Transactions on
Federated Engineering & Systems
VOL. 1, ISSUE. 1, AUGUST 2023

RESEARCH ARTICLE Open Access

Music Classification Using Convolutional Neural Systems

Arpitha K M . Nimrita Koul

School of Computer Science and Engineering
REVA University, Bangalore, Karnataka, India.

Received: 18 September 2023 / Revised: 19 November 2023 / Accepted: 5 December 2023
©Milestone Research Publications, Part of CLOCKSS archiving
DOI: **10.5281/zenodo.10279702**

Abstract – In this work we used two CNN models, to determine the genre of a particular musical composition, Lenet-5 and CNN-64 are trained on a dataset of audio samples. We evaluated the models based on accuracy and loss. We found that Lenet-5 achieved higher accuracy and lower loss than CNN-64, indicating its effectiveness. The outcome highlights potential for CNNs in retrieval of music data also demonstrate the utility of Lenet-5 for achieving high accuracy and low loss in genre classification. Thus we see that CNNs can be used in music-related applications, such as music recommendation systems and music transcription.

Index Terms – Music Genre, Convolutional Neural Networks (CNN), Lenet-5, CNN-64

I. INTRODUCTION

Classifying genres it's a critical task in data retrieving about music that involves automatic categorization of music into different genres. The process of music genre classification plays a crucial role in organizing large music collections and providing relevant music recommendations to users. The subjective character of music and the diversity of musical styles within each genre make it a difficult seek to categorize music into distinct genres. However, accurately categorizing music into different genres has numerous practical applications, including personalized music recommendations, automatic playlist generation, and music tagging for indexing and retrieval. In this work, we train and assess two CNN models, Lenet-5 and CNN-64, on a dataset of audio samples to explore the potential of CNNs for music genre classification. We outline the training and testing procedures for the models, including how accuracy and loss are used as performance indicators. This study's main objective is to compare the performance of various CNN models to see which one is best for classifying music genres. Lenet-5 scored better in the study's music genre classification task than CNN-64 in terms of accuracy and loss, proving its efficacy. The study's findings demonstrate CNNs potential for information retrieval in the field of music and lay the groundwork for further investigation in this area.

II. LITERATURE SURVEY

The performance of these algorithms is measured using accuracy scores and listening tests. Dr. P. Devaki [1]. The proposed model for classifying music genres has shown results that are superior to those of the existing waveform-based classification techniques and comparable to those of spectrum analysis techniques using KNN, SVC, neural networks, and CNN models. When the CNN model was tested with spectrogram representations, it achieved training accuracy of 98% but testing accuracy of just 57%. Yu-Huei Cheng ,Pang-Ching Chang [2]. CNN model is then trained using the converted Mel spectrum. This study the GTZAN datasets splitting to train and test , converting each audio file to its corresponding MFCC before sending it to model CNN . Librosa utility to processing audio signals, and it is used to convert audio during preprocessing in order to produce spectrograms Vishnupriya S,K.Meenakshi [3].

CNNs are employed in this instance for training and classification. The important part of feature extraction for audio analysis has been done. Technique uses feature vector extraction to categorize music into different genres. Results indicate that the proposed method, which aimed to enhance and ease the automatic classification of musical genres, has an accuracy level of about 76%. Ndiatenda Ndou, Ritesh Ajoodha [4]. The k-Nearest Neighbours (KNN) produced the greatest accuracy at 92.69% after training a number of classifiers, and it also had a short training time of 78 milliseconds. The three-second length feature set, which offers more training data, accounts for the better accuracy reached by kNN in comparison to comparable work. Decide that input features with a length of three seconds or less can offer more accuracy than those with a duration of thirty seconds. Anirudh Ghildiyal , Komal[5]. The proposed research project has examined a number of classification models and created a new model for CNN that is superior to earlier suggested methods. In this study, the suggested models were trained and compared using the GTZAN dataset. The majority of the models were learned using audio file trains, while a small number of models were trained using spectrograms.

III. PROPOSED METHODOLOGY

Datasets

The dataset we used contains 1000 audio files of 30 seconds each, divided into 10 different genres, with 100 audio files per genre. The sampling size for wav-formatted audio file is 22050 Hz.

