iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.54344

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: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

Analyzing the Performance of Machine Learning Models in Music Genre Classification

Lakshay Sharma[1], Prof. (Dr.) Amita Goel[2], Er.Nidhi Sengar[3], Dr.Vasudha Bahl[4]
Department of Information Technology, Maharaja AgrasenInstitute of Technology affiliated to Guru Gobind Singh Indraprastha University, Rohini, Delhi

Abstract: ***Music genre classification is a fundamental task in the field of music information retrieval (MIR) and has gained significant attention in recent years due to the rapid growth of digital music collections. This research paper presents a comprehensive review of the application of machine learning techniques for music genre classification. We explore various methodologies, feature extraction techniques, and classification algorithms used in the domain, highlighting their strengths, limitations, and recent advancements. The objective of this paper is to provide researchers and practitioners with a comprehensive understanding of the current state-of- the-art approaches, challenges, and future directions in music genre classification using machine learning.***

I. INTRODUCTION

A. Background and Motivation

With the rise of digital music and streaming platforms, there is a growing demand for automated music genre classification. Machine learning offers a solution by handling large datasets and extracting meaning features. It captures complex relationships among audio features, enabling accurate genre predictions. This approach enhancesuser experiences in music recommendation and playlist generation. Additionally, it promotes interdisciplinary collaboration in computer science, signal processing, musicology, and cognitive science, deepening our understanding of the cognitive processes underlying music genre classification.

B. Objectives of the Paper

The research paper aims to provide a comprehensive review of machine learning techniques for music genre classification. It explores feature extraction techniques, machine learning algorithms, datasets, and evaluation metrics used in the field. The paper highlights the strengths and limitations of different approaches, facilitating informed decision-making in system design. It presents experimental setups, results, and comparative analysis to assess performance. Additionally, recent advancements such as transfer learning and multimodal approaches are discussed, along with future directions and challenges in music genre classification. The paper serves as a guide forresearchers and practitioners, fostering innovation and development in the field.

C. Feature Extraction Techniques

Feature extraction techniques play a crucial role in music genre classification using machine learning. They involvetransforming the raw audio signal into a set of representative features that capture important characteristics of the music.
Here are some commonly used feature extraction techniques:

1) *Timbral Features:* Timbre refers to the tone color or quality of sound. Timbral features capture properties like brightness, roughness, and spectral centroid, providing insights into the texture and tonal characteristics of the music.

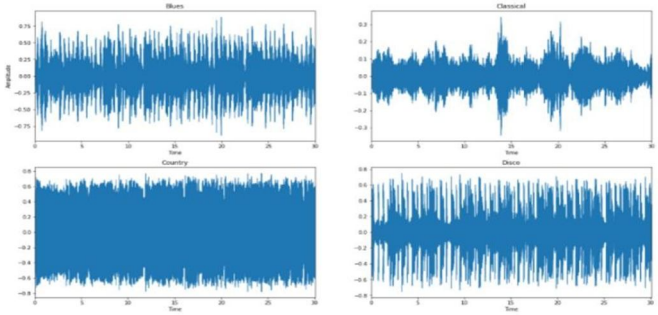


Figure 1: Plot raw wav file

2) *Rhythm and Tempo Features:* Rhythm features capture the rhythmic patterns in the music, including beat and tempo information. Tempo features,such as beat histogram and tempo histogram, quantify the tempo variations and rhythmic structure.
3) *Harmonic and Melodic Features:* Harmonic features describe the harmonic content of the music, including chord progressions, key profiles, and tonalstability. Melodic features capture the melodic characteristics, such as pitch contour, note duration, and pitch statistics.

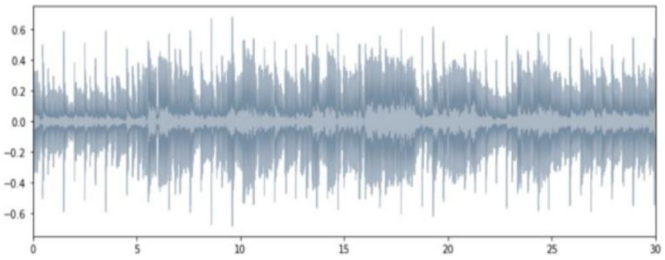


Figure 2: Spectral roll off of audio wav file

Spectral features provide information about the frequency content of the music. They include spectral centroid,spectral flux, and spectral roll-off, which convey details about the distribution of energy across the frequency spectrum.

