

# Machine Learning-Based Music Genre Classification with Pre-Processed Feature Analysis

Md Shofiqul Islam <sup>1</sup>, Md Munirul Hasan <sup>1</sup>, Md Abdur Rahim <sup>2</sup>, Ali Muttaleb Hasan <sup>1</sup>, Mohammad Mynuddin <sup>3</sup>, Imran Khandokar <sup>1</sup>, Md Jabbarul Islam <sup>4</sup>

<sup>1</sup> Faculty of Computing, Universiti Malaysia Pahang, 26600, Kuantan, Pahang, Malaysia.

<sup>2</sup> Department of Mechanical Engineering, College of Engineering, Universiti Malaysia Pahang, 26300 Gambang, Kuantan, Pahang, Malaysia

<sup>3</sup> Department of Civil, Transportation Engineering, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816 USA

<sup>4</sup> Department of Mathematics, National University, Gazipur-1704, Dhaka, India

## ARTICLE INFO

### Article history:

Received November 27, 2021  
Revised December 16, 2021  
Accepted January 18, 2022

### Keywords:

Artificial Intelligence;  
Classification;  
Filtering;  
Pre-processing;  
Machine learning;  
Music;  
Speech

## ABSTRACT

The growth of the entertainment industry around the world may be seen in the creation of new genres and the influx of artists and musicians into this field. Every day, a large amount of music is generated and released. The classification of this music based on genres and the recommendation of music to users is a crucial task for various music streaming platforms. Many artificial intelligence methods have been created to overcome this. Inadequate data for training is one of the biggest issues when it comes to building machine learning algorithms. A certain dataset contains a large number of redundant features, which may lead the models to overfit. Data filtering could be used to solve this issue. On the GTZAN data for music genre classification, this article constructed numerous Artificial Intelligence (AI) models and used a data filtering strategy. This study does a comparative analysis and discusses the results. The models developed and evaluated are Naive Bayes, Stochastic Gradient Descent, KNN, Decision trees, Random Forest, Support Vector Machine, Logistic Regression, Neural Nets, Cross Gradient Booster, Cross Gradient Booster (Random Forest), and XGBoost. Accuracy gained by Naive Bayes is 51.95%, Stochastic Gradient Descent 65.53%, KNN 80.58%, Decision trees 63.997%, Random Forest is 81.41%, Support Vector Machine 75.41%, Logistic Regression 69.77%, Neural Nets 67.73%, Cross Gradient Booster 90.22%, Cross Gradient Booster (Random Forest) 74.87%. Finally, XGBoost is the best-performed machine learning with an accuracy of 90.22%. This research outcome will help to analyze music in different areas.

This work is licensed under a [Creative Commons Attribution-Share Alike 4.0](https://creativecommons.org/licenses/by-sa/4.0/)



### Corresponding Author:

Md Shofiqul Islam, Faculty of Computing, Universiti Malaysia Pekan, 26600, Kuantan, Pahang, Malaysia.  
Email: [shafiqcseiu07@gmail.com](mailto:shafiqcseiu07@gmail.com)

## 1. INTRODUCTION

The most important and well-known element of any song is its genre. Many music fans make a playlist based on genre. Many music streaming apps, such as Apple Music, Spotify, Wynk, and others, suggest new songs to these customers based on their favorite genres. Because the music industry is rapidly evolving and establishing new song categorization can be challenging, technology can help solve this challenge. A song can be represented as an audio signal [1] with bandwidth, frequency, spectral roll-off, and other characteristics. The frequency under which a certain proportion of the total spectral energy is lost is referred to as spectral roll-off. The roll-off frequency is the frequency at which a certain percent of age (cutoff) of something like the spectrum's amount of energy is contained. The gap between the top and bottom frequencies in a continuous range of frequencies is known as bandwidth. It's possible to tell the difference between harmonics (below roll-off) and noisy noises using the roll-off wavelength (above roll-off). These characteristics identify two audio

transmissions. Using these qualities, we used the notion of data filtering to classify the songs into their various genres. The main issue with dealing with datasets is that the data will either be incomplete, i.e. incorrect data, or it will be outlier values. As a result, pre-processing of this data requires data filtering. Data filtering refers to a variety of approaches or tactics for fine-tuning datasets to meet the needs of users. Data filtering removes any noisy, undesirable, and duplicate values from the information, resulting in clean, ready-to-process data. To acquire the best result and show how data filtering improves accuracy, a comparison study is carried out.

Consumers can better grasp social networking sites and feedback by mining public emotions. Sentiment analysis is a type of computer-based mining that saves time and money by applying various voice analysis technologies. As the human brain learns from sample data, a machine learning system predicts fresh unlabeled data quickly. Machine learning algorithms outperform alternative methods or approaches based on handcraft aspects. There has been a lot of research and development in the field of voice classification using machine learning, and the results have been rather good. Some of the earlier systems have successfully achieved qualitative and meaningful knowledge; nevertheless, they are limited to coherence, and information extraction presents a significant challenge [2]. In terms of accuracy and performance, deep attempting to learn procedures, on the other hand, outperform other similar techniques. In the classification of image [3][4], video [5][6], speech [7], and text [2][8][9][10], IoT based analysis [11][12], deep learning outperformed. Several neural network models have already been proved superior to others. For example, Convolution Neural Network (CNN) [13], Long Short Term Memory (LSTM) [14], Bidirectional LSTM [15], Gated Recurrent Unit (GRU) [16] have improved textual grading significantly.

The remainder of the paper is laid out as follows: Section 2 formulates the problem and examines the context. Section 3 is for data and its preprocessing. The experimental technique and development are detailed in Section 4. The results are presented in Section 5, followed by Section 6 provides the conclusion.

