Automated Determination of Music Genre and Linguistic Categorization

Mrs.A.V.Lakshmi Prasuna
Project Guide
(Assistant Professor)
Avlakshmiprasuna_it@mgit.ac.in

Mali Rakesh
Student(MGIT)
Mrakesh_it201230@mgit.ac.in

Bachu Sumanth
Student (MGIT)
bsumanth_it201208@mgit.ac.in

Seetha Sathvik Ganesh
Student(MGIT)
ssathvikganesh_it201248@mgit.ac.in

Abstract-The main objective of the project is to build a system that allows users to predict genre as well as language out of some music . The Automated determination of musical genre and linguistic categorization is a machine learning based prediction model which is used to predict the genres and languages of an audio . Musical genres are categorical labels created by humans to characterize piece of music. A musical genre is characterized by common characteristics typically related to the instrumentation, rhythmic structure and harmonic content of the music. The performance and relative importance of the proposed features are investigated by training the statistical pattern recognition classifiers using real world audio collections and using machine learning algorithm.In this project, we built systems called music genre and language classification and the effects of different feature extraction methods and their various combinations were also investigated. This task is of great interest in music information retrieval and recommendation systems. In recent years, several machine learning algorithms have been developed to tackle this task, using features such spectral, temporal, and tonal characteristics of audio signals. Music genre prediction and language classification is a task that involves automatically predicting a piece of music into a specific genre category as well as language based on its audio features.

Keywords-machine-learning,music,music genre and language classification

1. INTRODUCTION

This introduction recognizes the importance of contextual understanding, both historical and contemporary, in decoding the complexities of music genres and linguistic frameworks. Drawing parallels with the achievements in other technological spheres, such as the cryptocurrency space with Bitcoin's milestones, it highlights the role of research, innovation, and technological breakthroughs in shaping the trajectory of automated determination systems.

Similar to the challenges encountered in adapt to the evolving nature of cultural and linguistic expressions.

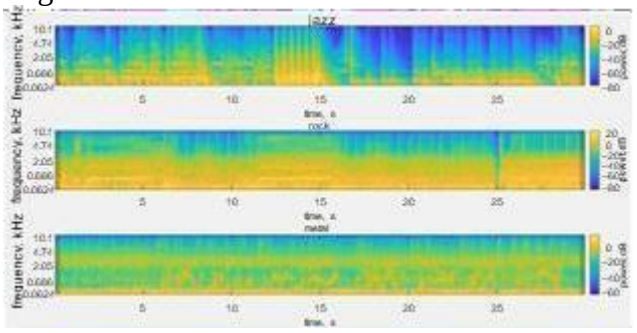
In alignment with the relentless progress witnessed in technological domains, the Automated Determination of Music Genre and Linguistic Categorization field leverages advancements in machine learning, linguistic analysis tools, and sophisticated music recommendation systems. These tools contribute to the refinement of automated systems, enhancing their ability to discern and categorize the ever-changing landscape of musical genres and linguistic structures.

This introduction recognizes the importance of contextual understanding, both historical and contemporary, in decoding the complexities of music genres and linguistic frameworks. Drawing parallels with the achievements in other technological spheres, such as the cryptocurrency space with Bitcoin's

milestones, it highlights the role of research, innovation, and technological breakthroughs in shaping the trajectory of automated determination systems.

As we delve into this multifaceted field, the text emphasizes the global attention and adoption of automated systems for music genre determination and linguistic categorization. Notable achievements, akin to Bitcoin's milestones, are explored, underlining the ongoing efforts of countries to integrate these automated processes into their systems, fostering secure and efficient linguistic processing.

In essence, the Automated Determination of Music Genre and Linguistic Categorization stands at the intersection of technology, culture, and creativity, promising a nuanced understanding of musical and linguistic trends



(b)

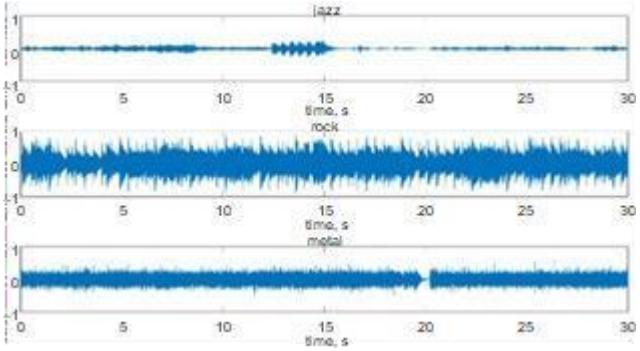


Fig. 1 (a) Example time domain representation of Jazz, Rock and Metal
(b) Time-frequency representations (spectrograms) of the signals given in Fig. 1(a)