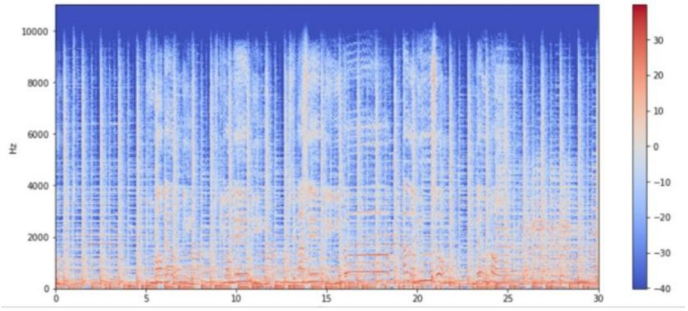


Figure 3: Spectrogram of audio wav file

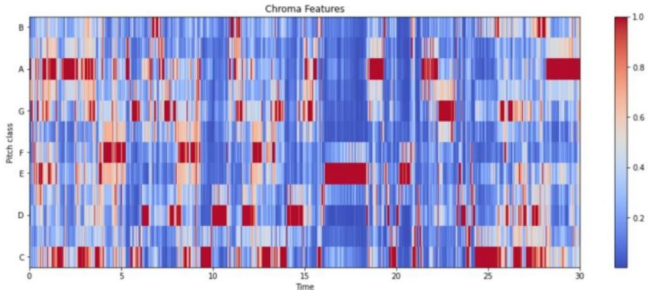


Figure: Chroma feature of audio wav file

4) Statistical Features: Statistical features encompass various statistical measures computed from the audio signal, such as mean, variance,skewness, and kurtosis. They capture statistical properties of the music and can provide insights into its complexityand variation.
5) Deep Learning-based Features: Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), can learn hierarchical representations directly from the raw audio signal. These models can automatically extract high-level features that capture both local and global patterns in the music.

The selection of feature extraction techniques depends on the specific requirements of the music genre classification task and the characteristics of the dataset. Researchers often combine multiple feature types to capture a comprehensive representation of the music. The extracted features serve as input to machine learning algorithms forgenre classification.

D. Machine Learning Techniques

Several machine learning algorithms can be employed for music genre classification. Here are some commonly usedalgorithms in this context:

1) Support Vector Machines (SVM)

SVM is a supervised learning algorithm that aims to find an optimal hyperplane to separate different music genres in the feature space. It works well with high-dimensional data and can handle both linear and non-linear classification problems. The goal of SVM is to divide thedata- sets into classes to find a maximum marginal hyperplane (MMH)An SVM model is basically a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the data-sets into classes to find a maximum marginal hyperplane (MMH).The main goal of SVMis to divide the data-sets into classes to find a maximum marginal hyperplane (MMH) and it can be donein the following two steps -> First, SVM will generate hyperplanes iteratively that segregates the classes in best way. Then, it will choose the hyperplane that separates the classes correctly.

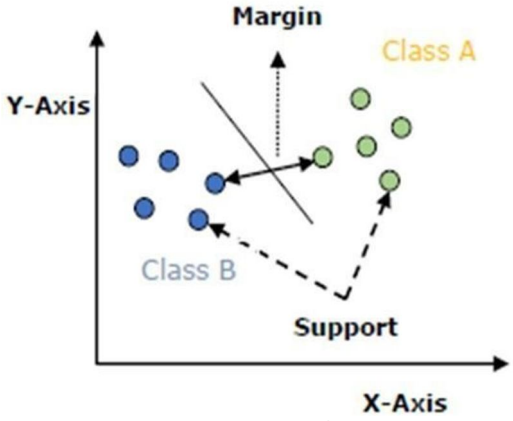


Figure 4: SVM

2) k-Nearest Neighbors (k-NN)

k-NN is a simple yet effective algorithm that classifies a new music sample based on the majority vote of its k nearest neighbors in the feature space. It is a non-parametric algorithm that does not make any assumptions about theunderlying data distribution.