## 2. RELATED STUDY

The music industry has grown rapidly during the previous decade, with new songs produced every day and new artists, bands, and songwriters joining the sector. This has resulted in a fascinating subject of study for scholars. Deep learning solutions have been employed by researchers for music genre recognition and music recommender systems. Using the GTZAN dataset, the author [17] created a residual neural network to train on audio snippets of 3 seconds duration. Certain qualities that overlapped were taken into account for various genres, and the author was able to obtain a 94% accuracy rate. For categorization, many writers have created convolution neural networks (CNN). Since spectrograms, chromatograms, and Mel-frequency cepstral coefficient are three alternative ways to visualize music (MFCC). The mel-spectrogram of the audios was utilized as input for the author's models in [18]. The researcher used a double convolution layer that went through multiple layers as well as data analysis. In [19], a comparison of two model classes that traveled through a CNN architecture in the sound and harmonic domains spectrum analyzer was performed (VGG-16). The author achieved a precision of 65 %. Author [20] conducted a comparative assessment of the benefits and drawbacks of various filter algorithms, using six different datasets for their research. In addition, the author created a new hybrid algorithm to improve feature selection. For predicting, the author [21] utilized various data filtering approaches on neural networks, with local regression filtering yielding the best results. For packaging the functionality, the author suggests an ensemble-based technique [22]. The data set was a very imbalanced distribution of the class, with the majority of the work being the development, evaluation, and application of several datasets. Traditional approaches failed to perform as well as the proposed method. The author [23] constructed a wrapper approach by extending an existing filter method and testing it on a variety of datasets. Another successful approach with some machine learning approaches and data filtering techniques, this method uses MLP tools and got 88.10% accuracy [24]. Another method used Single-Layer Feedforward Neural Network (SFNN). This method got 84.83% accuracy. A more successful method using Deep Attention Neural Network (DANN) got an accuracy of 90.00%.

The results of the preceding investigation reveal that present music analysis algorithms have issues with context and characteristics. It's envisaged that a new multi-label music analysis system with improved performance and adaptability will emerge. However, in order to design a current and highly accurate machine learning system for music analysis, we need to solve mentioned limitations of the existing methods. As a result, three objectives have been established to fulfill the research goal by resolving the problems:

1. To create a flexible, adaptive, and high-performance hybrid profound training approach with multi-label music analysis knowledge.
2. To propose a new technique to analyze the music using the hybrid machine learning algorithm.
3. To assess the performance of the proposed algorithm on the GTZAN dataset using current methodologies.

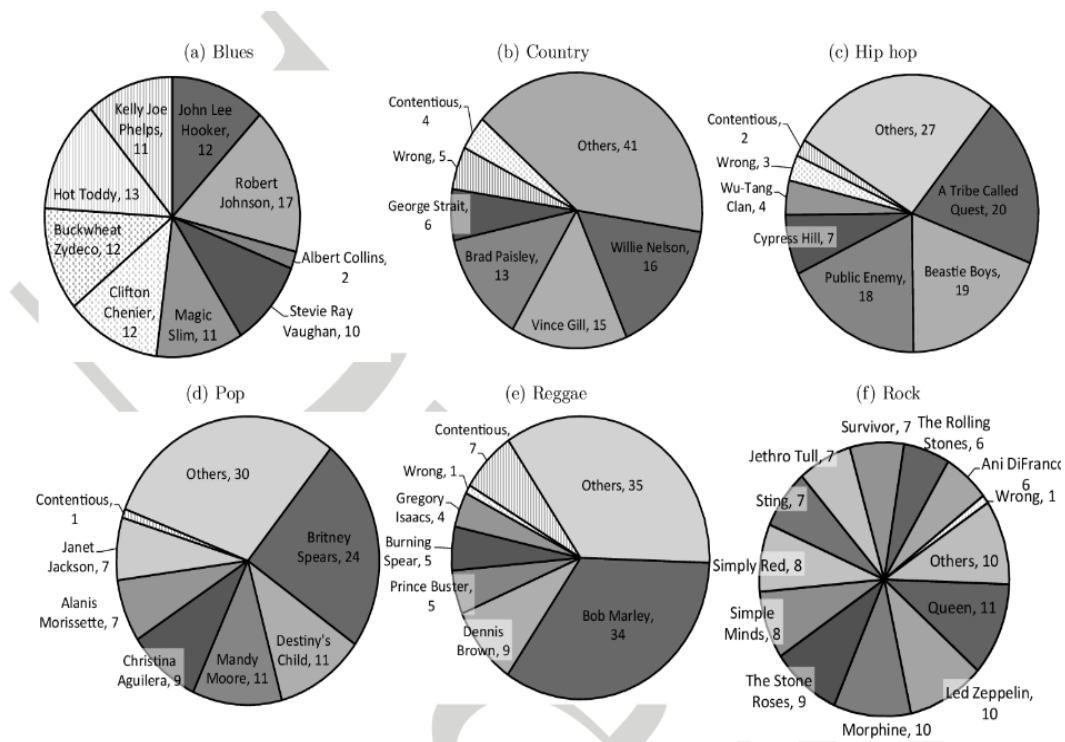
3. DATA

We employ GTZAN music information in our scientific implementation. The GTZAN dataset [25] is the most often utilized for music genre categorization. The GTZAN dataset has ten genres and 100 audio files for each class with 30 seconds duration. In this study, we used the GTZAN dataset, whose frequency is presented in Table 1.

**Table 1.** Overall summary of GTZAN Datasets.

Music Class	Number of classes label
Pop	1000
Reggae	1000
Metal	1000
Jazz	1000
Blues	1000
Disco	1000
Classical	1000
Hip hop	1000
Rock	1000
Country	1000

Fig. 1 displays the data representation of music amount in GTZAN Data. The amount of color labeled indicates the number of music separately exist in the dataset. As shown in Fig. 1, Pie chart visualization of GTZAN data, (a) Blues, (b) Country, (c) Hip hop, (d) Pop, (e) Reggae, and (f) Rock music amount with %age [25].



**Fig. 1.** Pie chart visualization of GTZAN data, (a) Blues, (b) Country (c) Hip hop, (d) Pop (e) Reggae, and (f) Rock music amount with percent of age.

Here, we have shown the Principal Component Analysis (PCA) illustration of Genre music. All colors indicate the type of music, and its label is given in the figure. PCA in Fig. 2 illustrated to visualize possible groups of genres.

