Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Performance Analysis of Deep Learning and Machine Learning Methods for Music Genre Classification System

J Vigneshwar[1], Thirumahal R[2]

[1]PG Scholar, Department of Computer Science and Engineering, PSG College of Technology, Coimbatore.

[2]Assistant Professor, Department of Computer Science and Engineering, PSG College of Technology, Coimbatore.

E-mail: [1]vigneshwarjayabal@gmail.com, [2]trk.cse@psgtech.ac.in

Abstract

Classification plays a crucial role in numerous applications within the music industry, spanning from content management to personalized playlists and music recommendation systems. While previous research has explored various machine learning frameworks for this purpose, such as support vector machines (SVM), a comprehensive comparison analysis of convolutional neural networks (CNN) and k-nearest neighbor (KNN) remains unknown. This study aims to address this gap by analyzing and contrasting the performance of SVM, KNN, and CNN in music genre classification. Each algorithm was carefully trained using different music genres from a protected dataset, employing feature extraction methods to capture the appropriate qualities of audio signals. The models underwent extensive training with a limited number of samples, and their performance was evaluated using industry standards such as accuracy, precision, recall, and F1 scores. Experimental results included SVM, KNN, and CNN for music genre designation. This study contributes significantly to existing literature by providing a comparative analysis of these algorithms. The findings highlight the strengths and limitations of each approach, offering guidance for researchers and practitioners in choosing the most suitable approach for their specific needs. The insights gained from this research have

DOI: https://doi.org/10.36548/jscp.2024.2.001
Received: 12.03.2024, received in revised form: 13.04.2024, accepted: 24.04.2024, published: 06.05.2024

the potential to enhance music genre classification systems, ultimately improving the user experience across various music-related contexts.

Keywords: Music genre classification, SVM, KNN, CNN, Audio feature

1. Introduction

The complexity of navigating a vast array of digital music genres necessitates sophisticated tools for programming and analysis. To address this challenge, this research conducts a comparative analysis of three primary approaches: Support Vector Machines (SVM), k-nearest neighbors (KNN), and Convolutional Neural Networks (CNN). SVMs excel in discerning subtle interactions between audio quality and genre labels, leveraging optimal boundaries to maximize class separations through kernel functions. Meanwhile, KNN, with its simplicity and success, relies on proximity in feature space to classify data points, effectively distributing genres. CNNs revolutionize genre classification by automatically extracting hierarchical representations from raw audio signals. They are adept at distinguishing localized and ubiquitous patterns, thus enhancing music analysis and classification across genres.

The paper employs SVM, KNN, and CNN algorithms to evaluate their efficacy in music genre classification, aiming to assess performance metrics like accuracy, precision, recall, and F1-score. Through this analysis, valuable insights are gleaned to enhance automatic genre classification in music systems, ultimately improving recommendations and performance across various applications.

2. Literature Review

The paper titled "Comparison of SVM, KNN, and NB Classifier for Metadata-Based Genre Music Classification" [1] focuses on the comparison of three popular classification algorithms for the task of metadata-based genre music classification. The authors' goal is to determine which algorithm performs better in this specific context. The study uses metadata that refers to information about music tracks, such as artist, album, year, and length, to classify music into different genres. The authors probably collected a dataset containing music tracks with corresponding metadata for experimentation. The paper provides a comparison of the performance of three classifiers in terms of performance metrics. These metrics are often used to evaluate the effectiveness of classification algorithms. The paper also says that music recommendations are essential for music streaming services and that categorizing music genres is an important first step in the music genre recommendation process. According to this paper, the SVM classifier has the best classification performance with an accuracy of 80%, followed by KNN with 77.18% and NB with 76.08%.

The paper titled "Music Genre Classification Using Machine Learning Algorithms: A Comparison" [2] investigates machine learning algorithms for predicting song genres to identify superior models. Utilizing the Free Music Archive (FMA) dataset, the study constructs classification models and assesses their performance in terms of prediction accuracy, lacking recent literature reviews. Various machine learning and deep learning algorithms, including mel-spectrograms and audio characteristics, are employed to categorize songs into genres, with spectrogram-only models achieving the highest accuracy. The article offers a comprehensive comparative analysis of machine learning algorithms, valuable for music information retrieval and machine learning researchers. Notably, the research streamlines genre classification, potentially benefiting music recommendation systems. However, the study's limited exploration of model interpretability represents a significant drawback, hindering applications where transparency and interpretive insights are crucial.