3) Decision Trees

Decision trees partition the feature space based on a series of binary decisions, resulting in a hierarchical structure. Each internal node represents a decision based on a specific feature, leading to the classification of the music sampleat the leaf nodes.

4) Random Forests

Random forests combine multiple decision trees to form an ensemble model. Each tree is trained on a random subsetof features and samples, and the final classification is determined by the majority vote of the individual trees. Random forests are known for their robustness and ability to handle high-dimensional data.

5) Neural Networks

Neural networks, especially deep learning architectures, have gained significant popularity in music genre classification. Convolutional neural networks (CNNs) can automatically learn hierarchical representations fromspectrogram or wavelet transforms, capturing local and global patterns. Recurrent neural networks (RNNs) are effective in modeling temporal dependencies in music.

6) Deep Learning Models

Deep learning models, such as deep neural networks and recurrent neural networks, can be used for end-to-endlearning, directly mapping the audio signal to genre labels. These models have the capability to learn complex representations and have shown promising results in music genre classification tasks.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

The choice of machine learning algorithm depends on factors such as the dataset size, dimensionality of the features, desired classification accuracy, and computational resources available. It is often beneficial to compare and combinemultiple algorithms to identify the best approach for a specific music genre classification task.

II. DATASETS

Several datasets have been created and used for music genre classification research. These datasets provide labeled examples of music tracks across various genres, allowing researchers to train and evaluate machine learning models. Here are some commonly used datasets: GTZAN Genre Collection: The GTZAN dataset is one of the most widely used datasets for music genre classification. It contains 1,000 audio excerpts of 30 seconds each, evenly distributed across ten genres, including rock, pop, jazz, classical, hip-hop, and others. The dataset is manually annotated by experts and its size is around 1 GB. These datasets provide a foundation for training and evaluating machine learning models for music genre classification. Researchers often use a combination of these datasets or create their own datasets tailored to theirspecific research goals and genres of interest. We will be using GTZAN for our project.

A. Experimental Setup And Evaluation

In our research, we utilized the Librosa Python library for audio analysis and data processing. Librosa provided uswith the necessary tools and techniques to extract relevant data from audio samples, which was crucial for solving our music genre classification problem. To extract features from the audio samples, we employed various functions available in the Librosa library. Some ofthe features we focused on include the zero-crossing rate, spectral centroid, spectral rolloff, Mel Frequency Cepstral Coefficients (MFCCs), and chroma feature. The zero-crossing rate measures the rate of sign changes in the signal, while the spectral centroid calculates the weighted mean of the frequencies present in the sound. The spectral rolloff indicates the frequency below which a specified percentage of spectral energy lies. MFCCs are a set of features that describe the shape of the spectral envelope and are commonly used in audio analysis. We calculated 20 MFCCs over 97 frames and performed feature scaling. The chroma feature represents the12 semitones of the musical octave.

1) *Mel-Frequency Cepstral Coefficients (MFCC):* MFCCs were introduced in the early 1990s by Davis and Mermelstein and have since been widely used in tasks like speech recognition. The process involves taking the Short-Time Fourier Transform (STFT) of the signal using parameters such as n_fft=2048, hop size=512, and a Hann window. The power spectrum is then computed, followed by applying a triangular MEL filter bank to mimic human sound perception. The MFCCs are obtained by taking the discrete cosine transform of the logarithm of the filterbank energies. In this study, the number of filter banks (n_mels) was set to 20.
2) *Spectral Centroid:* The spectral centroid represents the frequency around which most of the energy is concentrated in each frame. It is calculated as the magnitude-weighted frequency using the formula:

   fc = (∑k S(k)f(k)) / (∑k f(k)),

   where S(k) is the spectral magnitude of frequency bin k, and f(k) is the frequency corresponding to bin k. Spectral Bandwidth: The spectral bandwidth corresponds to the moment about the spectral centroid and providesinformation about the spread of frequencies. It is calculated using the formula:

   [∑k (S(k)f(k) - fc)^p]^(1/p),

where p represents the order of the moment, and p = 2 is equivalent to a weighted standard deviation.