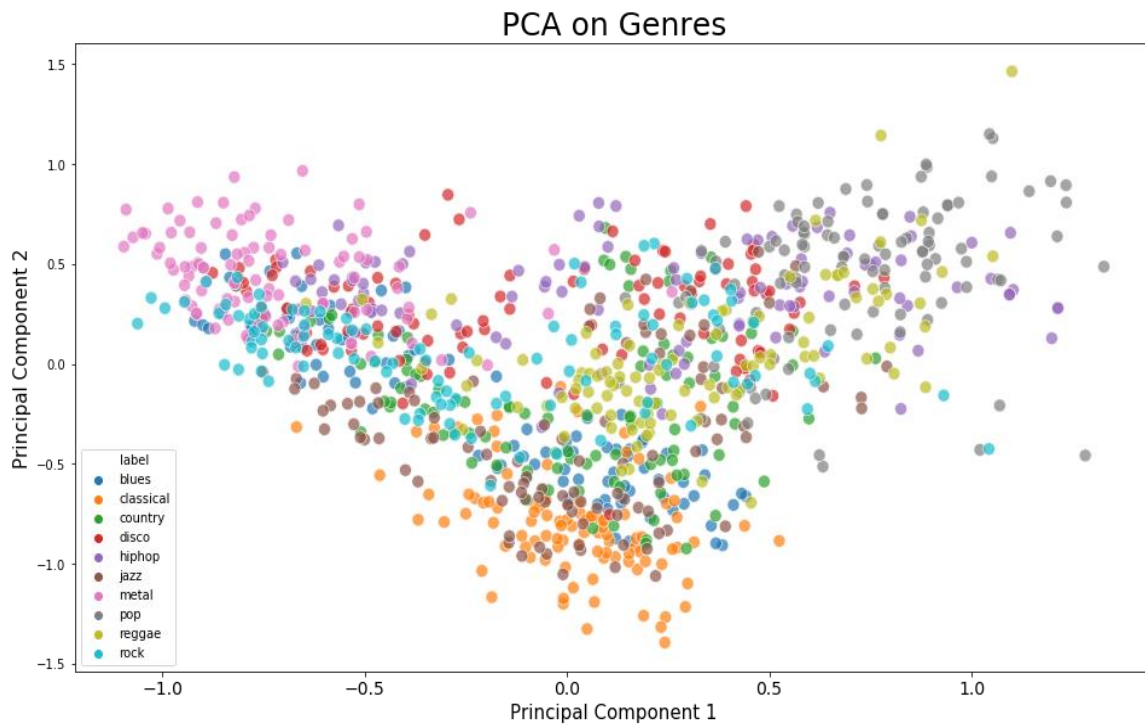


Fig. 2. PCA illustration of Genre music.

### 3.1. Preprocessing

When visualized, each audio can be represented as a sound signal [24]. This signal allows us to discriminate between two audios. As a result, we were able to extract several properties from these audio waves [24]. To tackle the problem, relevant audio features can be extracted [26]. Two subcategories of acoustic files can be found.

#### 3.1.1. Feature Extraction

##### a) Time Domain Feature

There are two of the Time Domain Features we used in our method. Point “a” state Root Mean Square Energy and the point “b” state definition of point “a.”

##### a. Root Mean Square Energy:

The Root Mean Square Energy(RMSE) determines how loud the audio is (RMSE). Because RMSE is computed frame by frame, the mean, standard deviation is calculated across all frames. The energy of the Root Mean Square is illustrated in the equation (1), where  $x$  is a set of  $N$  values:

$$x_{rmse} = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + x_3^2 + \dots + x_N^2)} \quad (1)$$

##### b. Zero-Crossing Rate:

The signaling rate, or the number of times the sign shifts from positive to negative and, conversely, is altered from positive to negative or negative to positive. The implementation of Zero Crossing Rate(ZCR) is done using equation (2). Let signal  $s$  with the time duration  $T$ , ZCR will be as follows:

$$ZCR = \frac{1}{T} \sum_{t=1}^T |s(t) - s(t-1)| \quad (2)$$

##### b) Frequency Domain Features

We use some frequency domain features in our method. These are listed below:

- a. Mel-Frequency Capstral Coefficient: The Mel-Frequency Capstral Coefficient is a set of factors (approximately 10-20) that describe the contour of an audio stream.
- b. Chroma Features: Each pitch class, i.e., the 12 unique semitones of the musical octave, has a feature vector of size 12. It establishes the degree of similarity between musical works.
- c. Spectral Centroid: It denotes the signal's 'center of mass,' or the wavelength at which the spectrum's energy is centered.
- d. Spectral Roll-off: It defines the signal's form. The frequency where the frequency components are set to 0 is determined by this variable.

Fig. 3 shows the Metal music Mel Spectrogram visualization in 30 seconds time duration.

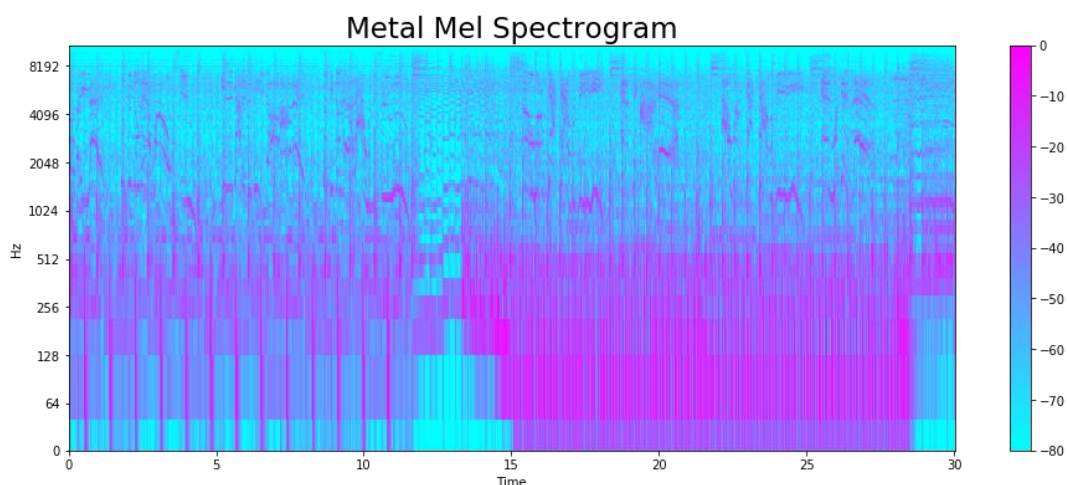


Fig. 3. Metal Mel Spectrogram.

### 3.1.2 Data Filtering

The basic objective of the feature is to exclude predictors that aren't useful or repetitive from the model. Data filtering is critical in machine learning [24]:

#### a) Parsimony (or Simplicity)

Advanced forms are harder to interpret than simple ones, especially while inferring.

#### b) Time is Money

As the amount of features is lower, the computation time falls, and thus the training time reduces.

#### c) Avoiding the Curse of Dimensionality

A highly accurate model trained with a big dataset can be deceiving since it may indicate overfitting and will not generalize to samples.