The paper titled "Music Genre Classification: A Review of Deep Learning and Traditional Machine Learning Approaches" [3] conducts a comprehensive comparative analysis of deep learning and traditional machine learning models for automated music genre classification. Utilizing the GTZAN dataset with ten music genres, the study evaluates model performance using features extracted from three-second and thirty-second durations. Key techniques highlighted include Mel Frequency Cepstral Coefficients (MFCC) aggregation, with the Information Gain Ranking algorithm used to identify influential features. Three distinct sets of properties are examined, and a confusion matrix is presented for "Phase A" using linear logistic regression. Notably, the k-Nearest Neighbors (kNN) algorithm achieves the highest classification accuracy of 92.69% with three-second input features. The paper contributes to understanding feature importance and model performance, enhancing discourse on music genre classification systems. Strengths include exploring machine learning and deep learning methods for audio signal classification, potentially improving efficiency in music search and categorization. However, the absence of real-time implementation investigation represents a limitation, hindering practical applicability in dynamic scenarios where timely music categorization is crucial. Addressing this gap would enhance the usability and effectiveness of the proposed classification models.

The paper titled "A Convolutional Neural Network Approach for Music Genre Classification" [4] emphasizes the growing importance of rapid music genre classification in the era of streaming platforms, aiming to address the time-consuming nature of conventional manual classification methods. Leveraging deep learning models, specifically convolutional neural networks (CNNs), the authors propose an efficient music genre classification system using Mel-spectrograms derived from original audio files. Through a majority voting mechanism from 10 classifiers, their model achieves an average accuracy of 84% on the GTZAN dataset. Visual aids such as graphs and diagrams enhance the clarity of their approach, which includes reviewing related work on deep learning techniques for music genre classification. Notably, the study highlights the advantages of CNNs in processing time series data efficiently. However, limitations regarding the model's adaptability to diverse musical genres and audio data should be considered.

The paper titled "Classification of musical genres by neural networks" [5] conducts a thorough review of literature on neural network applications in music genre classification, addressing challenges such as variations in song libraries and machine learning techniques. Various approaches including convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are explored, with the authors contributing their own research on genre classification using spectrogram images as inputs to neural networks. They suggest RNNs are beneficial for capturing sequential context, particularly in music analysis. While briefly mentioning other neural network types like generative adversarial networks (GANs), the primary focus remains on genre classification. This survey aims to synthesize existing knowledge and offer insights into complexities and potential solutions in using neural networks for music genre classification.

The paper proposes future research directions, highlighting RNNs for improved sequential context capture, promising advancements in music classification tasks. However, its exclusive focus on neural networks neglects exploration of alternative methodologies, limiting a comprehensive understanding of music genre classification techniques.

The paper titled "Music genre classification using transfer learning" [6] introduces a novel approach to classify 11 Western music genres based on sound through transfer learning. The authors identify limitations in previous methods, predominantly employing hand-crafted features and traditional machine learning algorithms. They advocate for transfer learning as a solution to address data sparsity issues in music informatics research. By adapting neural networks trained on source tasks to target tasks, transfer learning enhances classification performance. Drawing from a study by Liang et al., which utilized features from pre-trained convolutional neural networks (CNNs), the authors demonstrate the effectiveness of transfer learning in the target task of pedal on/off classification from acoustic piano recordings. The paper underscores transfer learning's potential to advance music genre classification by improving both accuracy and scalability. Notably, it establishes a benchmark for future research and contributes significantly to the field of music informatics. However, reliance on pre-trained neural networks and the absence of comparative analyses with other state-of-the-art methods are notable limitations, although the increasing availability of pre-trained models in music informatics may mitigate this issue over time.

The paper titled "Neural Network Model for Recommending Music Based on Music Genres" [7] begins with an extensive literature survey aimed at defining the problem and identifying gaps in existing research, emphasizing the necessity of effective music recommendation systems in the digital music consumption era. Various methodologies used in previous music recommendation research, such as collaborative filtering, content-based filtering, and hybrid approaches, are discussed, with the authors highlighting the potential for improvement in accuracy and efficiency. The article focuses on the use of neural networks, particularly convolutional neural networks (CNN) and recurrent neural networks (RNN), in music recommendation, citing their ability to capture temporal patterns and adaptability to different data types. The literature review lays the foundation for the authors' proposed approach, which utilizes a neural network model based on music genre to enhance musical recommendation accuracy.

