iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37669

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Music Genre Classification for Indian Music Genres

Neha Kumari[1], Tushar Shukla[2], Swati[3], Kumar Satyam[4], Balachandra K.[5]
[1, 2, 3, 4, 5]*Department of Telecommunication, BMS College of Engineering, Bengaluru, India*

Abstract: ***Due to the enormous expansion in the accessibility of music data, music genre classification has taken on new significance in recent years. In order to have better access to them, we need to correctly index them. Automatic music genre classification is essential when working with a large collection of music. For the majority of contemporary music genre classification methodologies, researchers have favoured machine learning techniques. In this study, we employed two datasets with different genres. A Deep Learning approach is utilised to train and classify the system. A convolution neural network is used for training and classification. In speech analysis, the most crucial task is to perform speech analysis is feature extraction. The Mel Frequency Cepstral Coefficient (MFCC) is utilised as the main audio feature extraction technique. By extracting the feature vector, the suggested method classifies music into several genres. Our findings suggest that our system has an 80% accuracy level, which will substantially improve on further training and facilitate music genre classification.***

Keywords: Music Genre Classification, CNN, KNN, Music information retrieval, feature extraction, spectrogram, GTZAN dataset, Indian music genre dataset.

I. INTRODUCTION

Due to the increase in data consumption, music which has been the source of entertainment for centuries uncounted, has also take the online route to reach its patrons. The sheer volume of data being uploaded on these cloud platforms however, are not automatically assigned a genre. Automatic music genre classification is the essence for organizations that sell music or let users stream their content for a fee. The music database as a result would need to provide accessibility to its users to search, retrieve, and make automatic recommendations to the listeners based on their taste. Due to the very difficult task of extraction of acceptable audio features, music classification is regarded as a difficult undertaking. While unlabeled data is widely available, there is a scarcity of music with properly classified genre. The two phases in music genre categorization are the extraction of audio features i.e. speech processing and the classification of the music data in the servers to increase the accessibility for the users. The first stage involves in the extraction of multiple audio features. The characteristics extracted from the training data are used to build a classifier in the second stage. There have been a variety of techniques to categorizing music into different genres. With so much music data available on the web, the organizations require an automatic music genre a classification system. Various types of feature extraction are used in each implementation. Some studies use pitch, timber, and beat as classification criteria. Deep learning algorithms are a popular classification technique that is used to train a large database.

Using Convolution Neural Networks and K-Nearest Neighbors technique, we offer a unique approach for automatic music genre classification. Music genre classification forms a basic step for building a strong recommendation system. The idea behind this project is to see how to handle sound files in python, compute sound and audio features from them, run Machine Learning Algorithms on them, and see the results. In a more systematic way, the main aim is to build and train a machine learning model, which serves the need of classifying music samples into their respective genres. It aims to predict the genre using an audio file as its input. The objective of automating the music classification is to make the selection of songs quick and less cumbersome. If one has to manually classify the songs or music, one has to listen to a whole lot of songs and then select the genre. This is not only time-consuming but also difficult. Automating music genre classification can be optimized to help to find and tag audio files with their genres, and artists easily.

II. LITERATURE SURVEY

[1]. In Hareesh Bahuleyan's work, the study presents a technique to classify music automatically by offering distinct tags to the songs available in the user's collection using Machine Learning methodologies. It discovers both Neural Network and classic methods of employing Machine Learning algorithms to accomplish their goals. The first concept makes use of Convolutional Neural Networks. In this the properties of Mel Spectrograms (pictures) of the audio stream are used to train the system from start to finish. The second method employs several Machine Learning methods such as Logistic Regression and Random Forest, among others, to extract hand-crafted features from the audio recordings' temporal and frequency domains.

ML methods such as Logistic Regression, Support Vector Machines (SVM), Random Forest, and Gradient Boosting are used to classify the music into its genres utilising the features which have been extracted manually; MFCC and Chroma features. When they compared the two techniques separately, they discovered that the VGG-16 CNN model had the best accuracy. By putting together ensemble classifier of VGG-16 CNN and XGB the model with 0.894 accuracy was built.