#### 3.1.2.1 Data Filtering Methods

There are few data filtering methods in speech preprocessing. Some of it's we used in our method are listed below:

##### a) Filter Methods

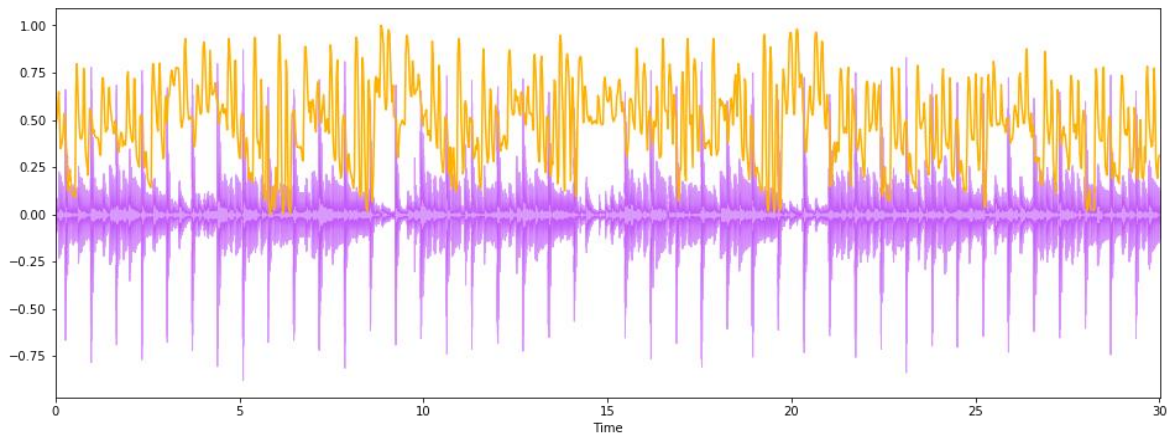
Filter approaches employ statistical calculations to assess the importance of predictors outside of prediction models and keep only those that meet a set of criteria. The kinds of data included, both in predictors and outcomes - numeric and categorical are factors to take into account when selecting filter techniques [24].

##### b) Wrapper Methods

Wrapper approaches analyze certain machine learning techniques to determine the best features. Wrapper methods for increased datasets take a long time to compute. Because machine learning models are trained with diverse combinations and features, there is a considerable risk of overfitting. Forward Selection (starts with one prediction and includes additional repeatedly), Backward Selection (starts with all indicators and removes one by one iterative manner), and Step-Wise Selection (starts within all predictors and removes one by one iteratively) are the three potential directions of processes (bi-directional) [24].

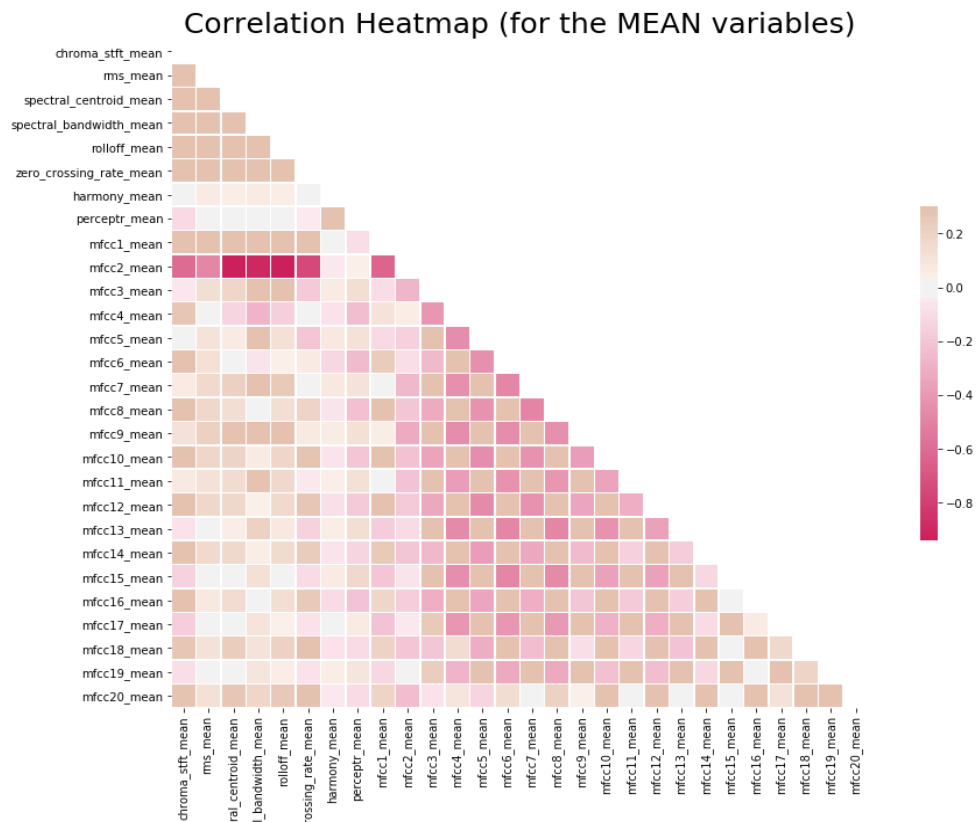
**c) Embedded Methods**

The characteristics of both the wrapper feature selection methods are combined in reinforcing elements. The classification algorithm in their models happens automatically throughout the model fitting procedure. The most common embedding method uses tree-based techniques, such as a decision tree as well as an alleged woodland. The general concept of function choice in the split node is related to information acquisition. Other embedded approaches include the LASSO charge and the L2 penalty for Ridge for creating a linear model [24]. Fig. 4 shows the spectral roll-off, which is a measurement of the signal's form. It denotes the frequency at which a certain proportion of overall spectral energy, such as 85%, is found.



**Fig. 4.** Spectral Roll off of speech signal.

In Fig. 5, we have added a Correlation Heat map for feature means, all mean, or other features are selected based on pre-processed features.



**Fig. 5.** Correlation Heat map for feature means

#### 4. METHOD

The fundamental model proposed consists of multiple parts. First, it collects data from a source and then pre-processes it in various ways mentioned in the previous section. Then, using the proposed methods, feed the which was before speech signal. Following the execution of many algorithms, the final prediction and analysis are completed. Certain genres have always been easily classed by all of the modeling techniques, while others, such as hip-hop as well as disco, were more difficult to categorize because the sound of these art forms had characteristics of other genres, i.e., the audio was a mix of multiple genres rather than a single genre, making it difficult to classify. Several artificially intelligent algorithms were presented, in which input was transmitted without even any filter techniques have been applied first, and with filter acts on the object second. When data filtering procedures were employed vs. when they weren't, there was a considerable difference in accuracy. The models were developed and evaluated with ten traditional machine learning algorithms. Finally, they compared their results to recommend the best algorithm in music classification.

