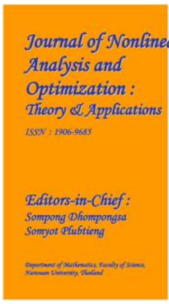
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**



CATEGORISATION OF ONLINE MUSIC GENRES FOR THE VISUALISATION AND ANALYSIS OF SONG TIMELINES

S SUNDEEP KUMAR, PG SCHOLAR, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, SREEDATTHA INSTITUTE OF ENGINEERING AND SCIENCE SHEERIGUDA, IBRAHIMPATNAM HYDERABAD, TELANGANA, INDIA.
Dr. S.V. ACHUTA RAO, PROFESSOR, DEAN ACADEMICS & DIRECTOR RECRUITMENT IN DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, SREE DATTHA INSTITUTE OF ENGINEERING AND SCIENCE, SHERIGUDA IBRAHIMPATNAM HYDERABAD TELANGANA, INDIA.

ABSTRACT—
An online tool for retrieving and categorising music from YouTube is shown in this project. Models trained using Audioset's musical collecting data form the basis of the technique described in this paper. Classifiers from many ML paradigms, including Support Vector Machines (SVMs), Feed-forward and Recurrent Neural Networks, and Probabilistic Graphical Models (Naive Bayes), have been used for this objective. These models were all trained using a scenario with multiple labels. Genres might change as a song progresses, therefore we classify every 10 seconds. Audioset makes this possible by providing 10-second samples. Presented in stacked area charts, the visualisation output shows the results of the categorisation, with the top ten scores shown in real time, all while the music video is playing in the background. outputs the labels acquired for each piece. The technical and theoretical underpinnings of the issue and the suggested classifiers are thoroughly explained. We then demonstrate the application's functionality by investigating three separate songs; we compare and contrast their classifications with those found online in order to talk about the models' efficacy and the difficulties of music genre identification.
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. The main objective in MGC is to classify a musical piece into one or more musical genres.

As simple as it sounds, the field still presents challenges related to the lack of standardization and vague genre definitions. Public databases and on tologies do not usually agree on how each genre is defined.

Moreover, human music perception, subject to opinions and personal experiences, makes this agreement even more difficult. For example, when a song includes swing rhythms, piano, trumpets and improvisation, we would probably define it as jazz music. However, if we introduce synthesizers in the same song, should the song be classified as electronic music as well? If we only consider acoustic characteristics, the answer is probably yes. But different listeners can perceive the piece from their own perspective. Whereas some might categorize the song as jazz, others might consider it electronic music or even a combination of both.

In an effort to provide a tool that gives more insights about how each genre is perceived, we have trained several classification models and embedded them in a web application that allows the user to visualize how each model
``senses" music in terms of music genre, at particular moments of a song. Note that experimentation details for each model are beyond the scope of this article and can be found in. These models have been built using common machine learning techniques, namely, Support Vector Machines (SVM), Naive Bayes classiers, Feed forward deep neural networks and Recurrent neural networks. Whereas Bayesian and SVM methods have historically delivered good results as general purpose machine learning models, the results achieved with deep learning techniques in artificial perception (artificial vision, speech recognition, natural language processing, among others) have delivered remarkable results, approaching human-like accuracy. By comparing deep learning with more traditional machine learning techniques, we also aim to compare its performance for music genre classification.

II. LITERATURE SURVEY

A) music information retrieval

The status of AR systems is covered in the Survey of Music Information Retrieval systems, presented at the Sixth International Conference on Music Information Retrieval in 2005.27 In illustrating a summary of 'Music Information Retrieval (MIR)', a distinction is made between the content- based search systems of general 'audio data' and search systems for 'music based on the notes'. Alongside these are the 'hybrid' systems, which in the early treatment of any type of audio data were converted into a symbolic version of the notes. With reference to music databases, content-based search has different perspectives. Search-by- humming allows users to search for pieces by humming, or strumming from memory. The traditional search-by-example, according to the type of similarity required, is useful for musicologists searching for pieces inspired by a melody. Lastly come searches orientated towards comparing whole soundtracks or their parts, proving useful in 'investigations' for copyright purposes into cases of plagiarism or quotation. AR techniques have numerous practical applications: identifying songs transmitted by broadcasters, also via a 'common receiver' connected to a treatment system; search for 'suspicious' sounds recorded by surveillance systems; and sound analysis of video and any type of application in television, radio or other media industry archive. 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. 'Distance measure' and 'indexing' are processes closely linked to the degree of matching, set each time for the document's retrieval, and the more broad and generic it is, the more the system can easily estimate the similarity between the parameters of the notes being compared, or between a parameter and indexing terms used. For audio data not based on systems of notes, other features need to be singled out, even by 'segmenting' sound tracks into parts representative of their structure.
These automatically detectable features are those typical to each sound object, namely tempo, frequency, amplitude, timbre, tone etc. The problem is in finding a scheme capable of composing the results of a track's analysis in order to obtain a satisfactory and reliable enough model of its audio features. This is feasible, for example, by composing vectors such as audio-fingerprint, or as it is known, a 'Self-organizing Map (SOM)'. This panorama continues with quick descriptions and comparisons of the 17 most advanced AR systems, and the differing needs and characteristics of users. The authors take into account three classes of user, namely 'industrial, professional, and general consumers'. 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