[2] In the second study that we reviewed Tom LH Li et al., They attempted to understand the fundamental features that truly contribute to the construction of the ideal model for Music Genre Classification utilising automatic musical feature extraction using convolutional neural networks. The main motive of this research is to build a novel method for extracting musical elements from audio files using Convolution Neural Networks. Their main goal is to investigate the many applications of CNN in music information retrieval (MIR). Their findings and studies reveal that CNN is capable of extracting low-level informative elements from a variety of musical patterns. The statistical spectral characteristics, rhythm, and pitch retrieved from audio recordings are less dependable, resulting in less accurate models. As a result of their approach to CNN, musical data has similar qualities to picture data and hence requires very little prior knowledge. GTZAN was the dataset used. It contained ten genres, each with 100 audio recordings. Each audio file is 30 seconds long, with a sample rate of 22050 Hz and a 16-bit resolution. The musical patterns were assessed using the WEKA tool, which included a variety of classification methods. The accuracy of the classifier was 84 percent at first, and it gradually improved. In comparison to the MFCC, chroma, and temp characteristics, the CNN features yielded better results and were more trustworthy. When calculating on other combinations of genres at the same time, the accuracy can still be improved.

[3] Tzanetakis and Cook were the first to use a machine learning algorithm to classify music genres. They generated the GTZAN dataset, which is still used as the gold standard for genre classification today. ChangshengXu et al. [4] demonstrated how support vector machines (SVM) can be used for this. For music genre classification, the authors used supervised learning approaches.

In [5], Scheirer devised a real-time beat tracking system for audio waves accompanied by music. A filter bank is used with a network of combination filters in this model to offer an idea of the main beat and its potency by tracing the signal periodicities. The author describes a real-time beat tracking system based on a multitudinous agent architecture that tracks many beat hypotheses in [6]By generating more data from the baseline dataset, GTZAN, Tao in [7] illustrates how to employ limited Boltzmann machines to achieve better outcomes than a regular multilayer neural network. In this study, a data distribution problem in the dataset is proposed and described, demonstrating that using only the GTZAN dataset, it is difficult to reliably identify more than four classes. This paper also suggests using MFCC spectrograms for song pre-processing.

[8] The work of Gwardys et al. demonstrates an intriguing strategy incorporating transfer learning. They used ILSVRC-2012 [9] to train the model for image recognition, and then reused it for genre classification on MFCC spectrograms. The model employed in this article has five layers, the first two of which have max pooling while the last one does not. Finally, there are three layers that are entirely joined.

To achieve better results, the authors in [10] employed midi, pitch, and duration as low-level elements of the music to accomplish the classification. We present a system for automatically classifying music into different genres based on the following literatures. The sections that follow will explain this.

III. DATASETS

A. The GTZAN genre collection dataset is very well known is and used for genre classification and by G. Tzanetakis and P. Cook. The authors wrote a paper on music genre classification and created a dataset specifically for this, they collected the audio files in the year 2000-2001 during their study from a various sources including music CDs and radio shows. The dataset encompasses 1000 music files each having with the duration of 30 seconds each. Thy authors created 10 different genres each containing hundred audio tracks. The audio files have been stored in .wav format.

B. The second dataset used in our project is Indian Music Genre dataset that encompasses different Indian music genres in the form of mp3 files. The dataset was collected for just the purpose of using it to train Machine learning models into classifying Indian music genres. Since, Indian music genres are vastly different in forms and techniques from the western music we felt, it was necessary to use a dataset that incorporates the best data available. The dataset contains 5 genres, each represented by 100 tracks. The tracks are 45 seconds long which were truncated with appropriate window size during the training of the model.

TABLE 1 DATASET

Genres		No. of Records
INDIAN MUSIC GENRE DATASET	CARNATIC	100
	GHAZAL	100
	SEMI-CLASSICAL	100
	SUFI	100
	BOLLY-POP	100
GTZAN DATASET	HIP-HOP	100
	CLASSICAL	100
	COUNTRY	100
	DISCO	100
	BLUES	100
	JAZZ	100
	METAL	100
	POP	100
	REGGAE	100
	ROCK	
TOTAL AUDIO RECORDS		1500

IV. WORKFLOW

A. Model

In FIG. 1 we see a workflow of this project. This is the minimum path taken by the model of our project.