After the data preprocessing, the data is partitioned into two parts, 80% for training and 20% for validation. Implementing the proposed model is a crucial task, and then the result is analyzed to predict music class. The main operational flow chart of our model is shown in Fig. 6.

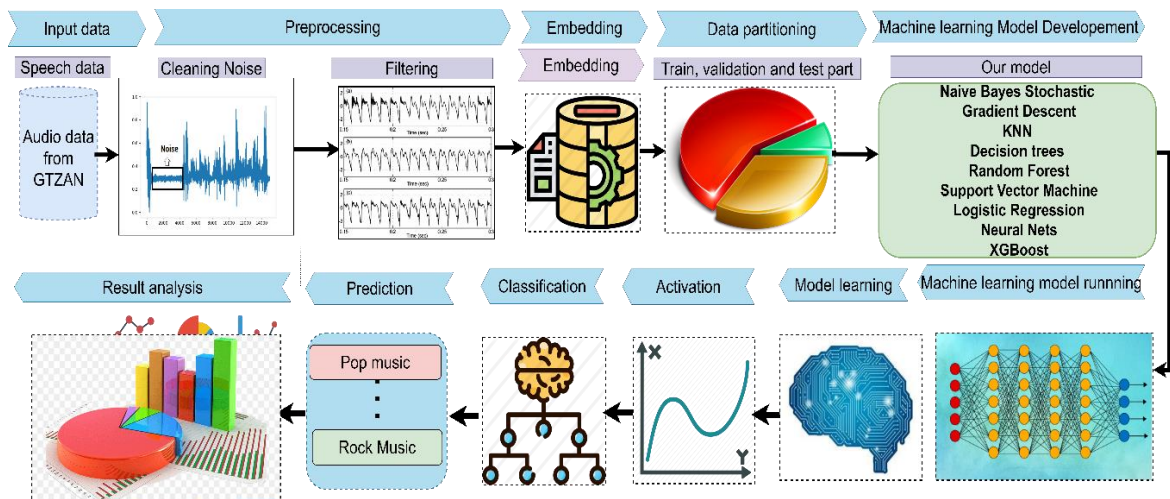


Fig. 6. Proposed machine learning model to classify music.

##### a) Naive Bayes

A Naive Bayes classifier is built using the Bayes Theorem. It estimates member probabilities for each classification, such as the probability that a particular record or data point belongs to that group. The most probable class is the one with the greatest probability.

##### b) Stochastic Gradient Descent

The iterative approach of stochastic gradient descent (commonly shortened SGD) is used to optimize an objective function with sufficient smoothness criteria (e.g., differentiable or sub differentiable). Because it substitutes the actual gradients (derived from the complete data set) with an approximation, it can be considered stochastic approximations of steepest descent efficiency (calculated from a randomly selected subset of the data). This minimizes the computing cost, especially for high optimization algorithms, allowing for quicker iterations in exchange for a reduced convergence speed.

##### c) K-Nearest Neighbors (KNN)

The supervised learning method k-nearest neighbors (KNN) is a simple, easy-to-implement technique that may be used to address both classification and regression issues. The KNN algorithm believes that objects that are similar are close together. To put it another way, related items are close together. A clustering procedure known as K Means separates observations into k groupings. Because the number of nodes may be specified, it can be simply applied to categorization when data is divided into clusters that are equal to or greater than the number of classes. The k-means clustering technique tries to divide a given anonymous data set (one that has no information about class identity) into a collection of k clusters. A total of k centroids are picked at the start. A centroid is an imaginary or actual piece of information in the center of a clustering.

**d) Decision trees**

A binary tree is used to depict the pattern. The very first node was indeed the root node. A flow path is constructed by a series of dataset-related inquiries (character traits as well as accompanying attributed information), the pathway is divided up, and the outcome is then anticipated. The many dataset classes are represented by the leaves. The decision tree is mostly employed in multiclass categorization.

**e) Random Forest**

Because of the large number of decision trees involved in the process, random forests are regarded as a highly robust and accurate technique. It is not affected by the problem of overfitting. The fundamental reason for this is that it averages all of the forecasts, canceling out any biases.

**f) Support Vector Machine**

Every data item is plotted as a spot in n-dimensional dimensions (where n is the number of characteristics you have), with both the value of each feature becoming the result of a certain position in the SVM algorithm. Then we accomplish classification by locating the hyperplane that best distinguishes the two classes. Even though SVM is commonly used only for the classification model, the result for all ten classes was calculated using the single-vs-rest approach. Each class is independently assessed, and the most engaged class is chosen.

**g) Logistic Regression**

A classification model is a linear model. Logistics regression is a statistical analytical method that forecasts data based on past dataset observations. Even though the expansions are more complicated, it exploits a logistical feature to represent a binary variable.

**h) Neural Nets**

A machine learning model is based on brain activity. Neurons, including such central nervous systems, are present in each step of the network. Whenever information passes thru these synapses, the model is capable of understanding it and providing the desired result. A complex link between outputs and inputs is often used as a non-linear framework.

**i) XGBoost**

The gradient boosted trees technique is designed in XGBoost, a popular and efficient open-source implementation. The weak learners in gradient boosting for prediction are regression trees, and each regression tree transfers input data points to one of its leaves containing a continuous value.

**4.1. Evaluation Metrics**

For the training of the model, validation, and simulation time with implementation, we find measures named accuracy in the evaluation. The operational formulae for these are given below. We use curvature values, mean correctness, mean recall, and a means F1 score of receiving operational properties to assess the determine incidence of the architectural outputs in this categorized risks (ROC). The submitted equation's correctness is evaluated in equation (3).

$$Accuracy = \frac{\text{Number of exactly matched instances}}{\text{Total instances}} \quad (3)$$

**4.2. Main algorithm to classify Music**

On GTZAN information, this section explains how the core algorithm works. The algorithm accepts raw information from the database and forecasts target mood categories based on data. Our model's operating processes at each phase are detailed in this core algorithm. After preprocessing raw voice or audio, it takes input and forwards it to multiple levels.