2. LITERATURE SURVEY

1,Music genre classification and music recommendation by using deep learning[1] A. Elbir✉ and N. Aydin

A flow chart of the conducted experiment is illustrated in Fig. 2. Initially, each music in the data set is divided into six parts with a duration of 5 s. Melspectrogram is generated from sampled each 5 s music and saved as an image. Then, this image is applied to the proposed MusicRecNet for training. The MusicRecNet, which is shown as the last block in Fig. 2, is a type of CNN that new layers and artificial dropout features have been added to minimise validation error. Specifically, in our experiments, MusicRecNet is designed to have three layers. Each layer consists of a two-dimensional convolution, an activation function (rectified linear unit deep neural network model), a two-dimensional maximum pooling operation and a dropout operation. The parameters used in the proposedMusicRecNet

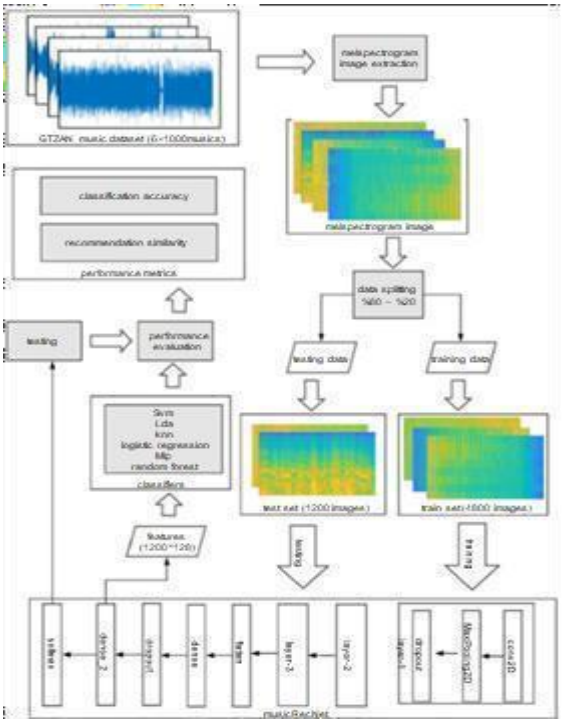


Fig. 2 Block diagram of the proposed study

After the training the MusicRecNet, the model is used for genre classification. Additionally, the last layer

of the model named as Dense_2 is used as a feature vector of the test music samples for music genre classification, music similarity and music recommendation. We implemented classification algorithms, some of which are ensemble and model based methods given in Table 1.

Classifier	Avg accuracy, %	
	Five-fold	Ten-fold
MLP	91.2	92.2
logistic regression	97.6	98.4
random forest	87.7	88.7
LDA	96.1	96.3
KNN	88.6	87.6
SVM(Poly-3)	97.6	97.9

Table 1: Classification performance accuracy

Experimental results: Since GTZAN data set consists of ten different genres, accuracy was used as the main performance metric. Although it can be assumed as a subjective measure in view of a listener, the average percentage of music similarity is also used as a metric for quality of music recommendation. Additionally, a confusion matrix, from which precision, recall and F-measure scores

can be obtained, is used. An example of the confusion matrix is illustrated in Fig. 4. Mean accuracy of the proposed MusicRecNet is 81.8% when it is used as a standalone classifier. In Table 3, obtained MusicRecNet classification results are compared to the results of other studies. According to these results, it is obvious that the MusicRecNet outperformed the other classifiers.

2. Music Recommendation System Miao Jiang, Ziyi Yang, Chen Zhao,
"An RNN-based music recommendation system." [1] In the very recent years, development of music recommendation system has been a more heated problem due to a higher level of digital songs consumption and the advancement of machine learning techniques. Some traditional approaches such as collaborator filtering, has been widely used in recommendation systems, have helped music recommendation system to give music listeners a quick access to the music. However, collaborative filtering or model-based algorithm have
limitations in giving a better result with the ignorance of combination factor of lyrics and genre. In our paper, we will propose an improved algorithm based on deep neural network on measure similarity between different songs. The proposed method will make it possible that it could make recommendations in a large system to make comparisons by "understand" the content of songs.