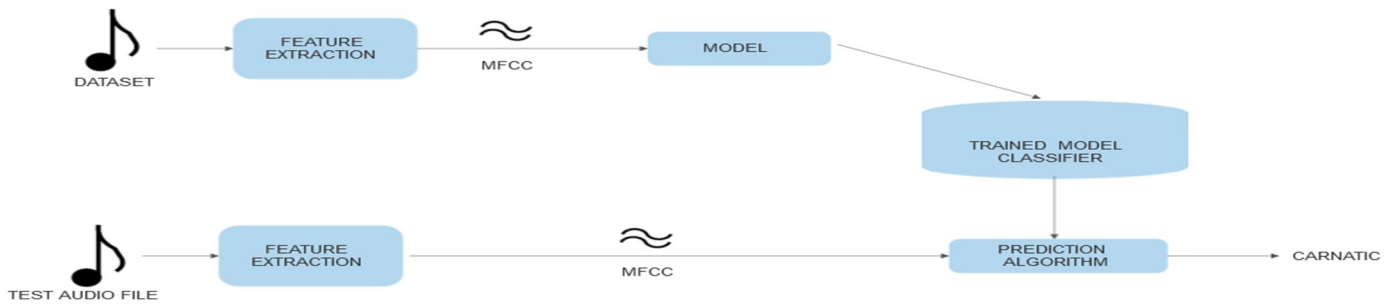


FIG. 1 Workflow of The Model

Firstly, the known 10 genres of western music and 5 genres of Indian music, which we are using in our project goes under feature extraction. We are using MFCC to obtain the spectrograms of the audio which is then succeeded by model training employing CNN and KNN algorithms. The prediction of the music genre takes place after the model is trained and the subsequent data is generated. The new input audio file is then fed to the trained classifier which then proceeds to make a prediction based on the spectral features extracted.

B. Feature Extraction

The step one for music genre classification would be to extract features and additives from the audios. It consists of figuring out the linguistic content and discarding noise.

1) Mel Spectrogram: A Mel Spectrogram differs from a conventional Spectrogram in that it plots Frequency vs. Time in two ways. On the y-axis, it employs the Mel Scale instead of Frequency. It indicates colours using the Decibel Scale rather than the Amplitude Scale. We commonly use this instead of a standard Spectrogram for deep learning models. By using mel scale instead of frequency and decibel scale instead of amplitude (loudness), we can see a much clear spectrogram as we can see in Fig 2.

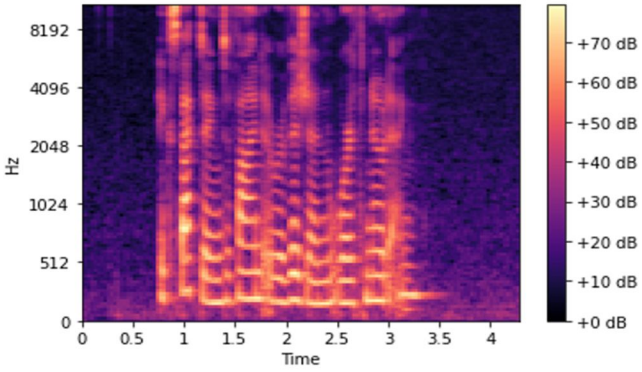


FIG. 2 MEL Spectrogram

2) Mel Frequency Cepstral Coefficients (MFCC)

These are present day capabilities used in automatic speech and speech popularity studies. There are a hard and fast of steps for generation of these features:

a) The audio signals keep on changing constantly; the first step is to divide these into smaller sets. Each set is around 20-40 ms long and we try to perceive unique frequencies found in every frame.

b) Now we separate linguistic frequencies from the noise

c) And the final step involves discarding the noise; the discrete cosine transform (DCT) of those frequencies is calculated in the next step. By using DCT we preserve particular sequence of frequencies which have a high probability of information.

MFCC is the coefficient of an MFC and the extraction procedure starts with the aid of windowing the signal, making use of the Discrete Fourier Transform (DFT), taking the log of the value, after which mapping the frequencies on a mel scale and then using the inverse discrete cosine transformation (DCT). The first few coefficients hold most information for 39 coefficients in our model.