S. No	Class	Clips
1	Blues	100
2	Classical	100
3	Country	100
4	Disco	100
5	Hiphop	100
6	Jazz	100
7	Metal	100
8	Pop	100
9	Reggae	100
10	Rock	100
	Total	1000

Fig. 1: Considered datasets

CNN Architecture

A. LENET-5

This architecture has 7 layers, the first convolutional section detects basic features in the input image, such as edges and corners. The second convolutional layer combines these features to detect more complex features. The output's size is shrink by the pooling layer, makes the model more computationally efficient. Finally, the fully connected layers classify the input image into one of the ten-digit classes. While LeNet-5 was initially developed for handwritten digit recognition, its architecture has been adapted and used for various computer vision applications, including facial recognition, object detection, and image segmentation.

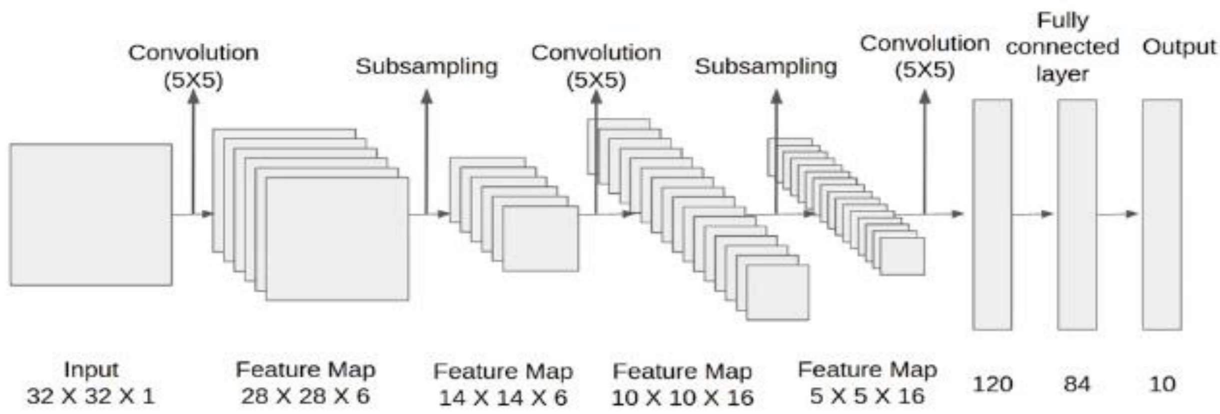

Fig. 2: Lener-5 Proposed Architecture

B. CNN-64

CNN-64 model can recognise patterns and elements in the photos that are typical of image classes throughout training, for as identifying various musical genres based on album covers. The CNN-64 architecture starts with several layered convolutions, which are followed with layers of pooling to shrink the feature maps' spatial dimensions. Following these layers are fully linked layers that provide the network's final output, which is a probability distribution across the various picture classifications. CNN-64 has demonstrated impressive results in a variety of image classification applications, identifying picture classes with a high degree of accuracy. However, the architecture and hyper parameters of the model may need to be adjusted depending on the specific task and dataset.

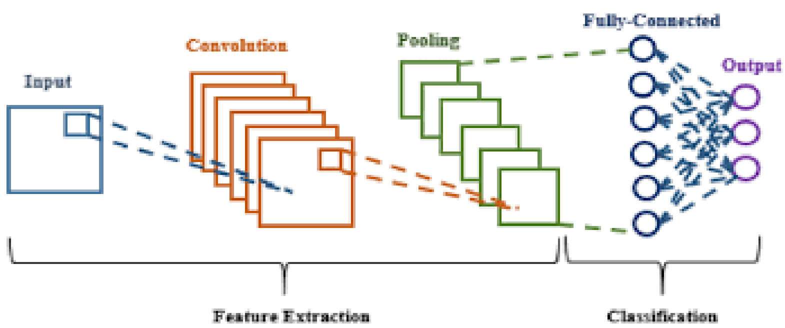

Fig. 3: CNN-64 Architecture

Proposed System

In the proposed system we use the LeNet 5 and CNN-64 architectures to extract features from audio files in the time-frequency domain. The Mel-frequency computes audio files, which will serve as input to the CNN-based classifier. The system is trained on a large dataset of music files spanning different genres and styles. We compared with other existing music genre classification systems to assess its effectiveness. The proposed system has many applications in the field of music. It can be used to automatically organize music collections based on genre, facilitate music search based on genre, and recommend music to users based on their listening preferences. It can also be useful in music streaming services that rely on automatic classification of music genres to enhance user experience.