Sushmita G. Kamble and Asso. Prof. A. H. Kulkarni, "Facial Expression Based Music Player." [2] Conventional method of playing music depending upon the mood of a person requires human interaction. Migrating to the computer vision technology will enable automation of such system. To achieve this goal, an algorithm is used to classify the human expressions and play a music track as according to the present emotion detected.

Jie Liu, Liang Jia, "The Application of Computer Music Technology in Music Education" [3] With the rapid development of Internet technology, computers are widely used in all walks of life, and among them, computer music technology has become one of the important tools in modern music education. Analysis of the traditional music teaching mode can be found, most of the time is limited to the teacher performance, student imitation, however, for the new era of music education, the combination of computer technology and music knowledge has brought great convenience for the development of music education

George Tzanetakis, Student Member, IEEE, and Perry Cook, Member, IEEE, "Musical Genre Classification of Audio Signals. "[4] Three feature sets for representing timbral texture, rhythmic content and pitch content are proposed. Features is investigated by training statistical pattern recognition classifiers using real-world audio collections.

Liang Stamatakis, Jie Liu, "An exploration of the application of computer music production software in music composition" [5] The gradual implementation of computer music production software within the current stage of music composition and traditional composition methods for integration is important for the creation of excellent musical works with unique styles and novel ideas.

Federico Simonetta, Stavros Ntalampiras, "Multimodal Music Information Processing and Retrieval: Survey and Future Challenges" [6] Towards improving the performance in various music information processing tasks, recent studies exploit different modalities able to capture diverse aspects of music. Such modalities include audio recordings, symbolic music scores, mid-level representations, motion and gestural data, video recordings, editorial or cultural tags, lyrics and album cover arts.

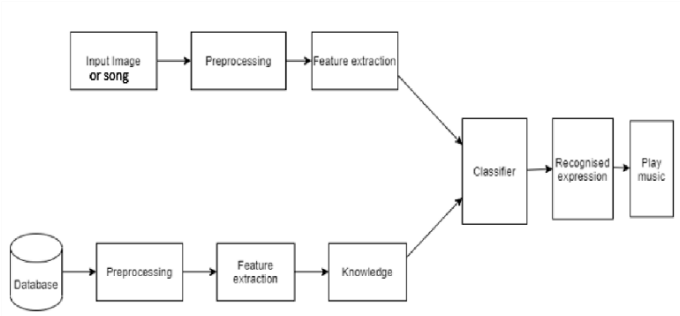


Fig .2 System Architecture

CNN stands for Convolutional Neural Networks, which are specialized for image and video recognition applications. Image recognition, object detection, and segmentation are among of the most common image analysis tasks that CNN is employed for. Convolutional Neural Networks have four different sorts of layers:

1) Convolutional Layer: Each input neuron in a conventional neural network is linked to the next hidden layer. Only a small portion of the input layer neurons connect to the hidden layer neurons in CNN.

2) Pooling Layer: The pooling layer is used to minimize the feature map's dimensionality. Inside the CNN's hidden layer, there will be several activation and pooling layers.

3) Flatten: Flattening is the process of transforming data into a one-dimensional array for use in the next layer. To construct a single lengthy feature vector, we flatten the output of the convolutional layers.

4) Fully Connected Layers: Fully Connected Layers are the network's final layers. The output from the final Pooling or Convolutional Layer, which is flattened and then fed into the fully connected layer,

is the input to the fully connected layer.

Support Vector Machines:

It is supervised machine learning algorithm. It is used for classification ad regression. In this algorithm, features are extracted from input ad plotted it on dimensional space where n is number of features i.e., each coordinate is value of particular feature. SVM works as follow:

1) Plotting features on n-dimensional space
2) Identifying right hyper-plane
3) Classifying the data according to that hyperplane

Harcascade:

The algorithm can be explained in four stages:

1. Calculating Haar Features
2. Creating Integral Images
3. Using Adaboost
4. Implementing Cascading Classifiers

It's important to remember that this algorithm requires a lot of positive images of faces and negative images of non-faces to train the classifier, similar to other machine learning models. Calculating Haar Features

The first step is to collect the Haar features. A Haar feature is essentially calculations that are performed on adjacent rectangular regions at a specific location in a detection window. The calculation involves summing the pixel intensities in each region and calculating the differences between the sums

3. Multimodal Deep Learning for Music Genre Classification Sergio Oramas[*,‡], Francesco Barbieri[†], Oriol Nieto[‡] and Xavier Serra[*]

In this work they have proposed a representation learning approach for the classification of music genres from different data modalities, i.e., audio, text, and images. The proposed approach has been applied to a traditional classification scenario with a small number of mutually exclusive classes. It has also been applied to a multi-label classification scenario with hundreds of non- mutually exclusive classes. In addition, we have proposed an approach based on the learning of a multimodal feature space and a dimensionality reduction of target labels using PPMI.