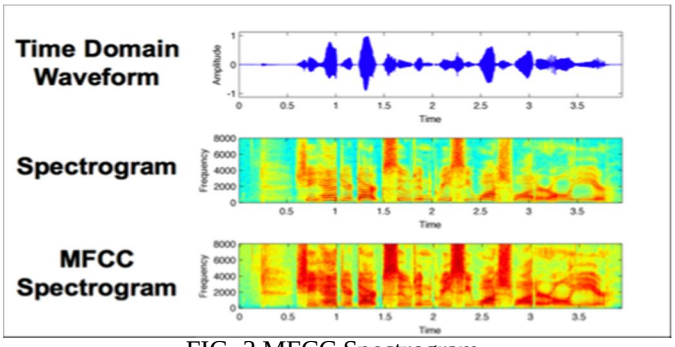


FIG. 3 MFCC Spectrogram

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

The formula to convert the frequency to mel scale is:

$$m=2915\cdot(1+f500)$$

Also, the formula for the Fourier Transform is given as:

$$X(f) = \int x(t) * e-i2\pi ftdt\infty-\infty$$

By expressing magnitudes on the logarithmic-axis, the log-spectrum already integrates perceptual sensitivity on the magnitude axis. The frequency axis is the other dimension.. We shall focus on the mel-scale in this context, most likely as a result of an obvious choice based on trend. The perceived margin between pitches of different frequencies is described by this scale.

V. MODEL TRAINING AND IMPLEMENTATION

A. CNN

Neural networks are one method for classifying data. Because NNs (like the CNN we'll be utilising) require some form of picture representation, the audio samples were converted to Mel Spectrograms to achieve this. To do so, we extracted the low level characteristics in time and frequency for audio files (.wav) that were contained in the dataset and mapped these low level features into a .json file and then this was used to train the model. Feature extraction was done using MFCC function which is mel-frequency cepstral coefficient as discussed earlier. Any incoming audio file was then identified on the basis of its low level features and comparing these with the low level features of audio files trained in the model and then the prediction of genre to which the new audio file belongs to was made.

The below image shows building of model using Keras which is an open source platform which provides python interface to neural networks. The image shows the code to build a CNN model using different layers of Keras. For the CNN model, we had used the Adam optimizer for training the model. The epoch that was chosen for the training model is 100. The RELU activation function is used in all of the hidden layers, whereas the output layer involves using the softmax function as an output layer. The sparse categorical crossentropy function is used to determine the loss.We chose the Adam optimizer because it gave us the best results after evaluating other optimizers.

The model accuracy can be increased by further increasing the epochs but after a certain period, we may achieve a threshold, so the value should be determined accordingly.

B. KNN

K-Nearest Neighbors is one of the most fundamental machine learning approaches. Despite its simplicity in idea, KNN is a complicated algorithm that produces reasonably good accuracy on most jobs. KNN tries out various K values to see which one produces the best results. KNN is a supervised learning algorithm, which means that the samples in the dataset must have labels/classes assigned to them. KNN is a non-parametric method, for starters. This means that when the model is utilised, no assumptions about the dataset are made. Rather, the model is built solely from the data provided. Second, when employing KNN, the dataset is not separated into training and test sets. When the model is asked to generate predictions, KNN makes no generalizations between the training and testing sets, therefore all of the training data is used.

VI. RESULTS AND ANALYSIS

A. Accuracy

The accuracy of the model is calculated using the formula

$$\textbf{ACCURACY \%} = \frac{\textbf{No. of songs correctly classified}}{\textbf{Total no of songs}} \textbf{ X 100}$$

Table 2 Prediction Accuracy

DATASET	CNN (50 epochs)	KNN (K=5)
GT-ZAN (using MFCC)	0.736	0.686
INDIAN MUSIC GENRE (using MFCC)	0.765	0.798