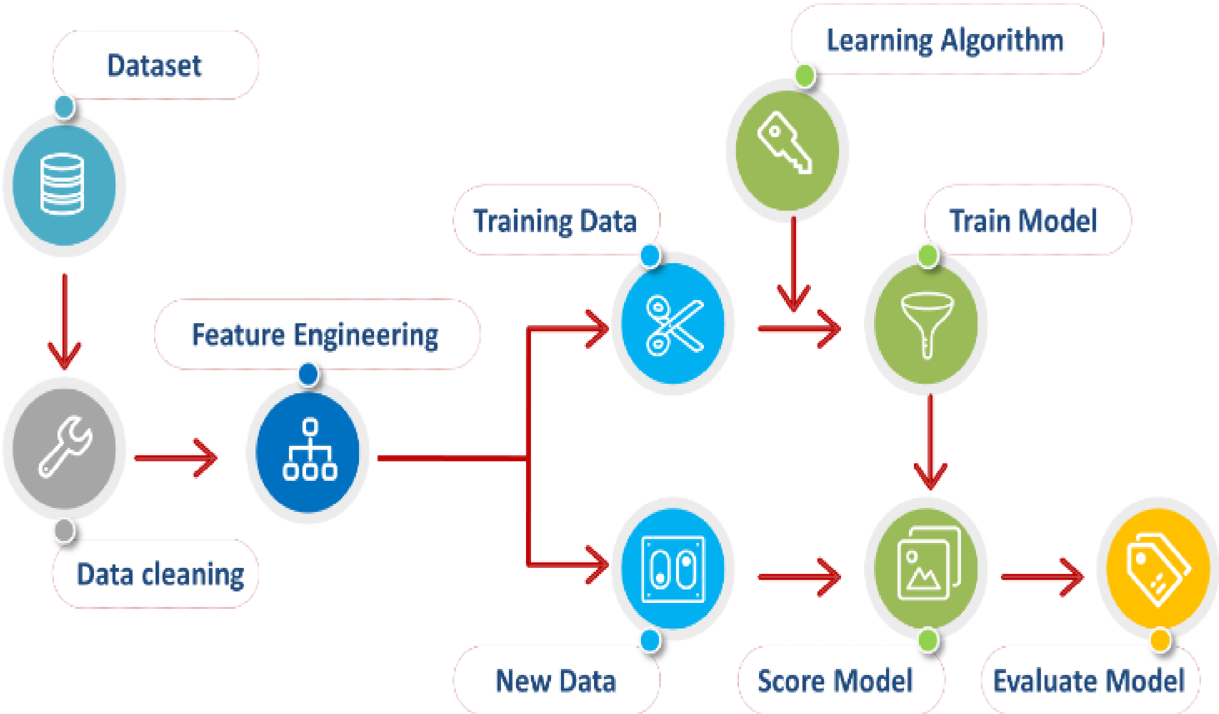

Fig.4: Proposed System Architecture

VI. RESULTS

LeNet-5– Classification Report

The confusion matrix for the LeNet-5 model showed an accuracy of 82.59%. This means that the model correctly predicted the class of 82.59% of the test samples. The loss value of 0.6758 indicates how well the model is able to minimize its prediction errors during training. A lower loss value indicates better performance.