Results show in both scenarios that the combination of learned data representations from different modalities yields better results than any of the modalities in isolation. In addition, a qualitative analysis of the results has shed some light on the behavior of the different modalities. Moreover, we have compared our neural model with a human annotator, revealing correlations and showing that our deep learning approach is not far from human performance.

In our single-label experiment we clearly observed how visual features perform better in some classes where audio features fail, thus complementing each other. In addition, we have shown that the learned multimodal feature space seems to improve the performance of audio features. This space increases accuracy, even when the visual part is not present in the prediction phase. This is a promising result, not only for genre classification, but also for other applications such as music recommendation, especially when data from different modalities are not always available for every item. However, more experimentation is needed to confirm this finding.

In our multi-label experiment we provide evidence of how representation learning approaches for audio classification outperform traditional handcrafted feature based approaches. Moreover, we compared the effect of different design parameters of CNNs in audio classification. Text-based approaches seem to outperform other modalities, and benefit from the semantic enrichment of texts via entity linking. While the image-based classification yielded the lowest performance, it helped to improve the results when combined with other modalities. Furthermore, the dimensionality reduction of target labels led to better results, not only in terms of AUC, but also in terms of aggregated diversity.

To carry out the experiments, we have collected and released two novel multimodal datasets for music genre classification: first, ***MSD-I***, a dataset with over 30k audio tracks and their corresponding album cover artworks and genre annotations, and second, ***MuMu***, a new multimodal music dataset with over 31k albums, 147k audio tracks, and 450k album reviews.

To conclude, this work has deeply explored the classification problem of music genres from different perspectives and using different data modalities, introducing novel ideas to approach this problem coming from other domains. In addition, we envision that the proposed multimodal deep learning approach may be easily applied to other MIR tasks (e.g., music recommendation, audio scene classification, machine listening, cover song identification). Moreover, the release of the gathered datasets opens up a number of potentially unexplored research possibilities.

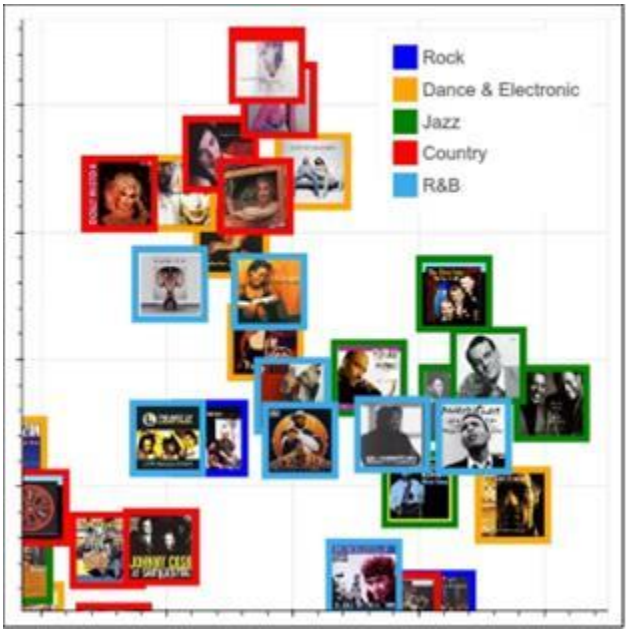


Figure 3: t-SNE visualization of image vectors from the ***single-parent-label*** subset

4. Music Genre Classification: A Review of Deep-Learning and Traditional Machine-Learning Approaches Ndiatenda Ndou , Ritesh Ajoodha, Ashwini Jadhav

This work aimed at automatic music genre classification using deep-learning and traditional machine-learning models. A review of related literature revealed the capability of these classifiers and a benchmark to compare the work of this research. We note that the reliability of a learning model is dependent on the quality of its ground truth, therefore, it is essential to ensure the ground truth is well-founded and motivated.