The advantages of the paper include the proposed model's ability to extract musical features through neural networks, thereby improving recommendation accuracy, and the comprehensive literature survey offering valuable insights for researchers in the field.

However, a notable limitation is the lack of discussion on the scalability of the proposed model, which may hinder its practical application in real-world scenarios where scalability is crucial..

The paper titled "Innovative Model for Automatic Music Genre Classification Using Deep Convolutional Neural Network and Self Adaptive Sea Lion Optimization" [8] presents a unique approach to automatic music genre classification (MGC), integrating feature extraction and classification stages. Various features, including non-negative matrix factorization (NMF), short-time Fourier transform (STFT), and pitch features, are extracted and classified using a deep convolutional neural network (DCNN) model. The DCNN model undergoes training with a novel SA-SLnO (Self Adaptive Sea Lion Optimization) weight optimization model to enhance classification accuracy. Evaluation of the proposed methodology against existing approaches includes error analysis and statistical measures, while the literature review covers advancements in MGC, such as attentional mechanisms, centroid optimization, and representation learning. The paper's strengths include the utilization of a DCNN model trained through the SA-SLnO model for optimized accuracy and a comprehensive literature survey showcasing current advancements. However, potential limitations may arise from the model's generalizability across different musical genres and datasets, as well as challenges in implementation due to computational resource requirements and expertise in deep learning techniques, possibly hindering adoption by researchers and practitioners with limited resources or familiarity with advanced methodologies.

The paper titled "Comparative analysis of three improved deep learning architectures for music genre classification" [9] offers a thorough literature review on the application of deep learning models in music information retrieval (MIR), particularly focusing on music genre classification. It critiques conventional hand-crafted feature extraction methods in MIR tasks, highlighting their limitations and explores the rising popularity of deep learning models due to their capacity to learn from examples and tackle complex perception tasks with high accuracy. The discussion covers three main deep learning models for music genre classification: Convolutional Neural Network (CNN), recurrent neural network (RNN), and hybrid convolutional recurrent neural network (CRNN), along with extended architectures like Bottom-up Broadcast Neural Network (BBNN), Independent Recurrent Neural Network (IndRNN), and CRNN in time and frequency dimensions (CRNN-TF), showcasing their state-of-the-art performance. The paper also addresses challenges faced by deep learning models in

MIR tasks and proposes solutions, including the adoption of new CNN architectures like BBNN and the integration of IndRNNs to enhance long-term dependency learning. While the paper offers innovative solutions, it lacks comprehensive technical explanations of deep learning models, assuming prior reader knowledge, and falls short in providing extensive experimental evaluations, limiting the generalizability and robustness of the proposed approaches across different datasets and scenarios.

The paper titled "Neural Network-Based Music Genre Classification" [10] delves into the development of an automated system utilizing neural networks to classify music genres, leveraging Mel Frequency Cepstral Coefficients (MFCC) as fundamental sound properties. Through a convolutional neural network (CNN) fed with short-time Fourier transform inputs, followed by another deep neural network for genre classification, the study utilizes the GTZAN dataset for training and testing. However, the implemented CNN exhibits signs of overfitting, with notably high accuracy on the training dataset compared to the test dataset. The paper also underscores the use of Python programming language alongside the Python Data Analysis Library (PANDAS) for efficient data manipulation and analysis. While the research presents a significant contribution to music genre classification through neural networks, it lacks a detailed analysis of study limitations, which could enhance its credibility, and misses insights into potential future research directions, limiting its guidance for further exploration in the field.

3. Proposed Work

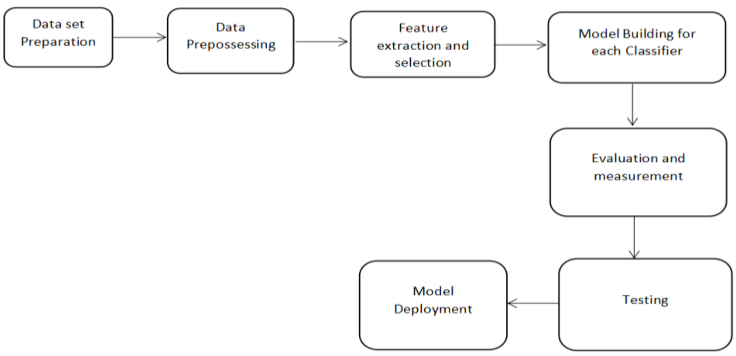

Figure 1. System Design

The system's design process is visually represented in Figure 1 through a flowchart, detailing the step-by-step progression from raw data processing to model training and deployment. This illustration serves as a valuable aid in understanding the overall design and workflow of the system, enabling efficient data processing and the development of models for classifying musical genres.