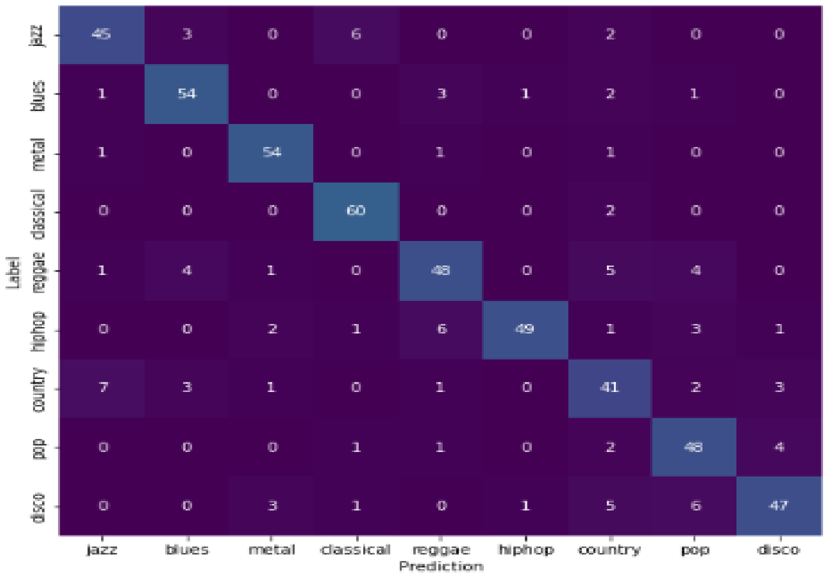

Fig. 5: Confusion report for LeNet-5

CNN-64 Confusion Matrix

The accuracy of the CNN-64 model is 82.41%, which means that the model correctly predicted the class of 82.41% of the samples in the dataset. The loss of the CNN-64 model is 1.0028.

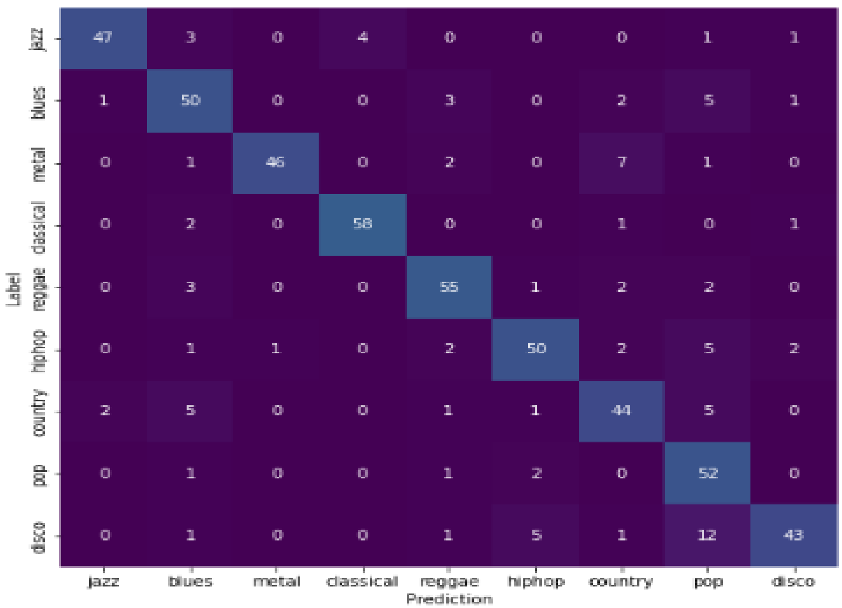

Fig. 6: Confusion matrix for CNN-64

ROC Graph

TPR set to 1 and an FPR to 0, signifying that it classifies all positive samples correctly and none of the negative samples as positive. The ROC graph offers insight into the model's performance across all potential thresholds. Figures 6 and 7 show the ROC curves of our lenet and CNN64 architectures.