Algorithm1():

Begin  
Input: Raw data  
Output: Predicted class of music  
Initialization of parameters: Class C:=10  
Partition of data  
Take single speech S from data  
Pre-process speech S



```

Set model parameter initially
Number of models M:=10, Number of exactly matched instances:=0
Set number of model M
For I in range (1, M)
  Train I-th model
  P:=Predict test S data with I-th model
  If P is true for any of C class
    Number of exactly matched instances:= Number of exactly matched instances+1
    Evaluate performance metrics of the I-th model
    Accuracy={(Number of exactly matched instances)}/{(Total instances)}
  End If
End For
End

```

### 4.3 Experimental setup

The model was developed on a laptop with a good Internet connection. Google Colab was implemented, which is a free cloud-based service for machine research and education. The Jupyter interface is identical. It comes with a fully configured environment for in-depth research and free access to a powerful GPU. The device acceleration was chosen as the standard-setting for the remainder of the options. The categorization was carried out on a computer with five Intel(R) 3.60 GHz processors, 16 GB of RAM, and Windows 10 installed. The better outcomes from all experimental experiments are shown in [Table 2](#) and [Table 3](#) of the results.

## 5. RESULT

Predefined genres were easily classified by all of the modeling techniques, while others, such as hip-hop and disco, were more difficult to categorize because the sound of these genres had characteristics of other genres, i.e., the audio was a mix of multiple genres rather than a single genre, making it difficult to categorize. Several artificial intelligence algorithms were presented, in which input was passed without any filtering techniques have been applied first, and with filtering products which contain second. When data filtering procedures were employed vs. when they weren't, there was a considerable difference in accuracy after the pre-processing of signal in various ways mentioned in the previous section. The models developed and evaluated are Naive Bayes, Stochastic Gradient Descent, KNN, Decision trees, Random Forest, Support Vector Machine, Logistic Regression, Neural Nets, Cross Gradient Booster, Cross Gradient Booster (Random Forest), and XGBoost. Accuracy gained by Naive Bayes is 51.95%, Stochastic Gradient Descent 65.53%, KNN 80.58%, Decision trees 63.997%, Random Forest is 81.41%, Support Vector Machine 75.41%, Logistic Regression 69.77%, Neural Nets 67.73%, Cross Gradient Booster 90.22%, Cross Gradient Booster (Random Forest) 74.87% and XGBoost is the best-performed machine learning with an accuracy of 90.22%.

[Table 2](#) shows a comparison of the current model with a previously constructed model. Because some authors utilized a different dataset or generated one from scratch, some people have only taken 5–6 classes in order to be classified. As a result, our work is put up for comparison, with all ten classes of the GTZAN dataset factored into the model architecture.

**Table 2.** Comparative accuracy analysis of existing models

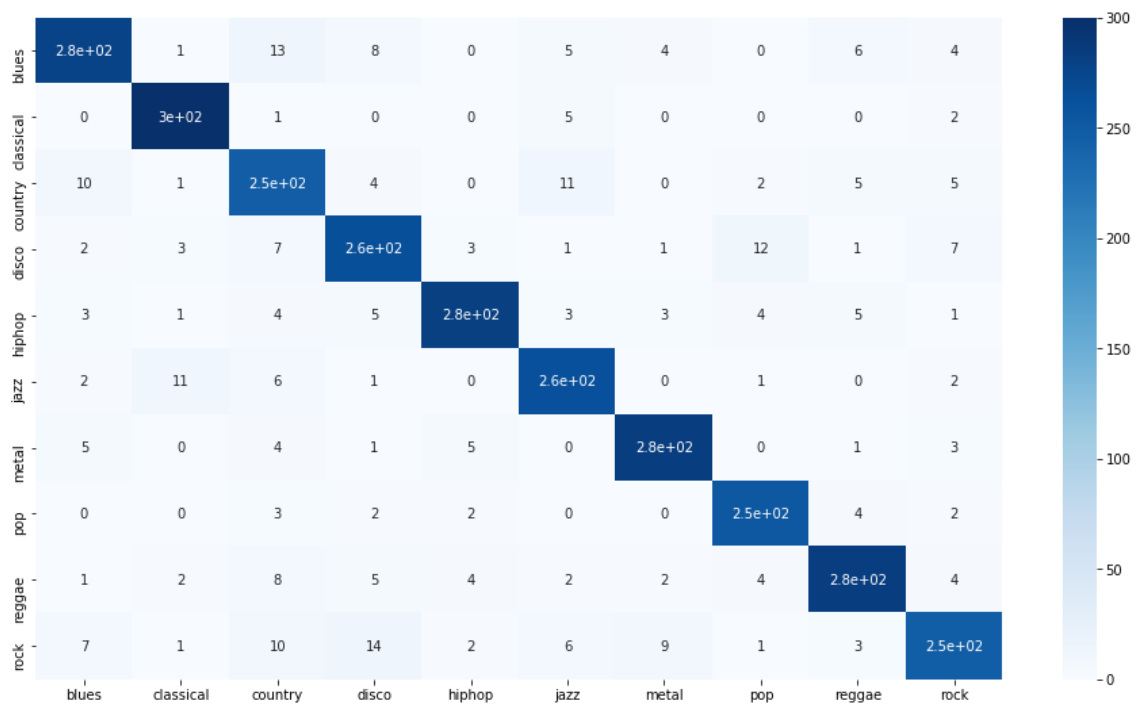
Model Authors	Model Name	Accuracy
[18]	Convolutional Neural Network (CNN)	65.00%
[27]	Single-Layer Feedforward Neural Network (SFNN)	84.83%
[25]	Recurrent Neural Network (RNN)	64.00%
[28]	Deep Belief Neural Network (DBNN)	84.30%
[29]	Support Vector Machine (SVM)	84.40%
[24]	Multi-Layer Perception (MLP)	88.10%
[30]	Deep attention Neural Network (DANN)	90.00%
<b>Proposed</b>	<b>XGBoost</b>	<b>90.22%</b>

We have implemented ten machine learning models by developing them, and our analyzed result is presented here. [Table 3](#) shows the comparative accuracy analysis of the proposed ten models. Based on the metric evaluation performance, it is clearly shown that the **XGBoost algorithm** performs better than other algorithms.

Here, Fig. 7 shows the correlation matrix of the XBoost algorithm result on the target result. Here the number of the target class is ten, and it is illustrated. In this figure, the relation between the target music class is clearly shown. This figure is designed from the test data result by the XBoost algorithm. For testing purposes, we took 20% of the data.