3.1 Dataset Description

A diverse compilation comprising ten distinctive genres is presented, encompassing 100 audio files per genre, each lasting 30 seconds. Concurrently, an accompanying set of original images captures the Mel-spectrogram visualization for every audio file, intended for input into a CNN model. Additionally, two CSV files have been meticulously crafted, encapsulating audio file features. The first CSV file provides mean and variance values computed across various features for each complete 30-second song. Meanwhile, the second CSV file follows a similar structure, but the songs have been pre-segmented into 3-second audio files. This comprehensive collection serves as a rich resource for further exploration and analysis within the realm of audio data processing.

The evaluation process for music genre classification using SVM, KNN, and CNN algorithms typically follows a structured approach. It begins with data preparation, where the GTZAN dataset is pre-processed to augment sample size by subdividing original 30-second audio clips into 3-second segments and generating Mel-spectrograms for CNN input. Essential features are extracted from these segments for SVM and KNN input. Next, the SVM, KNN, and CNN models are trained using the prepared data to discern patterns and relationships. Following training, the models are evaluated using testing data, comparing predictions with actual labels to assess performance. Performance metrics including confusion matrix, F1 score, accuracy, precision, and recall are computed to provide a comprehensive evaluation of the models' capability in precisely classifying music genres.

4. Experimental Result

The way the SVM, CNN, and CNN perform on the test data will determine the project's outcome in the end. The coding was done using python using the libraries scikit-learn and Matplotlib and executed in colab.

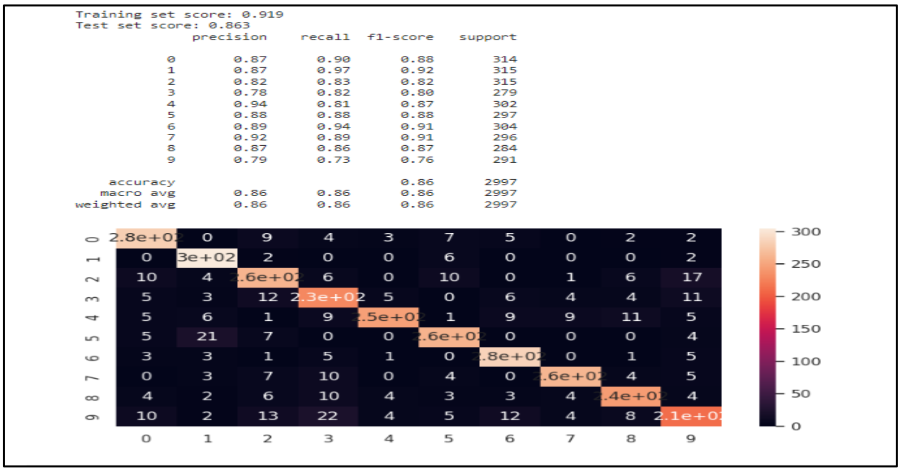

Figure 2. Output of Performance of SVM Model

Figure 2 displays the output screenshot of the Support Vector Machine (SVM) algorithm code applied to a 3-second audio clip from the GTZAN dataset. The output encompasses a detailed analysis of the SVM model's performance, inclusive of a confusion matrix and an array of evaluation criteria. The assessment metrics provide a thorough examination of the model's ability to classify by offering valuable insights into its accuracy, precision, recall, and F1-score. The confusion matrix visually represents the proportions of accurate and inaccurate classifications for each musical genre, offering a graphical depiction of the model's predictions.