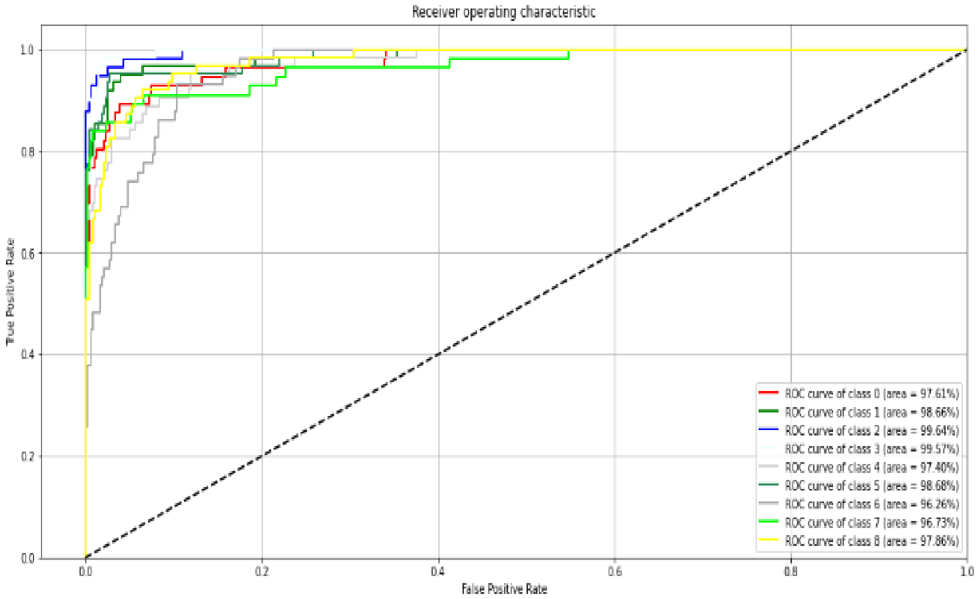

Fig 7: ROC Curve of lenet

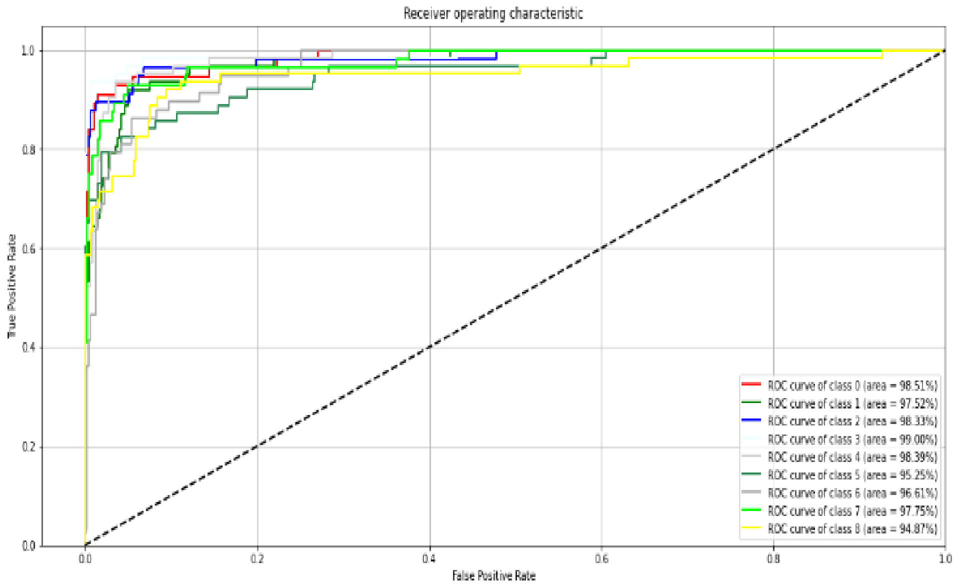

Fig 8: ROC Curve of CNN

VII. CONCLUSION

In this paper we used LeNet-5 and CNN 64 for prediction of music genres. LeNet-5 achieved an accuracy of 82.59%, while CNN-64 had an accuracy of 82.41%. This indicates that these models have the potention to be used in the task of music genre classification.

REFERENCES

1. Devaki, P., Sivanandan, A., Kumar, R. S., & Peer, M. Z. (2021, October). Music Genre Classification and Isolation. In *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)* (pp. 1-6). IEEE.

2. Srivastava, N., Ruhil, S., & Kaushal, G. (2022, November). Music Genre Classification using Convolutional Recurrent Neural Networks. In *2022 IEEE 6th Conference on Information and Communication Technology (CICT)* (pp. 1-5). IEEE.
3. Falola, P. B., & Akinola, S. O. (2021). Music Genre Classification Using 1D Convolution Neural Network. *International Journal of Human Computing Studies*, *3*(6), 3-21.
4. Ndou, N., Ajoodha, R., & Jadhav, A. (2021, April). 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.
5. Ghildiyal, A., Singh, K., & Sharma, S. (2020, November). Music genre classification using machine learning. In *2020 4th international conference on electronics, communication and aerospace technology (ICECA)* (pp. 1368-1372). IEEE.
6. Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing*, *22*(10), 1533-1545.
7. Gemmeke, J. F., Ellis, D. P., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., ... & Ritter, M. (2017, March). Audio set: An ontology and human-labeled dataset for audio events. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 776-780). IEEE.
8. Bhavatarini, N., Basha, S. M., & Ahmed, S. T. (2022). *Deep Learning: Practical approach*. MileStone Research Publications.
9. Chen, C., & Steven, X. (2021, March). Combined transfer and active learning for high accuracy music genre classification method. In *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)* (pp. 53-56). IEEE.
10. Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, *10*(5), 293-302.
11. Lidy, T., & Schindler, A. (2016). Parallel convolutional neural networks for music genre and mood classification. *MIREX2016*, *3*.