This research was conducted in three phases, namely, ‘phase A’, ‘phase B’, and, ‘phase C’. Each phase had a significance that aligns with the contribution made by

Table 2: Classification results and implementation details of each of the models employed during ‘phase A’ of this research. The columns list the accuracy, training time and hyperparameters related to the implementation of each classifier, [1](sic).

Classifier	Accuracy	Training Time (s)	Hyperparameters
Linear Logistic Regression	81.00%	25.2500	maximum number of itterations for LogitBoost=500
Random Forests	75.70%	18.0800	number of trees = 1000
Support Vector Machines	75.40%	3.8200	kernel degree=3, tolerance=0.001, epsilon for loss function=0.1, used polynomial kernal: $\gamma^{u_0}v+coef_o$, and did not normalize
Multilayer Perceptron	75.20%	27.480	number of hidden layers= number of hidden classes, learning rate=0.3, training time=500 epochs, validation threshold=20
k-Nearest Neighbour	72.80%	0.0100	number of neighbours=1, using absolute error for crossvalidation, and appied linear search algorithm
naíve Bayes	53.20%	0.5600	used normal distribution for numeric attributes and supervised discretization

Table 3: Classification results and implementation details of each of the models employed during 'phase B' of this research. The columns list the accuracy, training time and hyperparameters related to the implementation of each classifier, [16](sic).

Classifier	Accuracy	Training Time (s)	Hyperparameters
k-Nearest Neighbours	92.69%	0.0780	nearest neighbours=1
Multilayer Perceptron	81.73%	60.620	activation=ReLu solver lbfgs
Random Forests	80.28%	52.890	number of trees=1000, max depth=10, α= e^{-5}, and hidden layer sizes=(5000,10)
Support Vector Machines	74.72%	3.8720	decision function shape=ovo
Logistic Regression	67.52%	3.6720	penaty=12, multi class=multinomial

Table 4: Classification results and implementation details of each of the models employed during 'phase C' of this research. The columns list the accuracy, training time and hyperparameters related to the implementation of each classifier, [9](sic).

Classifier	Accuracy	Training Time (s)	Hyperparameters
Support Vector Machines	80.80%	0.3000	radial basis function kernel, tolerance=0.001, and regularization=0.17
Multilayer Perceptron	77.30%	0.2300	hidden layers=2, learning rate=0.02, activation=ReLu, max iterations=200, solver=adam, and tolerance=0.0001
Logistic Regression	75.80%	0.0800	solver=newton-cg and max itterations=500
Random Forests	72.40%	61.080	split function=gini, number of trees = 100, and max depth 100
k-Nearest Neighbour	69.70%	0.0110	k=7 with manhattan distance metric, weighting=distance
naíve Bayes	54.50%	0.0019	Gaussian naíve Bayes with smoothing

this research to the current body of work. We present music genre classification via machine-learning and deep- learning approaches, furthermore, this work provides a comparison of the accuracy of machine-learning models and deep-learning models in completing the classification task.

After training several classifiers, the k-Nearest Neighbours (kNN) provided the best accuracy at 92.69%, furthermore, the kNN had a relatively low training time of 78 milliseconds. The higher accuracy attained by kNN relative to related literature can be explained by the three- seconds duration feature set which provides more training data. Backed by these findings, We conclude that three- second duration input features can provide better accuracy than thirty-second duration input features.

Further noteworthy performances were provided by the Linear Logistic Regression and Support Vector Machines (SVM), attaining 81.00% and 80.80% respectively. The Convolutional Neural Network (CNN) implementations followed in this research provided relatively low accuracy, with the most accurate CNN implementation attaining 72.40%.

This work has shown that automatic music genre classification is possible, furthermore, traditional machine learning models tend to outperform deep-learning approaches.

3. CONCLUSION

The automated determination of music genre and linguistic categorization project represents a significant stride in the fusion of machine learning, audio signal processing, and natural language processing. The comprehensive analysis and classification of music based on both its acoustic features and accompanying textual content have resulted in a robust and versatile system.

In the domain of music genre determination, the employed machine learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have proven effective in capturing intricate patterns and nuances within audio signals. The utilization of spectrogram and Mel-frequency cepstral coefficients (MFCCs) as feature representations has facilitated a nuanced understanding of the temporal and spectral characteristics inherent in various musical genres. The model's ability to adapt to diverse musical styles underscores its potential applicability across a broad spectrum of the music landscape.