3) *Spectral Contrast:* Spectral contrast is calculated within pre-specified frequency bands, measuring the differencebetween the maximum and minimum magnitudes in each band.
4) *Spectral Roll-off:* The spectral roll-off represents the frequency below which a certain percentage (e.g., 85%) of thetotal energy in the spectrum lies.

Using Librosa, we extracted these features from the dataset based on both the time and frequency domains. Weselected the most informative features and stored them in a CSV file for further analysis and classification.

III. RESULTS

A. Comparitive Analysis Of Results

In this classification system, we used K-Nearest Neighbour (K-NN) and Support Vector Machine (SVM) which is developed in Convolutional Kernal with the help of Convolutional Neural Network (CNN) that provide more accuracy compared to the previous system. The testing data-set gives an accuracy of more than 95%.

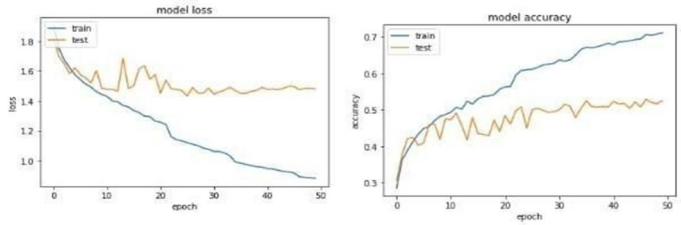


Figure 5: Comparitive analysis of SVM and KNNK-Nearest Neighbour(K-NN)

60.4% 66.4%(with hyperparameters tunning)
Support Vector Machine (SVM)
73.2% 76.4%(with hyperparameters tunning)
Convolutional Neural Networks (CNNs) are widely used in image classification tasks and can also be applied to music genre classification by treating spectrograms as 'images'. The CNN architecture used in this study is VGG-16, which achieved top performance in the ImageNet Challenge 2014.
Investigation was conducted to determine the most influential features in the music genre classification task. The XGB model was chosen based on the previous section's results. To rank the top 20 features, a scoring metric was employed, which measured how frequently a feature was used as a decision node in the gradient boosting predictor's individual decision trees. Figure illustrates the results, highlighting that Mel-Frequency Cepstral Coefficients (MFCC) were the most prominent among the important features. Previous studies have also recognized the significance of MFCCs in enhancing speech recognition systems. In the context of music genre classification, our experiments demonstrate that MFCCs contribute significantly. Additionally, the mean and standard deviation of spectral contrasts at different frequency bands, as well as the music tempo measured in beats per minute, were identified as important features within the top 20. Furthermore, an analysis was conducted to assess the model's performance when trained solely on the top N features. Table shows that even with just the top 10 features, the model achieved surprisingly good performance. Compared to the full model containing 97 features, using only the top 30 features resulted in only a slight decrease in performance (2 points on the AUC metric and 4 points on the accuracy metric).

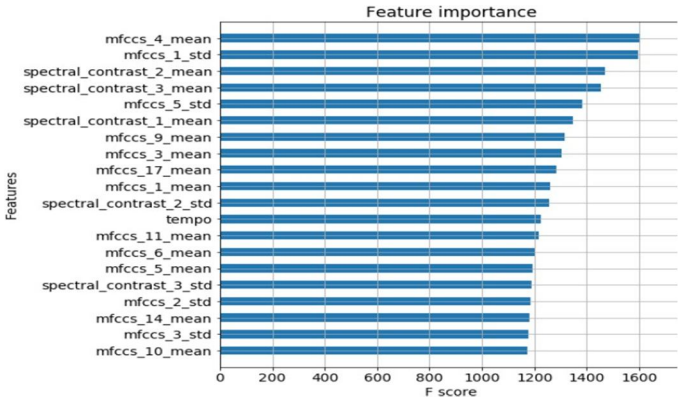


Figure 6: Features of MFCC analysis (XGBoost)A CNN block consists of the following operations:

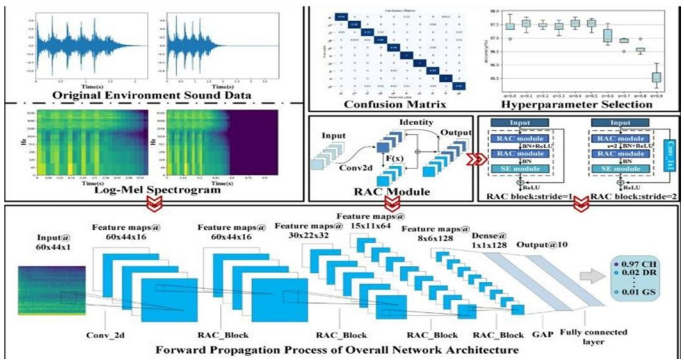


Figure 7: CNN Design for audio classification

Convolution: A matrix filter is slid over the input image (spectrogram) to compute a feature value through element- wise multiplication and summation. Multiple filters are used, and their values are learned during the network training.Pooling: This step reduces the dimensionality of the feature map obtained from the convolution. Max pooling is commonly used, where the maximum value among a window of elements is retained while sliding the window acrossthe feature map. Non-linear Activation: To introduce non-linearity and enhance the network's power, an activation function like ReLU is applied to each element of the feature map. In the context of music genre classification using spectrograms, the VGG-16 model is downloaded with pre-trained weights. The convolutional base (5 blocks) is extracted, followed by a new feed-forward neural network that predicts the music genre. This network has a final layer that outputs class probabilities using the softmax activation function. There are two settings for implementing the pre-trained model: Transfer learning: The weights in the convolutional base are fixed, and only the weights in the feed-forward network are tuned to predict the correct genre label. Fine-tuning: The pre-trained weights of VGG-16 are used, and all model weights are allowed to be tuned during training. The cross-entropy loss is computed to measure the difference between predicted and actual class probabilities. The loss is used to update the networkweights through backpropagation until convergence.

```
test_loss, test_acc  = model.evaluate(X_test, y_test, batch_size=128)
print("The test Loss is :",test_loss)
print("\nThe Best test Accuracy is :",test_acc*100)

26/26 [==============================] - 0s 2ms/step - loss: 0.5182 - accuracy: 0.9293
The test Loss is : 0.5182217955589294

The Best test Accuracy is : 92.93296933174133
```

Figure 8: CNN study

	With data processing			Without data processing		
	Train	CV	Test	Train	CV	Test
Support Vector Machine	.97	.60	.60	.75	.32	.28
K-Nearest Neighbors	1.00	.52	.54	1.00	.21	.21
Feed-forward Neural Network	.96	.55	.54	.64	.26	.25
Convolution Neural Network	.95	.84	**.82**	.85	.59	.53

Figure 9: Comparitive analysis

While Convolutional Neural Networks (CNN) can also be used for genre classification, this project employs an RNNwith LSTM due to the dataset size limitations and the accuracy drop observed with CNN as the number of genres increases. The dataset consists of strong and mild classes, with high and low amplitude audio files representing different genres, respectively. The Gtzan music dataset is used for training, and the Librosa library is employed to extract MFCC features from the raw audio data. These features, in the form of MFCC vectors, serve as input for the LSTM neural network model developed with Keras and TensorFlow. The blog explains the pipeline for extracting MFCC features using the Librosa library and provides code examples. The extracted MFCCs form a 2-D array, where one dimension represents time and the other represents different frequencies.

B. Long Short Term Memory Network (LSTM)

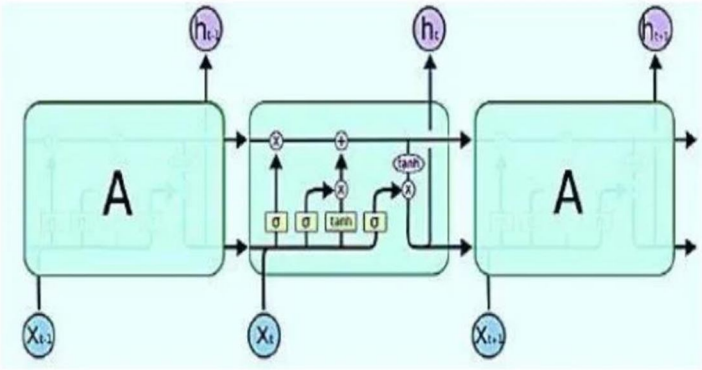


Figure 10: LSTM Cell

Instead of using Convolutional Neural Networks (CNN), the project explores the use of RNN with LSTM for Music Genre Classification. LSTM addresses the challenge of long-term dependencies and effectively utilizes past data to predict current outputs. The LSTM network used in the project consists of four layers, and its internal cell structure isdescribed.