B. Error Plot

FIG. 4 shows accuracy and error plots

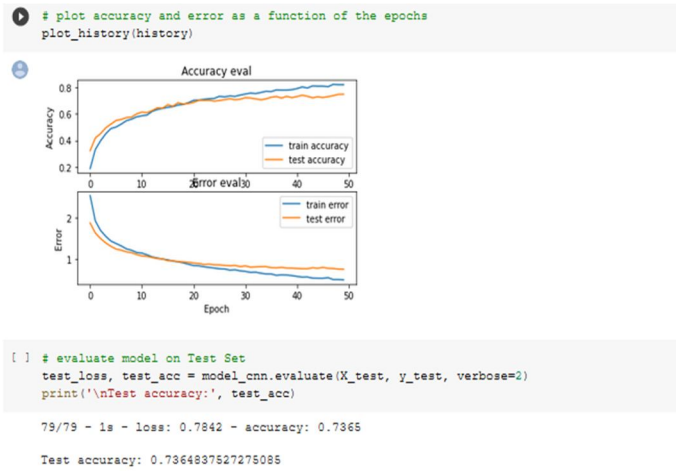


FIG. 4 Error Plot

C. Genre Prediction

After successfully validating our model on the test set, we created a cell for new song prediction that contained the code for pre-processing the new external audio file. The low-level elements of this input file were analysed and compared in order to estimate the genre of this audio file. The hop length was set to 512, the sampling rate was set to 22050, and the number of mfcc coefficients was set to 13.

The Librosa library, a python package for music and audio analysis, was used. It outlines the basic processes for developing music information retrieval systems. We can extract certain critical properties from audio samples using Librosa, such as tempo, Chroma Energy Normalized, Mel-Freqency Cepstral Coefficients, Spectral Centroid, Spectral Contrast and Spectral Roll off.

The classifier that we created to predict the genre of audio files using the K-nearest neighbours technique works by creating a dataset file that contains all of the raw audio data categorised. In this experiment, we used the GTZAN music genre categorization dataset and the Indian music genre dataset. We used a count of K as 5 in our deep learning research to create a K nearest neighbor.

VII. CONCLUSION AND FUTURE WORK

This project presents an application that uses Machine Learning techniques to do Music Genre Classification.

We studied quite a few audio feature extraction techniques and concluded that the best one to use for our project would be MFCC for feature extraction.

We implemented KNN and CNN algorithms to train our model and perform classification on our dataset. The python based librosa and python speech features package have been employed to extract the audio features and plot the spectrograms for further training

The peak accuracies that we have obtained after training our models for both K-NN and CNN are:

KNN is about 80% and CNN is 73%.