To make music composition more approachable, we designed the first AI- powered Google Doodle, the Bach Doodle, where users can create their own melody and have it harmonized by a machine learning model Coconet (Huang et al., 2017) in the style of Bach. For users to input melodies, we designed a simplified sheet-music based interface. To support an interactive experience at scale, we re-implemented Coconet in TensorFlow.js (Smilkov et al., 2019) to run in the browser and reduced its runtime from 40s to 2s by adopting dilated depth-wise separable convolutions and fusing operations. We also reduced the model download size to approximately 400KB through post-training weight quantization. We calibrated a speed test based on partial model evaluation time to determine if the harmonization request should be performed locally or sent to remote TPU servers. In three days, people spent 350 years worth of time playing with the Bach Doodle, and Coconet received more than 55 million queries. Users could choose to rate their compositions and contribute them to a public dataset, which we are releasing with this paper. We hope that the community finds this dataset useful for applications ranging from ethnomusicological studies, to music education, to improving machine learning models.

C) Deep learning techniques for music generation A survey

This paper is a survey and an analysis of different ways of using deep learning (deep artificial neural networks) to generate musical content. We propose a methodology based on five dimensions for our analysis: Objective - What musical content is to be generated? Examples are: melody, polyphony, accompaniment or counterpoint.
- For what destination and for what use? To be performed by a human(s) (in the case of a musical score), or by a machine (in the case of an audio file). Representation - What are the concepts to be manipulated? Examples are: waveform, spectrogram, note, chord, meter and beat. - What format is to be used? Examples are: MIDI, piano roll or text. - How will the representation be encoded? Examples are: scalar, one-hot or many-hot. Architecture - What type(s) of deep neural network is (are) to be used? Examples are: feedforward network, recurrent network, autoencoder or generative adversarial networks. Challenge - What are the limitations and open challenges? Examples are: variability, interactivity and creativity. Strategy - How do we model and control the process of generation? Examples are: single- step feedforward, iterative feedforward, sampling or input manipulation. For each dimension, we conduct a comparative analysis of various models and techniques and we propose some tentative multidimensional typology. This typology is bottom-up, based on the analysis of many existing deep-learning based systems for music generation selected from the relevant literature. These systems are described and are used to exemplify the various choices of objective, representation, architecture, challenge and strategy. The last section includes some discussion and some prospects.

III. PROPOSED SYSTEM

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

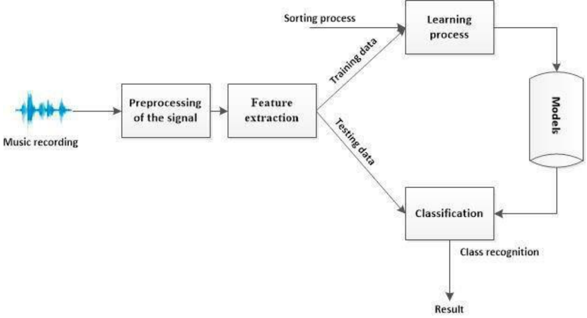


Fig. 1: System Overview

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:

- Using this module user can signup with the application and then can login

Train SVM:

- Using this module we extract features from dataset using MFCC algorithm and this extracted features will get train with SVM 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

Train Decision Tree:

- Using this module we extract features from dataset using MFCC algorithm and this extracted features will get train with Decision Tree 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

Train LSTM:

- Using this module we extract features from dataset using MFCC algorithm and this extracted features will get train with LSTM 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

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:

- Using this module user can upload test audio files from 'test Music Files' folder and then LSTM will predict/classify type of that uploaded music Genre

Implementation Algorithms

Support Vector Machine

- In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. An SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

Decision Tree

- Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems.
- It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions

LSTM

- Long Short-Term Memory is an improved version of recurrent neural network designed by Hochreiter & Schmidhuber.
- A traditional RNN has a single hidden state that is passed through time, which can make it difficult for the network to learn long-term dependencies. LSTMs model address this problem by introducing a memory cell, which is a container that can hold information for an extended period.
- LSTM architectures are capable of learning long-term dependencies in sequential data, which makes them well- suited for tasks such as language translation, speech recognition, and time series forecasting

IV. RESULTS

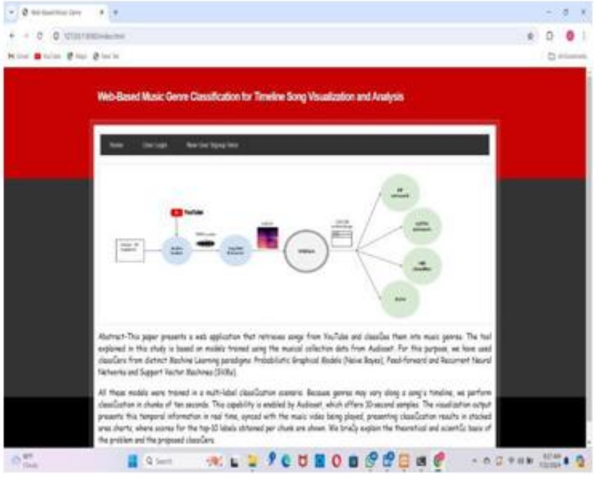


Fig. 2: Home Page

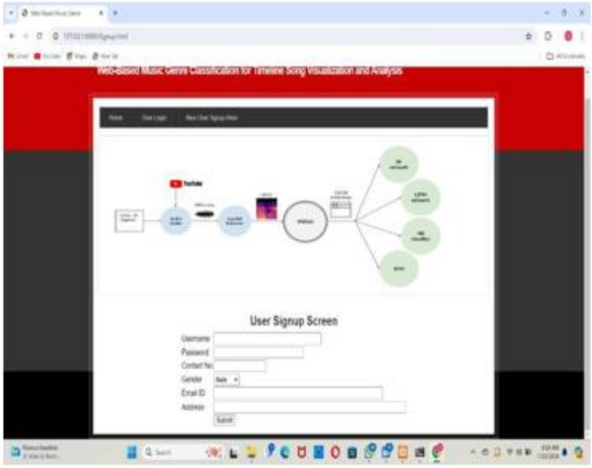


Fig. 3: Registration

Fig. 4: Upload Music File

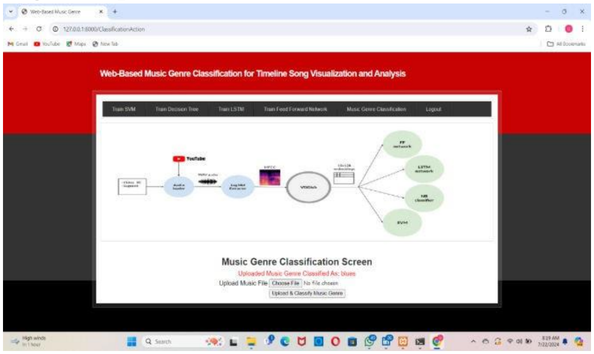


Fig. 5: Music Genre Classification

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. It is, in any case, a challenge to establish a formal way to validate genre predictions, particularly when trying to compare them with categorizations from other sources, such as online music platforms, because there is no standard or formal way of dening genres. Last.fm, to name an example, has a completely different set of tags, which, in many cases, do not correspond or exist in the Audioset ontology. The application is also a first step towards an eventual user-centered MGC tool, in which the users can submit feedback about the correctness of the predictions. To our knowledge, there is no visual tool that provides this level of verification on genre classification results for different fragments of the song.

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] J.-P. Briot, G. Hadjeres, and F.-D. Pachet, ``Deep learning techniques for music generationA survey,'' 2017, arXiv:1709.01620. [Online]. Available: http://arxiv.org/abs/1709.01620

[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] K. He, X. Zhang, S. Ren, and J. Sun, ``Delving deep into rectiers: Surpassing human-level performance on ImageNet classication,'' in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 10261034.

[8] R. Basili, A. Serani, and A. Stellato, ``Classication of musical genre: A machine learning approach,'' in Proc. 5th ISMIR Conf., Barcelona, Spain, 2004.

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

[10] T. D. Nielsen and F. V. Jensen, Bayesian Networks and Decision Graphs. New York, NY, USA: Springer, 2009.