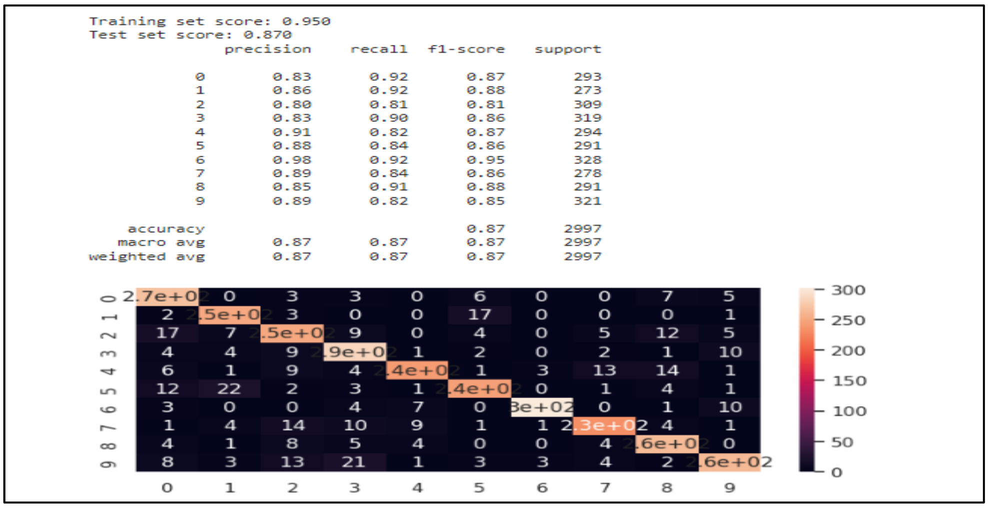

Figure 3. Output of Performance of KNN Model

Figure 3 presents the outcome screenshot of the K-nearest neighbors (KNN) algorithm code execution for a 3-second audio clip sourced from the GTZAN dataset. This visual representation includes a confusion matrix alongside additional assessment parameters. The confusion matrix visually represents the classification results, while the evaluation metrics offer detailed insights into the performance of the KNN algorithm. Collectively, these elements provide a comprehensive overview of the algorithm's effectiveness in accurately categorizing audio clips from the GTZAN dataset.

```
[ ] model.fit_generator(train_generator,epochs=10,validation_data=vali_generator)

<ipython-input-29-af1eb8ee7fa0>:1: UserWarning: `Model.fit_generator` is deprecated and will be removed in a future version. Please use `Model.fit`, which supports generators.
  model.fit_generator(train_generator,epochs=10,validation_data=vali_generator)
Epoch 1/10
71/71 [==============================] - 1224s 17s/step - loss: 1.4883 - accuracy: 0.5953 - get_f1: 0.5671 - val_loss: 2.5760 - val_accuracy: 0.1000 - val_get_f1: 0.0000e+00
Epoch 2/10
71/71 [==============================] - 1197s 17s/step - loss: 0.0989 - accuracy: 0.9733 - get_f1: 0.9726 - val_loss: 3.0735 - val_accuracy: 0.1000 - val_get_f1: 0.0000e+00
Epoch 3/10
71/71 [==============================] - 1227s 17s/step - loss: 0.0340 - accuracy: 0.9914 - get_f1: 0.9918 - val_loss: 2.9701 - val_accuracy: 0.1020 - val_get_f1: 0.1727
Epoch 4/10
71/71 [==============================] - 1274s 18s/step - loss: 0.0137 - accuracy: 0.9979 - get_f1: 0.9977 - val_loss: 2.2305 - val_accuracy: 0.2940 - val_get_f1: 0.3679
Epoch 5/10
71/71 [==============================] - 1269s 18s/step - loss: 0.0101 - accuracy: 0.9974 - get_f1: 0.9974 - val_loss: 0.7485 - val_accuracy: 0.7280 - val_get_f1: 0.7317
Epoch 6/10
71/71 [==============================] - 1225s 17s/step - loss: 0.0061 - accuracy: 0.9992 - get_f1: 0.9991 - val_loss: 0.1155 - val_accuracy: 0.9810 - val_get_f1: 0.9778
Epoch 7/10
71/71 [==============================] - 1267s 18s/step - loss: 0.0036 - accuracy: 0.9994 - get_f1: 0.9994 - val_loss: 0.0165 - val_accuracy: 0.9980 - val_get_f1: 0.9980
Epoch 8/10
71/71 [==============================] - 1264s 18s/step - loss: 0.0024 - accuracy: 0.9998 - get_f1: 0.9997 - val_loss: 0.0127 - val_accuracy: 0.9960 - val_get_f1: 0.9966
Epoch 9/10
71/71 [==============================] - 1233s 17s/step - loss: 0.0020 - accuracy: 0.9998 - get_f1: 0.9999 - val_loss: 0.0086 - val_accuracy: 0.9970 - val_get_f1: 0.9980
Epoch 10/10
71/71 [==============================] - 1226s 17s/step - loss: 0.0016 - accuracy: 0.9999 - get_f1: 0.9999 - val_loss: 0.0035 - val_accuracy: 0.9990 - val_get_f1: 0.9995
<keras.src.callbacks.History at 0x7ea4e6e598a0>
```