**Table 3.** Performance of different proposed machine learning algorithms on GTZAN data.

Algorithm Name	Accuracy
Naive Bayes	51.95%
Stochastic Gradient Descent	65.53%
KNN	80.58%
Decision trees	63.99%
Random Forest	81.41%
Support Vector Machine	75.41%
Logistic Regression	69.77%
Neural Nets	67.73%
Cross Gradient Booster (Random Forest)	74.87%
<b>XGBoost</b>	<b>90.22%</b>



**Fig. 7.** Correlation matrix of XBoost algorithm for the result among target result.

## 6. CONCLUSION

The suggested research uses the GTZAN dataset to construct multiple AI models for music genre classification. The audio characteristics for all of the audio recordings were included in the dataset. Several artificial intelligence methods were presented, in which input was passed without any filtering techniques have been applied first, and then with filtering techniques applied second. When data filtering procedures were employed vs. when they weren't, there was a significant difference of 2% in accuracy. The models developed and evaluated are Naive Bayes, Stochastic Gradient Descent, KNN, Decision trees, Random Forest, Support Vector Machine, Logistic Regression, Neural Nets, Cross Gradient Booster, Cross Gradient Booster (Random Forest), and XGBoost. Accuracy gained by Naive Bayes is 51.95%, Stochastic Gradient Descent 65.53%, KNN 80.58%, Decision trees 63.997%, Random Forest is 81.41%, Support Vector Machine 75.41%, Logistic Regression 69.77%, Neural Nets 67.73%, Cross Gradient Booster 90.22%, Cross Gradient Booster (Random Forest) 74.87% and XGBoost is the best-performed machine learning with the accuracy of 90.22%. In the future, we will improve our model to handle more classes of music on diverse datasets. We have also planned to handle real-time music or live music with less computational complexity.

## ACKNOWLEDGEMENT

This research is helped as a part by grants from FRGS by the Government of Malaysia to Universiti Malaysia Pahang (UMP). The grant numbers are FRGS/1/2018/ICT02/UMP/02/15 and FRGS/1/2019/ICT02/UMP /02/8.

## REFERENCES

- [1] M. McKinney and J. Breebaart, "Features for audio and music classification," 2003. <https://ismir2003.ismir.net/papers/McKinney.pdf>
- [2] M. S. I. Shofiqul, N. Ab Ghani, and M. M. Ahmed, "A review on recent advances in Deep learning for Sentiment Analysis: Performances, Challenges and Limitations," *COMPUSOFT: An International Journal of Advanced Computer Technology*, vol. 9, no. 7, pp. 3768-3776, 2020. <https://ijact.joae.org/index.php/ijact/article/view/1175>
- [3] I. Khandokar, M. Hasan, F. Ernawan, S. Islam, and M. Kabir, "Handwritten character recognition using convolutional neural network," in *Journal of Physics: Conference Series*, 2021, vol. 1918, no. 4: IOP Publishing, p. 042152. <https://doi.org/10.1088/1742-6596/1918/4/042152>
- [4] Purwono, A. Ma'arif and A. Wulandari, "Face Shape-Based Physiognomy in LinkedIn Profiles with Cascade Classifier and K-Means Clustering," *2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2021, pp. 347-353. <https://doi.org/10.23919/EECSI53397.2021.9624262>
- [5] M. S. Islam, S. Sultana, U. kumar Roy, and J. Al Mahmud, "A review on Video Classification with Methods, Findings, Performance, Challenges, Limitations and Future Work," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 6, no. 2, pp. 47-57, 2020. <https://doi.org/10.26555/jiteki.v6i2.18978>
- [6] M. S. Islam, S. Sultana, and M. J. Islam, "New Hybrid Deep Learning Method to Recognize Human Action from Video," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 7, no. 2, pp. 306-313, 2021. <https://doi.org/10.26555/jiteki.v7i2.21499>
- [7] L. Fan, Z. Yin, H. Yu, and A. Gilliland, "Using Data-driven Analytics to Enhance Archival Processing of the COVID-19 Hate Speech Twitter Archive (CHSTA)," *PrePrint*, 2020. <https://doi.org/10.31229/osf.io/gkydm>
- [8] M. M. Hasan, M. S. Islam, S. A. Bakar, M. M. Rahman, and M. N. Kabir, "Applications of Artificial Neural Networks in Engine Cooling System," in *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM)*, IEEE Computer Society, pp. 471-476, 2021. <https://doi.org/10.1109/ICSECS52883.2021.00092>
- [9] M. S. Islam, S. Sultana, U. K. Roy, J. Al Mahmud, and S. Jahidul, "HARC-New Hybrid Method with Hierarchical Attention Based Bidirectional Recurrent Neural Network with Dilated Convolutional Neural Network to Recognize Multilabel Emotions from Text," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 7, no. 1, pp. 142-153, 2021. <https://doi.org/10.26555/jiteki.v7i1.20550>
- [10] M. S. Islam and N. A. Ghani, "A Novel BiGRUBiLSTM Model for Multilevel Sentiment Analysis Using Deep Neural Network with BiGRU-BiLSTM," Singapore, 2022: Springer Singapore, pp. 403-414. [https://doi.org/10.1007/978-981-33-4597-3\\_37](https://doi.org/10.1007/978-981-33-4597-3_37)
- [11] M. A. Rahim, M. Rahman, M. A. Rahman, A. J. M. Muzahid, and S. F. Kamarulzaman, "A Framework of IoT-Enabled Vehicular Noise Intensity Monitoring System for Smart City," *Advances in Robotics, Automation and Data Analytics: Selected Papers from ICITES 2020*, vol. 1350, p. 194, 2021. [https://doi.org/10.1007/978-3-030-70917-4\\_19](https://doi.org/10.1007/978-3-030-70917-4_19)
- [12] L. C. Kiew, A. J. M. Muzahid, and S. F. Kamarulzaman, "Vehicle Route Tracking System based on Vehicle Registration Number Recognition using Template Matching Algorithm," in *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM)*, pp. 249-254. <https://doi.org/10.1109/ICSECS52883.2021.00052>
- [13] Y. Kim, "Convolutional neural networks for sentence classification," *MS thesis. University of Waterloo*, 2014. <https://core.ac.uk/reader/144148500>
- [14] J. Xu, D. Chen, X. Qiu, and X. Huang, "Cached long short-term memory neural networks for document-level sentiment classification," *arXiv preprint arXiv:1610.04989*, 2016. <https://doi.org/10.18653/v1/D16-1172>
- [15] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Twenty-ninth AAAI conference on artificial intelligence*, 2015. <https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/viewPaper/9745>
- [16] L. LI, A. ZHOU, Y. LIU, S. QIAN, and H. GENG, "Aspect-based sentiment analysis based on dynamic attention GRU," *Scientia Sinica Informationis*, vol. 49, no. 8, pp. 1019-1030, 2019. <https://doi.org/10.1360/N112018-00280>
- [17] D. Bisharad and R. H. Laskar, "Music Genre Recognition Using Residual Neural Networks," in *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, IEEE, pp. 2063-2068, 2019. <https://doi.org/10.1109/TENCON.2019.8929406>
- [18] H. Yang and W.-Q. Zhang, "Music Genre Classification Using Duplicated Convolutional Layers in Neural Networks," in *INTERSPEECH*, 2019, pp. 3382-3386. <https://doi.org/10.21437/Interspeech.2019-1298>
- [19] H. Bahuleyan, "Music genre classification using machine learning techniques," *arXiv preprint arXiv:1804.01149*, 2018. <https://arxiv.org/abs/1804.01149>
- [20] S. Das, "Filters, wrappers and a boosting-based hybrid for feature selection," *Proceedings of the Eighteenth International Conference on Machine Learning*, pp. 74-81, 2001. <https://dl.acm.org/doi/abs/10.5555/645530.658297>