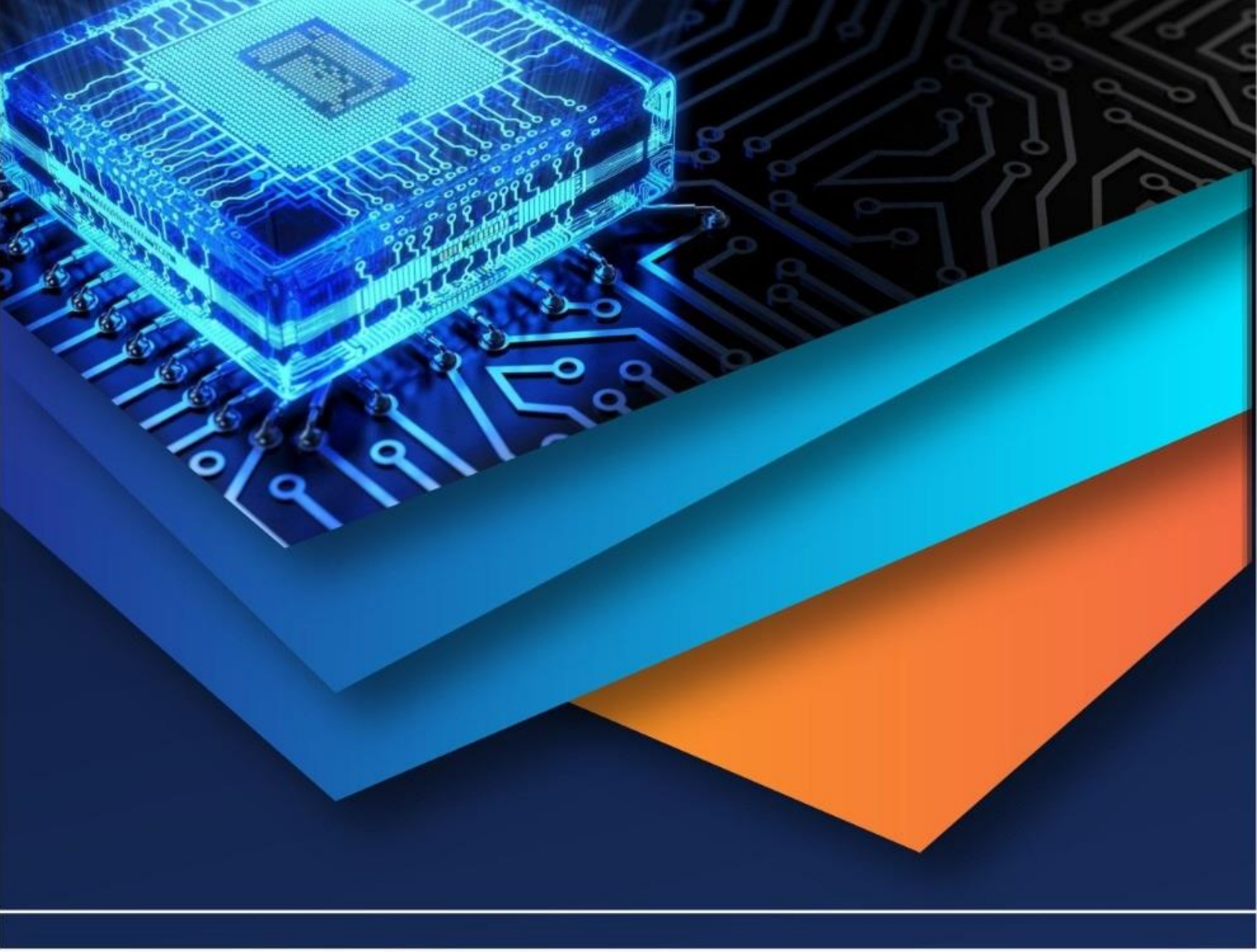
The accuracies can be affected by the quality and the vastness of the dataset. To achieve a prediction with decent accuracy the dataset needs to encompass a great variety of genres with large number of tracks for the model to train on. One problem we faced during the project was the lack of proper dataset for our model.

The future work in this domain is quite immense since the field of application of voice intelligence is wide and fairly new. We have tried to implement both the western and Indian genres of music. The constraint was of dataset and the window size for the Discrete Fourier transform. Upon creating a dataset large enough with a good number of audio tracks, the accuracy of the model is expected to increase further.

The performance of different features extracted from audio samples was investigated in order to identify useful feature descriptors for the Indian music genres of classical (Carnatic music) and folk music. This project could also be developed into a standalone we application deployable on websites to automatically classify the fed audio or video inputs.

REFERENCES

[1] Hareesh, B. "Music Genre Classification using machine learning techniques." 2018. www.researchgate.net/publication/324218667_Music_genre_Classification_using_Machine_Learning_Techniques
[2] Tom,Li & Antoni, C & Andy, C. "Automatic Musical Pattern Feature Extraction Using Convolutional Neural Network." (2010) Lecture Notes in Engineering and Computer Science 2180.
[3] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," in *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, July 2002, doi: 10.1109/TSA.2002.800560
[4] Changsheng, Xu & Namunu, M. & Xi, S. & Fang, C. & Qi, T. "Musical genre classification using support vector machines." (2003) 5. V - 429. 10.1109/ICASSP.2003.1199998.
[5] Eric, D. S. "Tempo and beat analysis of acoustic musical signals." The Journal of the Acoustical Society of America 103.1 (1998): 588-601
[6] S, Vishnupriya, and K. Meenakshi. "Automatic music genre classification using convolution neural network." 2018 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2018
[7] T. Feng. Deep learning for music genre classification. 2014.
[8] D. G. Grzegorz Gwardys. "Deep image features in music information retrieval." 2014 10.O
[9] Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei.ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3):211–252, 2015.
[10] Rebelo, Ana, et al. "Optical music recognition: state-of-the-art and open issues." International Journal of Multimedia Information Retrieval 1.3 (2012): 173-190.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY