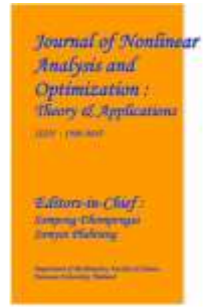
**J**ournal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**




# MUSIC GENRE CLASSIFICATION USING ADVANCED AUDIO ANALYSIS

**HIMAMBASHA SHAIK[1], JILLELLAMUDI KIRAN KUMAR[2]**

[1]Assistant Professor, Dept. of MCA, QIS College of Engineering and Technology, Ongole, Andhra Pradesh.

[2]PG Scholar, Dept. of MCA, QIS College of Engineering and Technology, Ongole, Andhra Pradesh.

**ABSTRACT—** A web application that pulls songs from YouTube and categorizes them into musical genres is presented in this project. The tool described in this study is based on models that were trained using Audioset's musical collection data. For this goal, we have used classifiers from diverse Machine Learning paradigms: Probabilistic Graphical Models (Naive Bayes), Feed-forward and Recurrent Neural Networks and Support Vector Machines (SVMs). These models were all trained in a situation involving multi-label categorization. We execute classification in 10-second chunks because genres can change throughout the course of a song. Audioset makes this possible by providing 10-second samples. In real time, the visualization output displays this temporal data in sync with the playing music video. The classification results are displayed as stacked area charts, which display scores for the top ten labels acquired for each chunk. We provide a thorough explanation of the problem's theoretical and scientific foundation as well as the suggested classifiers. In order to explain model performance and music genre classification issues, we first demonstrate the application's functionality in practice using three different songs as study cases. These are then examined and contrasted with internet categorizations.

Index Terms— Feature extraction, Music, Support vector machines, Classification algorithms

## I. INTRODUCTION

Research in Music Information Retrieval (MIR) comprises a broad range of topics including genre classification, recommendation, discovery and visualization. In short, this research line refers to knowledge discovery from music and involves its processing, study and analysis. When combined with Machine Learning techniques, we typically try to learn models able to emulate human abilities or tasks, which, if automated, can be helpful for the user. Computational algorithms and models have even been applied for music generation and composition Music genre classification (MGC) is a discipline of the music annotation domain that has recently received attention from the MIR research community, especially since the seminal study of Tzanetakis and Cook. Classifying a musical composition into one or more musical genres is the primary goal of MGC.

Even though it seems straightforward, the area nonetheless faces difficulties because of a lack of standards and imprecise definitions of genres. There is typically disagreement between public databases and tologies regarding the definition of each genre.

This agreement is further complicated by the fact that human perception of music is influenced by individual experiences and opinions. For instance, we would most likely classify a song as jazz music if it features improvisation, swing rhythms, piano, and trumpets. However, should a song also be categorized as electronic music if synthesizers are used in it? The answer is most likely yes if we solely take into account auditory properties. However, the piece might be interpreted differently by each listener. The song may be classified as jazz by some, electronic music by others, or perhaps a mix of the two.

We have trained a number of categorization models and integrated them into a web application that enables the user to see how each model ~`senses" music in terms of music genre, at specific points in a song, in an attempt to develop a tool that provides more insights about how each genre is perceived.

Note that each model's specific testing details are available elsewhere and outside the purview of this page. Common machine learning methods, such as Support Vector Machines (SVM), Naive Bayes classifiers, Feed Forward Deep Neural Networks, and Recurrent Neural Networks, were used to construct these models. Deep learning techniques in artificial perception (artificial vision, speech recognition, natural language processing, etc.) have produced amazing results that are almost human-like in accuracy, while Bayesian and SVM methods have historically produced good results as general-purpose machine learning models.

We also want to assess deep learning's performance in music genre classification by contrasting it with more conventional machine learning methods.

## II. LITERATURE SURVEY

### *A) music information retrieval*

The Survey of Music Information Retrieval Systems, which was presented at the Sixth International Conference on Music Information Retrieval in 2005, discusses the state of AR systems. 27 A difference between content-based search systems for generic "audio data" and search systems for "music based on the notes" is made in order to provide an overview of "Music Information Retrieval (MIR)." These are accompanied by "hybrid" systems, which were initially used to transform any kind of audio data into a symbolic representation of the notes. Different viewpoints exist about content-based search in relation to music databases. Users can look for pieces by humming or strumming from memory with search-by-humming. Depending on the kind of resemblance needed, musicologists can find compositions inspired by a melody using the conventional search-by-example method. Last but not least are searches focused on comparing entire soundtracks or specific tracks, which are helpful in "investigations" for copyright purposes into instances of quotation or plagiarism. Identifying songs broadcast by broadcasters, also through a "common receiver" linked to a treatment system, searching for "suspicious" sounds captured by surveillance systems, and sound analysis of video and any other application in television, radio, or other media industry archives are just a few of the many real-world uses for augmented reality techniques. Despite the novelty of its application, AR is making tasks faster and more efficient, and its applications are now present in a lot of commercial equipment. The survey moves on to describe the two techniques, AR or MIR, relative to ‘musical data’ structured on notes and ‘audio data’ in general. For musical data it is still necessary to distinguish between ‘monophonic and polyphonic melodies’. The most important issues in both cases are measuring differences between the compared data of the notes, which the system must be able to carry out automatically, and the construction of the data index, automatically or semi-automatically. The degree of matching is closely related to the "distance measure" and "indexing" processes, which are set each time for the retrieval of the document. The more general and broad the process, the easier it is for the system to estimate the similarity between the parameters of the notes being compared, or between a parameter and indexing terms used. Even "segmenting" sound records into sections that reflect their structure is necessary to identify other elements in audio data that is not reliant on note systems.

These automatically observable characteristics tempo, frequency, loudness, timbre, tone, etc. are typical to each sound item.

Finding a method that can combine the findings of a study of a track to provide a model of its audio characteristics that is both adequate and dependable is the challenge. This can be achieved, for instance, by creating vectors like an audio fingerprint, or "Self-organizing Map (SOM)". Quick summaries and comparisons of the top 17 AR systems, as well as the many user needs and traits, round out this panorama. Three user classes "industrial, professional, and general consumers" are considered by the writers. These classes, to varying degrees of research, need single sound outputs, full tracks, information about composers, musical genres and classes of sounds. Objectives can be varied: copyright protection; search for music based on tastes and styles; search for the works of a given artist; and identification of tracks, etc.

### *B) The batch doodle: Approachable music composition with machine learning at scale*

We created the Bach Doodle, the first AI-powered Google Doodle, to make music composition more approachable. It allows users to compose their own melody and has it harmonized in the style of Bach by a machine learning model called Coconet (Huang et al., 2017). We created a simplified interface based on sheet music so that users could enter melodies. We re-implemented Coconet in TensorFlow.js (Smilkov et al., 2019) to run in the browser and decreased its runtime from 40s to 2s by utilizing fusing operations and dilated depth-wise separable convolutions in order to offer an interactive experience at scale. We also used post-training weight quantization to lower the model download size to about 400KB. To ascertain whether the harmonization request should be handled locally or transmitted to distant TPU servers, we calibrated a speed test based on partial model assessment time. People played with the Bach Doodle for 350 years in three days, and Coconet received over 55 million requests. We are sharing this work as part of a public dataset where users can choose to rate their compositions. We believe that this dataset will be helpful to the community for a variety of applications, such as enhancing machine learning models, ethnomusicological research, and music education.

### C) *Deep learning techniques for music generation A survey*

In this research, various approaches to producing musical material utilizing deep learning (deep artificial neural networks) are surveyed and analyzed. For our analysis, we provide a technique based on five dimensions: Goal: What kind of music is to be produced? Melody, polyphony, counterpoint, and accompaniment are a few examples. For what purpose and to what end? to be played either by a machine (for an audio file) or by a human (for a musical score). Representation: Which ideas need to be changed? Waveform, spectrogram, note, chord, meter, and beat are a few examples. Which format should be used? Text, piano roll, and MIDI are a few examples. What encoding method will be used for the representation? Scalar, one-hot, and many-hot are some examples. Architecture: Which deep neural network type or types are to be used? Recurrent networks, feedforward networks, autoencoders, and generative adversarial networks are a few examples. Challenge: What are the restrictions and unresolved issues? Variability, interactivity, and inventiveness are a few examples. Strategy: How can the generation process be modeled and managed? Examples include input manipulation, sampling, iterative feedforward, and single-step feedforward. We perform a comparative study of several theories and methods for each dimension and suggest a preliminary multidimensional typology. This bottom-up typology was developed by analyzing numerous deep learning-based music generating systems that have been developed and chosen from the pertinent literature. The different choices of objective, representation, architecture, difficulty, and strategy are explained and illustrated using these systems. There is considerable discussion and some prospects in the final portion.

## III. PROPOSED SYSTEM

The overview of our proposed system is shown in the below figure.

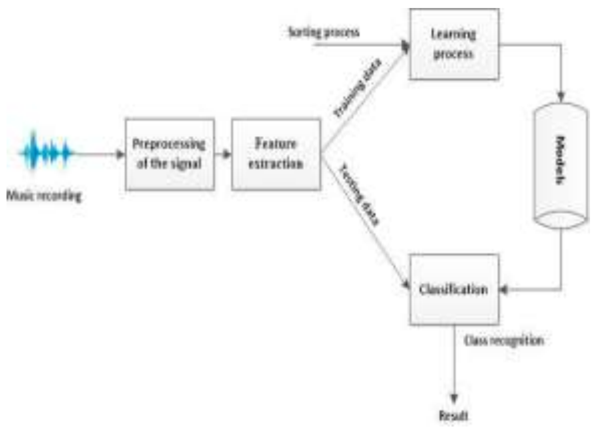



Fig. 1: System Overview

### *Implementation Modules*

**User Login:**

- Using this module user can login to application and after login can train with SVM, LSTM and then classify music genre

**New User Signup Here:**

- This module allows users to register for the application and subsequently log in.

**Train SVM:**

- This module uses the MFCC technique to extract features from the dataset. The retrieved features are then trained using SVM, and a confusion matrix graph is used to calculate accuracy, average precision, AUC, and recall. In this case, the extracted features dataset will be divided into train and test, with 20% going toward testing and 80% going toward training.

**Train Decision Tree:**

- This module uses the MFCC technique to extract features from the dataset. The retrieved features are then trained using a decision tree, and a confusion matrix graph is used to calculate accuracy, average precision, AUC, and recall. In this case, the extracted features dataset will be divided into train and test, with 20% going toward testing and 80% going toward training.

**Train LSTM:**

- This module uses the MFCC technique to extract features from the dataset. The retrieved features are then trained using LSTM, and a confusion matrix graph is used to calculate accuracy, average precision, AUC, and recall. In this case, the extracted features dataset will be divided into train and test, with 20% going toward testing and 80% going toward training.

**Train Feed Forward Network:**

- Using this module we extract features from dataset using MFCC algorithm and this extracted features will get train with Feed Forward Neural Network and then will calculate accuracy, average precision, AUC and recall with confusion matrix graph.

- Here extracted features dataset will be split into train and test where 80% data used for training and 20% for testing

**Music Genre Classification:**

- With the help of this module, users can upload test audio files from the "test Music Files" folder, and LSTM will identify the genre of the music they have uploaded.

## IV. RESULTS

Fig. 2: Home Page

Fig. 3: Registration

Fig. 4: Upload Music File

Fig. 5: Music Genre Classification

## V. CONCLUSION

The article presents a web application to discover music genres present in a song, along its timeline, based on a previous experimentation with different machine learning models. By identifying genres in each 10-second fragment, we can get an idea of how each model perceives each part of a song. Moreover, by presenting those data in a stacked area timeline graph, the application is also able to quickly show the behavior of the models, which at the same time, is an interesting way to detect undesired or rare predictions. We believe that this application could be a supporting tool for the traditional evaluation metrics in MGC, especially when manual introspection of questionable results is required beyond classic performance metrics, such as average precision or AUC. There is no standard or formal method of dening genres, so it is difficult to establish a formal way to validate genre predictions, especially when attempting to compare them with categorizations from other sources, like online music platforms. For instance, Last.fm has a completely different set of tags that, in many cases, do not correspond or exist in the Audioset ontology. The application is also a prelude to a future user-centered MGC tool where users can provide feedback regarding the accuracy of the predictions. To our knowledge, no visual tool offers this level of verification on genre classification results for various song fragments.

## REFERENCE


[1] J. S. Downie, ``Music information retrieval,'' Annu. Rev. Inf. Sci. Technol., vol. 37, no. 1, pp. 295340, 2003.

[2] C.-Z. A. Huang, C. Hawthorne, A. Roberts, M. Dinculescu, J. Wexler, L. Hong, and J. Howcroft, ``The bach doodle: Approachable music composition with machine learning at scale,'' 2019, arXiv:1907.06637. [Online]. Available: http://arxiv.org/abs/1907.06637

[3] M. A. Shaik, Y. Sahithi, M. Nishitha, R. Reethika, K. Sumanth Teja and P. Reddy, "Comparative Analysis of Emotion Classification using TF-IDF Vector," 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2023, pp. 442-447, doi: 10.1109/ICSSAS57918.2023.10331897.

[4] H. Li, ``Piano automatic computer composition by deep learning and blockchain technology,'' IEEE Access, vol. 8, pp. 188951188958, 2020.

[5] G. Tzanetakis and P. Cook, ``Musical genre classication of audio signals,'' IEEE Trans. Speech Audio Process., vol. 10, no. 5, pp. 293302, Jul. 2002.

[6] J. Ramírez and M. J. Flores, ``Machine learning for music genre: Multifaceted review and experimentation with audioset,'' J. Intell. Inf. Syst., vol. 59, pp. 469499, Nov. 2019.

[7] M. A. Shaik, M. Azam, T. Sindhu, K. Abhilash, A. Mallala and A. Ganesh, "Hand Gesture Based Food Ordering System," 2023 International Conference on Self Sustainable Artificial Intelligence

Systems (ICSSAS), Erode, India, 2023, pp. 867-872, doi: 10.1109/ICSSAS57918.2023.10331637.

[8] "Classication of musical genre: A machine learning approach," in Proceedings of the 5th ISMIR Conference, Barcelona, Spain, 2004, by R. Basili, A. Serani, and A. Stellato.

[9] In Proc. 8th ISMIR Conf., Vienna, Austria, 2007, pp. 425430, J.-J. Aucouturier, F. Pachet, P. Roy, and A. Beurivé, "Signal C context= better classication."

[10] M. Parveen and M. A. Shaik, "Review on Penetration Testing Techniques in Cyber security," 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2023, pp. 1265-1270, doi: 10.1109/ICAISS58487.2023.10250659.

**AUTHORS Profile**

**Mr. Himambasha Shaik** is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his Master of Computer Applications (MCA) from Anna University, Chennai. With a strong research background, He has authored and co-authored research papers published in reputed peer-reviewed journals. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing, and Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

**Mr. JILLELLAMUDI KIRAN KUMAR** has received him B.S.C (Computer Science) and degree from ANU 2022 and pursuing MCA in QIS College of Engineering and Technology affiliated to JNTUK in 2023-2025.