- [21] W. Wettyaprasit, N. Laosen, and S. Chevakiadagarn, "Data filtering technique for neural networks forecasting," in *Proceedings of the 7th WSEAS International Conference on Simulation, Modelling and Optimization*: World Scientific and Engineering Academy and Society (WSEAS), pp. 225-230, 2007. <https://dl.acm.org/doi/abs/10.5555/1353862.1353902>
- [22] P. Yang, W. Liu, B. B. Zhou, S. Chawla, and A. Y. Zomaya, "Ensemble-based wrapper methods for feature selection and class imbalance learning," in *Pacific-Asia conference on knowledge discovery and data mining*: Springer, pp. 544-555, 2013. [https://doi.org/10.1007/978-3-642-37453-1\\_45](https://doi.org/10.1007/978-3-642-37453-1_45)
- [23] L. Talavera, "An evaluation of filter and wrapper methods for feature selection in categorical clustering," in *International Symposium on Intelligent Data Analysis*: Springer, pp. 440-451, 2005. [https://doi.org/10.1007/11552253\\_40](https://doi.org/10.1007/11552253_40)
- [24] A. Ghildiyal and S. Sharma, "Music Genre Classification Using Data Filtering Algorithm: An Artificial Intelligence Approach," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*: IEEE, pp. 922-926, 2021. <https://doi.org/10.1109/ICIRCA51532.2021.9544592>
- [25] S. Zhang, H. Gu, and R. Li, "Music Genre Classification: Near-Realtime Vs Sequential Approach," *PrePrint*, 2019. <http://cs229.stanford.edu/proj2019spr/report/3.pdf>
- [26] A. Ghildiyal, K. Singh, and S. Sharma, "Music genre classification using machine learning," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*: IEEE, pp. 1368-1372, 2020. <https://doi.org/10.1109/ICECA49313.2020.9297444>
- [27] S. Das and A. K. Kolya, "A Theoretic Approach to Music Genre Recognition from Musical Features Using Single-Layer Feedforward Neural Network," in *Emerging Technologies in Data Mining and Information Security*: Springer, pp. 145-155, 2019. <https://doi.org/10.1007/978-981-13-1951-8>
- [28] P. Hamel and D. Eck, "Learning features from music audio with deep belief networks," in *ISMIR*, vol. 10: Citeseer, pp. 339-344, 2010. <https://ismir2010.ismir.net/proceedings/ismir2010-58.pdf>
- [29] A. Zlatintsi and P. Maragos, "Comparison of different representations based on nonlinear features for music genre classification," in *2014 22nd European Signal Processing Conference (EUSIPCO)*: IEEE, pp. 1547-1551, 2014. <https://ieeexplore.ieee.org/abstract/document/6952549>
- [30] U. Riaz, S. Aziz, M. Umar Khan, S. A. A. Zaidi, M. Ukasha, and A. Rashid, "A novel embedded system design for the detection and classification of cardiac disorders," *Computational Intelligence*, vol. 37, no. 4, pp. 1844-1864, 2021. <https://doi.org/10.1111/coin.12469>

## BIOGRAPHY OF AUTHORS

**Md Shofiqul Islam** is doing Masters (Research-based), a student at University Malaysia Pahang (UMP), Pahang, Malaysia. He has completed the B. Sc. at CSE from Islamic University, Kushtia, Bangladesh, in 2014. Now he is a research assistant at University Malaysia Pahang (UMP). He is also a teacher at CSE under the faculty of FST at ADUST university, Dhaka. He is also in the teaching profession since 2015. His research field is Deep learning, Machine learning, Natural Language Processing, Image Processing. He has published a lot of papers in my field. Email: [shafiqseiu07@gmail.com](mailto:shafiqseiu07@gmail.com)

**Md Munirul Hasan** is doing Ph.D. at UMP. He is working in the field of image processing, video processing, and text processing. Email: [monirul.iuc@gmail.com](mailto:monirul.iuc@gmail.com)

**Md Abdur Rahim** is doing a Ph.D. in Mechanical Engineering at UMP. He is working in the field of Mechanical engineering, vehicle monitoring, and text processing. Email: [rahimul7421@gmail.com](mailto:rahimul7421@gmail.com)

**Ali Muttaleb Hasan** is doing a Ph.D. in CSE at UMP. He is working in the field of image processing, video processing, and text processing. Email: [alimatlab65@yahoo.com](mailto:alimatlab65@yahoo.com)

**Mohammad Mynuddin** is doing a Ph.D. at the Department of Civil, Transportation Engineering, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816 USA. He is working in the field of image processing, video processing, and text processing. Email: [mm40771@georgiasouthern.edu](mailto:mm40771@georgiasouthern.edu)

**Imran Khandokar** is doing a B. Sc. at UMP. He is working in the field of image processing, video processing, and text processing. Email: [imrankhandokarump@gmail.com](mailto:imrankhandokarump@gmail.com)

**Md Jabbarul Islam** has been completing bachelor's degrees from the Department of Mathematics, National University Gazipur-1704, Dhaka, Bangladesh. He is doing his research work in the field of Graph theory, Statistics, Machine learning, image processing, video processing, and text processing. Email: [abdujabbbar11061997@gmail.com](mailto:abdujabbbar11061997@gmail.com)