Figure 4. Output of Performance of CNN Model

Figure 4 visually represents the performance of a Convolutional Neural Network (CNN) model in classifying musical genres using the GTZAN dataset. The image provides a thorough evaluation of the model's accuracy in correctly categorizing musical genres, employing various assessment criteria such as accuracy, precision, recall, and F1-score. Through this comprehensive analysis, the image effectively illustrates the CNN model's capability to accurately identify and classify musical genres, showcasing its potential in the development of music analysis and recommendation systems. The Table.1 shows the comparison of performance scores of the machine learning and the deep learning algorithms.

Table 1. Comparison of Performance Scores

Models	Training Score	Test Score
SVM	0.919	0.863
KNN	0.950	0.870
CNN	0.999	0.990

5. Conclusion and Future Works

The music genre classification project utilized the GTZAN dataset and employed three established machine learning and deep learning algorithms, showcasing commendable performances: SVM achieved 86% accuracy, KNN reached 87%, and CNN delivered an impressive 99% accuracy rate. While all algorithms demonstrated effectiveness, CNN exhibited superior performance, showcasing deep learning's capability in extracting essential features autonomously from audio data. Its exceptional accuracy underscores its potential in content labeling and recommendation systems within the music industry. This research emphasizes the significance of deep learning, particularly CNNs, in musical genre classification, with the CNN algorithm's 99% accuracy rate illustrating its proficiency in extracting intricate features without human feature engineering. The findings contribute to advancing deep learning techniques in music analysis, enriching the understanding of music genre classification and related domains.

References

[1] Rahardwika, Dewangga Satriya, Eko Hari Rachmawanto, Christy Atika Sari, Candra Irawan, Desi Purwanti Kusumaningrum, and Swapaka Listya Trusthi. "Comparison of SVM, KNN, and NB classifier for genre music classification based on metadata." In 2020 international seminar on application for technology of information and communication (iSemantic), pp. 12-16. IEEE, 2020.

[2] Ghildiyal, Anirudh, Komal Singh, and Sachin Sharma. "Music genre classification using machine learning." In 2020 4th international conference on electronics, communication and aerospace technology (ICECA), pp. 1368-1372. IEEE, 2020.

[3] Ndou, Ndiatenda, Ritesh Ajoodha, and Ashwini Jadhav. "Music genre classification: A review of deep-learning and traditional machine-learning approaches." In 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1-6. IEEE, 2021.

[4] Cheng, Yu-Huei, Pang-Ching Chang, and Che-Nan Kuo. "Convolutional neural networks approach for music genre classification." In 2020 International Symposium on Computer, Consumer and Control (IS3C), pp. 399-403. IEEE, 2020.

[5] Pelchat, N. and Gelowitz, C.M.,"Neural network music genre classification". Canadian Journal of Electrical and Computer Engineering, 43(3), pp.170-173.,2020.

[6] Liang, B. and Gu, M.,"Music genre classification using transfer learning." In 2020 IEEE conference on multimedia information processing and retrieval (MIPR) pp. 392-393. IEEE.,2020, August.

[7] Singh, J. and Bohat, V.K.,"Neural Network Model for Recommending Music Based on Music Genres." In 2021 International Conference on Computer Communication and Informatics (ICCCI) pp. 1-6. IEEE.,2020, August.

[8] Kumaraswamy, B. and Poonacha, P.G., "Deep convolutional neural network for musical genre classification via new self adaptive sea lion optimization." Applied Soft Computing, 108, p.107446.,2021

[9] Rafi, Q.G., Noman, M., Prodhan, S.Z., Alam, S. and Nandi, D., "Comparative analysis of three improved deep learning architectures for music genre classification." International Journal of Information Technology and Computer Science, 13(2), pp.1-14.,2021.

[10] Raval, M., Dave, P. and Dattani, R., "MUSIC GENRE CLASSIFICATION USING NEURAL NETWORKS." International Journal of Advanced Research in Computer Science, 12(5).,2021.pp 12-18