C. Dataset and Pre-processing

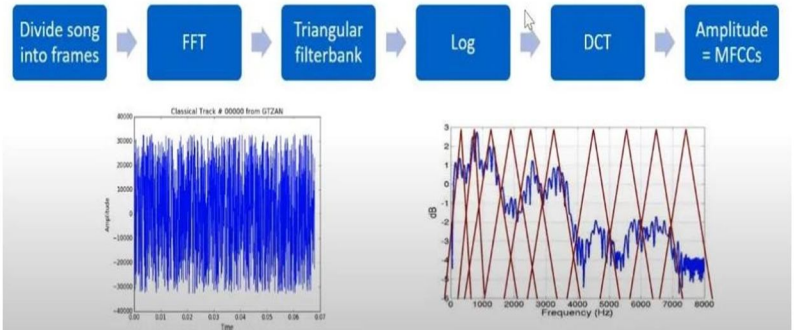


Figure 11: Pipeline for MFCC

The GTZAN dataset from marsyas.info is utilized, containing 1000 samples for each of the 10 music genres. The dataset is randomly split into training and testing sets, with no overlap. The training dataset is further divided into training and validation subsets. Waveform analysis reveals similarities between certain genres. Preprocessing involves extracting useful features from audio signals, where the MFCC is chosen due to its ability to define soundbrightness and timbre. Librosa is used to convert audio files into MFCC features the model is trained for around 28epoc.