On the linguistic categorization front, the integration of natural language processing techniques has enabled the system to extract valuable insights from textual data associated with music, such as lyrics. Sentiment analysis, topic modeling, and word embeddings have collectively enhanced the understanding of the linguistic context, contributing to more accurate genre categorization. The ability to consider both audio and text components concurrently has provided a holistic perspective on music classification, incorporating not only the auditory experience but also the lyrical and semantic dimensions.

While the project has showcased promising results, there are challenges that warrant further exploration. Ambiguities in genre boundaries and the evolving nature of music styles pose ongoing challenges. Additionally, the interpretability of the model's decisions and addressing potential biases in training data require continuous scrutiny.

In summary, the automated determination of music genre and linguistic categorization project has laid a solid foundation for advancing our understanding of how machine learning can comprehensively analyze and categorize music. Its success opens doors to applications in music recommendation systems, content organization, and contributes to the broader exploration of the intersection between artificial intelligence and creative domains. Continued research and refinement will undoubtedly drive the evolution of such systems, bringing them closer to real-world deployment and enhancing the user experience in the ever-expanding world of digital music

4. REFERENCES

[1] Music genre classification and music recommendation by using deep learning[1] A. Elbir✉ and N. Aydin,2021

[2] Music Recommendation System Miao Jiang,Ziyi,Chen Zhao,2020

[3] Multimodal Deep Learning for Music Genre classification Sergio Oramas,Francesco Barbieri,Oriol Nieto and Xavier Serra.

[4] E.Music genre Classification:A Review of Deep-learning and Traditional machine-Learning Approaches Ndiatenda Ndou,Ritesh Ajoodha,Ashwini Jadhav

[5] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kégl, "Aggregate features and adaboost for music classification," Machine Learning, vol. 65,
pp. 473–484, 12 2006.

[6] K. Choi, G. Fazekas, and M. Sandler, "Explaining deep convolutional neural networks on music classification," 2016.

[7] F. Chollet et al., "Keras," https://github.com/fchollet/keras, 2015.

[8] I. Fujinaga, "Adaptive optical music recognition," Ph.D. dissertation, McGill University, CAN, 1997, aAINQ29937.

[9] D. S. Lau and R. Ajoodha, "Music genre classification: A comparative study between deep-learning and traditional machine learning approaches," in Sixth International Congress on Information and Communication Technology (6th ICICT). Springer, 2021, pp. 1–8.

[10] J. H. Lee and J. S. Downie, "Survey of music information needs, uses, and seeking behaviours: preliminary findings." in ISMIR, vol. 2004. Citeseer, 2004, p. 5th.

[11] A. Lerch, An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics. Wiley Online Library, 10 2012.

[12] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification," in Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, ser. SIGIR '03. New York, NY, USA: Association for Computing Machinery, 2003, p. 282–289. [Online]. Available: https://doi.org/10.1145/860435.860487

[13] T. Lidy, A. Rauber, A. Pertusa, and J. Iñesta, "Combining audio and symbolic descriptors for music classification from audio," 2007.

[14] C. McKay, R. Fiebrink, D. McEnnis, B. Li, and I. Fujinaga, "Ace: A framework for optimizing music classification," in ISMIR, 2005.

[15] C. McKay and I. Fujinaga, "Musical genre classification: Is it worth pursuing and how can it be improved?" in ISMIR, 2006.

[16] T. Nkambule and R. Ajoodha, "Classification of music by genre using probabilistic graphical models and deep learning models," in Sixth International Congress on Information and Communication Technology (6th ICICT). Springer, 2021, pp. 1–6.

[17] A. Olteanu. Gtzan dataset - music genre classification. [Online]. Available: https://www.kaggle.com/andradaolteanu/gtzan- dataset-musicgenre-classification

[18] F. Pedregosa, G. Varoquaux, A. Gramfort,
V. Michel, B. Thirion, O. Grisel, M. Blondel, A.

Müller, J. Nothman, G. Louppe, P. Prettenhofer,
R. Weiss, V. Dubourg, J. Vanderplas, A. Passos,
D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay, "Scikit-learn: Machine learning in python," 2018.

[19] B. L. Sturm, "Alexander lerch: An introduction to audio content analysis: Applications in signal processing and music informatics," Computer Music Journal, vol. 37, no. 4, pp. 90–91, 2013.

[20] B. L. Sturm, "On music genre classification via compressive sampling," in 2013 IEEE International Conference on Multimedia and Expo (ICME), 2013, pp. 1–6.

[21] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5,
pp. 293–302, 2002