L. Elmazouzi, and B. Mounir, "Emotion recognition based on the energy distribution of plosive syllables," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 6, pp. 6159-6171, 2022, doi: 10.11591/ijece.v12i6.pp6159-6171.
- [5] D. Liang, H. Gu, and B. O'Connor, "Music genre classification with the million song dataset," *Final Report*, Carnegie Mellon University: Pittsburgh, USA, 2011.
- [6] C. Monterola, C. Abundo, J. Tugaff, and L. E. Venturina, "Prediction of potential hit song and musical genre using artificial neural networks," *International Journal of Modern Physics C*, vol. 20, no. 11, pp. 1697-1718, 2009, doi: 10.1142/S0129183109014680.
- [7] C. Sanden, C. Befus, and J. Z. Zhang, "Clustering-based genre prediction on music data," in *Proceedings of the 2008 C3S2E conference on - C3S2E '08*, 2008, pp. 117-119, doi: 10.1145/1370256.1370274.
- [8] E. Zangerle, R. Huber, M. Vötter, and Y.-H. Yang, "hit song prediction: leveraging low- and high-level audio features," in *Proceedings of the 20th ISMIR Conference*, pp. 319-326, 2019.
- [9] H. Bahuleyan, "Music genre classification using machine learning techniques" *arXiv-Computer Science*, pp. 1-12, 2018.
- [10] M. A. A. Mamun, I. Kadir, A. S. A. Rabby and A. Al Azmi, "Bangla music genre classification using neural network," *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, Moradabad, India, pp. 397-403, 2019, doi: 10.1109/SMART46866.2019.9117400.
- [11] H. Ali, M. F. Hossain, S. B. Shuvo, and A. A. Marouf, "Banglasenti: a dataset of Bangla words for sentiment analysis," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1-4, doi: 10.1109/ICCCNT49239.2020.9225565.
- [12] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimedia Tools and Applications*, vol. 82, no. 3, pp. 3713-3744, 2022, doi: 10.1007/s11042-022-13428-4.
- [13] S. Kibria, A. M. Samin, M. H. Kobir, M. S. Rahman, M. R. Selim, and M. Z. Iqbal, "Bangladeshi Bangla speech corpus for automatic speech recognition research," *Speech Communication*, vol. 136, pp. 84-97, 2022, doi: 10.1016/j.specom.2021.12.004.
- [14] M. I. Shuvo, S. A. Shahriyar, and M. A. Akhand, "Bangla numeral recognition from speech signal using convolutional neural network," *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2019, pp. 1-4, doi: 10.1109/ICBSLP47725.2019.201540.
- [15] M. M. Hasan, M. A. Islam, S. Kibria, and M. S. Rahman, "Towards lexicon-free Bangla automatic speech recognition system," *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2019, pp. 1-6, doi: 10.1109/ICBSLP47725.2019.201544.
- [16] M. R. Kotwal, F. Hassan, and M. N. Huda, "Speech feature evaluation for Bangla automatic speech recognition," *Technical Challenges and Design Issues in Bangla Language Processing*, Pennsylvania: IGI Global, pp. 169-208, 2013.
- [17] M. F. Hossain, M. A. Supto, Z. Chowdhury, H. S. Chowdhury, and S. Abujar, "BAAD: a multipurpose dataset for automatic bangla offensive speech recognition," *Data in Brief*, vol. 48, pp. 1-9, 2023, doi: 10.1016/j.dib.2023.109067.
- [18] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357-366, August 1980, doi: 10.1109/TASSP.1980.1163420.
- [19] T. Ganchev, N. Fakotakis, and G. Kokkinakis, "Comparative evaluation of various MFCC implementations on the speaker verification task," *Proceedings of the SPECOM*, 2005, pp. 1-4,
- [20] D. O'Shaughnessy, "Automatic speech recognition," in *Speech Communications: Human and Machine*, 2000, pp.367-435, doi: 10.1109/9780470546475.ch10.
- [21] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning*, pp. 448-456, 2015.
- [22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929-1958, 2014.
- [23] D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," in *International Conference on Learning Representations*, pp. 1-15, 2015.
- [24] K. Janocha and W. M. Czarnecki, "On loss functions for deep neural networks in classification," *Schedae Informaticae*, vol. 25, pp. 49-59, 2017, doi: 10.4467/20838476SI.16.004.6185




- [25] T. Schaul, S. Zhang, and Y. LeCun, "No more pesky learning rates," in *30th International Conference on Machine Learning*, 2013, pp. 343-351.

## BIOGRAPHIES OF AUTHORS






**Mariam Akter**    received the B.Sc. degree in Computer Science and Engineering from Daffodil International University, Dhaka, Bangladesh in 2023. Her research interests include natural language processing (NLP), machine learning, deep learning, and human computer interaction (HCI). She can be contacted at email: [mariam15-12754@diu.edu.bd](mailto:mariam15-12754@diu.edu.bd).






**Nishat Sultana**    received the B.Sc. degree in Computer Science and Engineering from Daffodil International University, Dhaka, Bangladesh in 2023. Her research interests include natural language processing (NLP), machine learning, deep learning, and human computer interaction (HCI). He can be contacted at email: [nishat15-12396@diu.edu.bd](mailto:nishat15-12396@diu.edu.bd).



**Sheak Rashed Haider Noori**    is a professor and associate head in the Department of CSE at the Daffodil International University, Dhaka. He received his Ph.D. in Information & Communication Technology from the University of Trento, Italy, and studied Masters in Software Engineering of Distributed System at KTH, Sweden. He has several publications in reputed journals and international peer reviewed conferences and book chapters and is a reviewer for a number of international journals and conferences. He has several years of experience working in various European and Asian IT companies and Universities. His work includes R&D, software development, training, and consulting. He also served as a Visiting Academic at Anglia Ruskin University, UK, and the University of West Of Scotland, UK. He can be contacted at email: [drnoori@daffodilvarsity.edu.bd](mailto:drnoori@daffodilvarsity.edu.bd).



**Md Zahid Hasan**    received a Ph.D. degree in Computer Science and Engineering from Jahangirnagar University. He is currently working as an associate professor with the Department of CSE, Daffodil International University, Bangladesh. His research interests include computer vision, health informatics, machine learning, artificial intelligence-based systems, and decision theory. He can be contacted at email: [zahid.cse@diu.edu.bd](mailto:zahid.cse@diu.edu.bd).