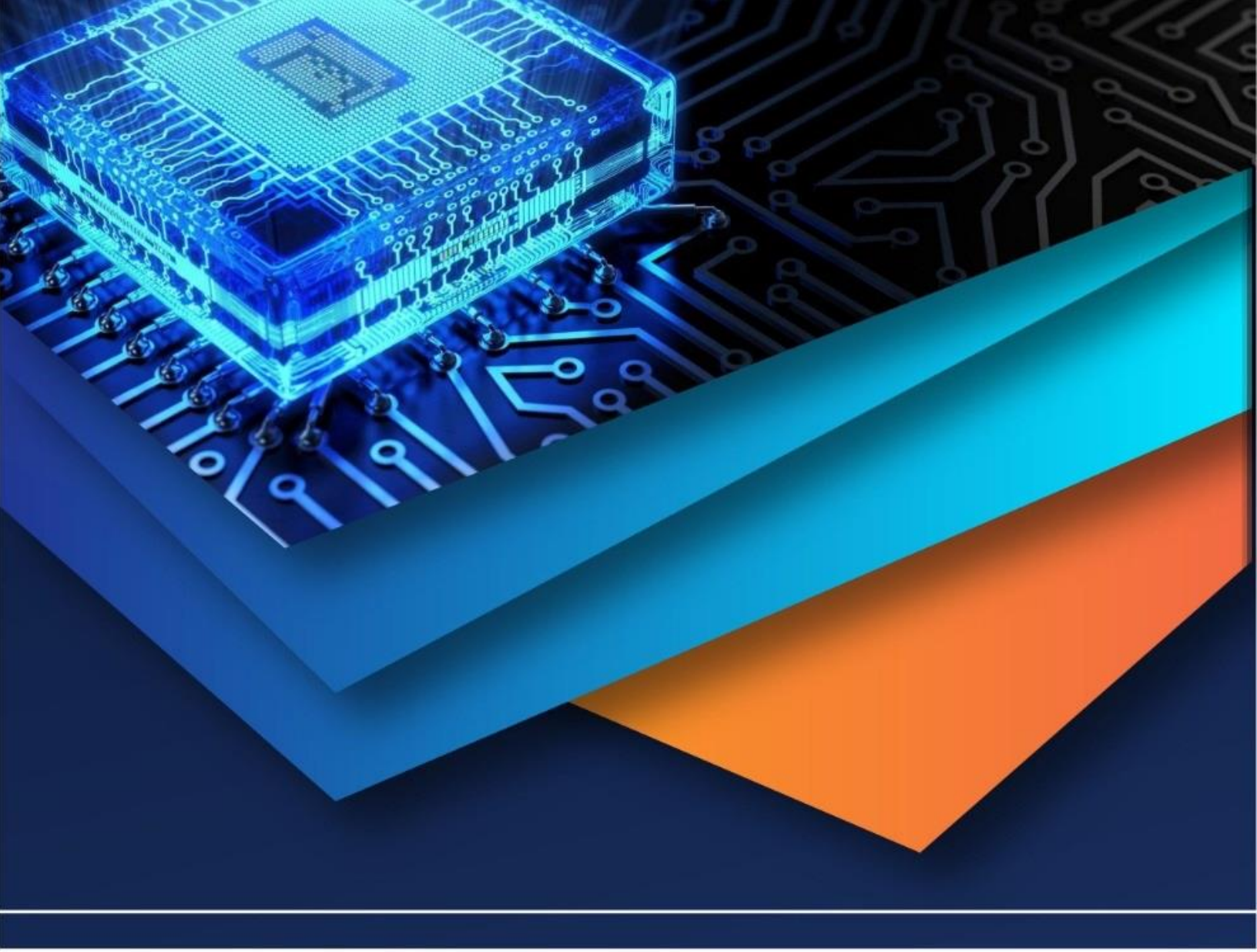
IV. CONCLUSION

In this study, different approaches were explored for music genre classification. The first approach involved treatingthe audio signal as an image by generating a spectrogram. A CNN-based image classifier, specifically VGG-16, wastrained on these spectrogram images to predict music genres. The second approach focused on extracting time and frequency domain features from the audio signals and using traditional machine learning classifiers. XGBoost was identified as the most effective feature-based classifier, and important features were identified. Other few machine learning algorithms and techniques were also used such as KNN , SVM Randomforest for comparitive study as a result of outcomes CNN is performing better. The results indicated that the CNN-based deep learning models outperformed the feature-engineered models. It was also found that ensembling the CNN and XGBoost models yielded additional benefits. The study utilized a dataset consisting of audio clips from YouTube videos, **which** generally have high levels of noise. Future research could explore preprocessing techniques to improve the performance of machine learning models on noisy data. The paper primarily utilized four machine learning models (k-nearest neighbors, SVM, logistic regression, and random forest). Initially, without employing proper techniques for hyperparameter tuning, the models performed poorly. However, by using techniques like GridSearchCV and RandomizedSearchCV to find optimal hyperparameter combinations, significant improvements in accuracy were achieved. Additionally, selecting relevant features improved the performance of the models. Moving forward, the researchers plan to explore other deep learning approaches and investigate the impact of parameter optimization on the predictions of deep learning models. The study focused on music genre classification using the Free Music Archive small (fma_small) dataset. A simple approach was proposed, and comparisons were made with more complex models. Two types of inputs were used: spectrogram images for CNN models and audio features stored in aCSV file for logistic regression and ANN models. The simple ANN model achieved the best performance among the feature-based classifiers, while the CNN model outperformed the other spectrogram-based models. Image-based classification consistently performed better than feature-based classification.

REFERENCES

[1] Music Genre Classification Using Independent Recurrent Neural Network, Wenli Wu ; Fang Han ;GuangxiaoSong ; Zhijie Wang, 2018 Chinese Automation Congress (CAC)
[2] Bahuleyan Hareesh. Music genre classification using machine learning techniques. University of Waterloo, n.d.
[3] Thomas Lidy and Andreas Rauber. 2005. Evaluation of feature extractors and psycho-acoustic transformationsfor music genre classification. In ISMIR. pages 34–41.
[4] Thomas Lidy and Alexander Schindler. 2016. Parallel convolutional neural networks for music genre and moodclassification. MIREX2016 .
[5] Aaron Van Den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. Wavenet: A generative model for raw audio. arXivpreprint arXiv:1609.03499 .
[6] Lonce Wyse. 2017. Audio spectrogram representations for processing with convolutional neural networks. arXivpreprint arXiv:1706.09559 .

[7] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu. 2014. Convolutional neural networks for speech recognition. IEEE/ACM Transactions on audio, speech, and languageprocessing 22(10):1533–1545.
[8] Alan V. Oppenheim. A speech analysis-synthesis system based on homomorphic filtering. Journal of the Acostical Society of America, 45:458–465, Februar Kristopher West and Stephen Cox. Features and classifiers for the automatic classification of musical audio signals. In International Symposium on Music Information Retrieval,2004.
[9] Bisharad, D.; Laskar, R.H. Music genre recognition using convolutional recurrent neural networkarchitecture. Expert Syst. **2019**
[10] Huang, A.; Wu, R. Deep Learning for Music. arXiv **2016**, arXiv:1606.04930.
[11] Abdoli, S.; Cardinal, P.; Koerich, A.L. End-to-end environmental sound classification using a 1D convolutional neural network. Expert Syst. Appl. **2019**, 136, 252–263.
[12] Murad, A.; Pyun, J.-Y. Deep Recurrent Neural Networks for Human Activity Recognition. Sensors **2017**, 17, 2556.
[13] H. G. Kim, N. Moreau, and T. Sikora, "Audio classification basedon MPEG-7 spectral basis representation," IEEETrans. Circuits Syst. Video Technol., vol. 14, no. 5, pp. 716–725, May 2004.
[14] H. G. Kim, N. Moreau, and T. Sikora, MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval. New York: Wiley, 2005.
[15] F. Mörchen, A. Ultsch, M. Thies, and I. Löhken, "Modeling timbre distance with temporal statistics from polyphonic music," IEEE Trans. Audio, Speech, Lang. Process., vol. 14, no. 1, pp. 81–90, Jan. 2006.
[16] T. Lambrou, P. Kudumakis, R. Speller, M. Sandler, and A. Linney, "Classification of audio signals using statistical features on time and wavelet transform domains," in Proc. IEEE Int. Conf. Acoustics,Speech, Signal Processing, 1998,vol.6,pp.3621–3624.
[17] Choi, K., Fazekas, G., & Sandler, M. (2017). Convolutional recurrent neural networks for music classification. In 18th International Society for Music Information Retrieval Conference (ISMIR).
[18] Hamel, P., Eck, D., & Schmidhuber, J. (2010). Learning genre-specific beat patterns for music classification and retrieval. IEEE Transactions on Audio, Speech, and Language Processing, 18(6), 1313-1323.
[19] Humphrey, E. J., Bello, J. P., & LeCun, Y. (2012). Feature learning and deep architectures: New directions for music informatics. Journal of Intelligent Information Systems, 38(1), 3-20.
[20] Van Den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. In Advances in Neural Information Processing Systems (pp. 2643-2651).
[21] Lee, C. M., & Slaney, M. (2009). Acoustic feature design for audio classification tasks. Journal of New Music Research, 38(3), 211-222.
[22] Pons, J., Nieto, O., Prockup, M., Schmidt, E. M., Ehmann, A. F., & Serra, X. (2017). End-to-end learning for music audio tagging at scale. In 25th European Signal Processing Conference (EUSIPCO) (pp. 1133-1137).
[23] Humphrey, E., Bello, J. P., & LeCun, Y. (2013). Feature learning and deep architectures: New directions for music informatics. In 14th International Society for Music Information Retrieval Conference (ISMIR).
[24] Sigtia, S., Benetos, E., Boulanger-Lewandowski, N., & Dixon, S. (2015). An end-to-end neural network for polyphonic piano music transcription. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(12), 2043-2056.
[25] Lee, K., Lee, K., & Choi, J. (2018). Music genre classification using convolutional neural network. In International Conference on Innovative Computing Technology (pp. 67-74).
[26] Music Genre Classification using Machine Learning Techniques Hareesh Bahuleyan University of Waterloo, ON, Canada hpallika@uwaterloo.ca
[27] Music Genre Classification UsingMachineLearning 1M.D.Nevetha, 2A.Nithyasree, 3A.Parveenbanu, 4Mrs.Jetlin CP 1Student, 2Student, 3Student, 4Assistant Professor Agni College of Technology
[28] Music Genre Classification using Machine Learning Algorithms: A comparison Snigdha Chillara1, Kavitha A S2, Shwetha A Neginhal3, Shreya Haldia4, Vidyullatha K S5 1,2,3,4,5Department of Information Science and Engineering, Dayananda Sagar College of Engineering, Bangalore – 560078, Karnataka, India



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY