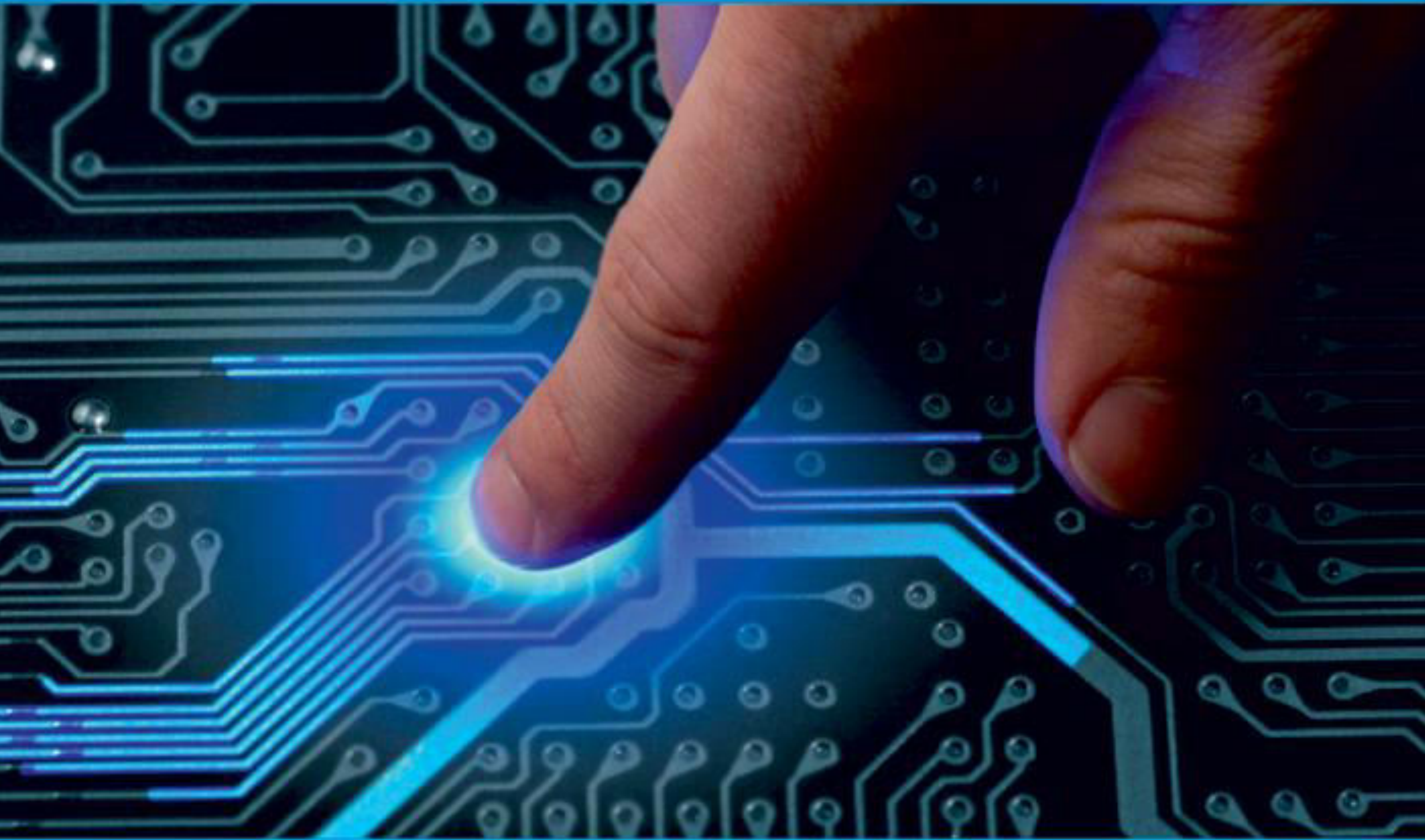
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 2, February 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202007 |

Classification of Music Genre using Machine Learning

Mrs. Preethi Kolluru Ramanaiah[1], **Dr. Ramanaiah**[2]

Cloud Architect, Ernst and Young, New York, USA[1]

Head of Department of Mathematics (Retd), Jamal Mohammed College, Trichy, India[2]

ABSTRACT: This study investigates the use of convolutional neural networks (CNN), specifically CNN-64 and Lenet-5, for the classification of musical genres. The process entails using measures like accuracy and loss to assess and train the models.Lenet-5 performs better than CNN-64, demonstrating its effectiveness in categorizing genres. The research highlights the potential of CNNs in obtaining musical data, highlighting the higher precision and reduced loss of Lenet-5. The results provide important new understandings of CNNs for music genre classification and establish a framework for further investigation. It is suggested that more research be done on the potential of CNNs for wider music applications, such as recommendation systems and transcribing.

KEYWORDS: ***Music*** *Genre, ConvolutionNeural Networks (CNN), Lenet-5, CNN-64.*

I. INTRODUCTION

Sorting music into different types, like pop, rock, or hip-hop, is an important part of organizing music information. It means using technology to automatically group songs into various categories based on their style.This process is crucial for efficiently organizing vast music collections and providing consumers with tailored song recommendations. The challenge lies in the subjective nature of music and the wide range of styles within each genre. Despite these challenges, accurately classifying music into genres has practical applications such as personalized recommendations, automatic playlist generation, and music tagging for indexing.

With the exponential growth of digital music collections, the need for swift and precise genre classification has become more pressing. Ongoing research aims to develop efficient methods for this task. In this particular study, we delve into the potential of Convolutional Neural Network (CNN) models, specifically Lenet-5 and CNN-64, trained on a dataset of audio samples for music genre classification. The study details the training and testing procedures, with accuracy and loss serving as performance indicators.

This study aims to see how well different computer models can tell music genres apart. The goal is to figure out which model works best for classifying genres.Lenet-5 did better than CNN-64 in terms of accuracy and loss, showing that it works well for classifying music genres. This suggests that CNNs have a lot of promise in finding and organizing music, and there's potential for more research in this area.

II. RELATED WORK

People have been studying how to classify music genres for a long time, especially since the early days of the internet. Tzanetakis et al. [1] tackled this by using computer techniques to learn from examples. They used things like Gaussian Mixture models and k-nearest neighbor classifiers. They also came up with three different types of features to help with this: pitch content, rhythmic content, and timbral structure.The exploration extended to support vector machines utilizing diverse distance metrics, as well as hidden Markov models (HMMs) for the purpose of classifying musical genres [2] [3] [4].

In their investigation, Lidy and Rauber [5] emphasized the significance of psychoacoustic elements, particularly the Short-Time Fourier Transform (STFT) on the Bark Scales [6], for discerning musical genres. Tzanetakis et al. utilized properties such as MFCCs, spectrum contrast, and spectral roll-off [7]. Nanni et al. [8] used SVM and AdaBoost classifiers and taught them using a mix of visual and auditory information.

More recent approaches have involved the use of neural networks with deep layers for music genre classification [9][10]. The fast rate at which audio is recorded makes it tricky to process it for neural networks when considering time.Van Den Oord et al. [11] addressed this issue in the context of audio creation tasks. Another prevalent representation is the spectrogram, which is conducive to training convolutional neural networks (CNNs) capable of capturing both temporal and frequency information [12].

Li et al. [13] created a computer model (CNN) that guesses music genres. They used a raw MFCC matrix as the starting information. In study by Li et al. [14] they used a constant Q-transform (CQT) spectrogram as the initial data for the CNN in the same task of predicting music genres.

III. DATASET

The collection of data includes 1000 sound files. Each file is 30 seconds long and falls into one of 10 different types of music, such as Rock, Pop, Reggae, Hip-Hop, Metal, Country, Disco, Jazz, and Blues. With 100 files per genre, the dataset provides a balanced representation. All audio files maintain a sample rate of 22050 Hertz and are formatted in .wav. A crucial component is the accompanying metadata file, containing detailed information about each audio file, including genre, artist, and title. This metadata enhances the dataset's utility by providing additional context for the audio content. This dataset is useful for studying music genres because it has standardized features and includes a variety of music types. It's a good source for teaching and testing computer models, providing a wide range of characteristics that can help improve how we classify and find music.

IV. METHODOLOGY

1.1 Lenet-5

The LeNet-5 model as shown in the Figure 1 was created in 1998 by the Convolutional Neural Network (CNN) pioneers Yann LeCun, Leon Bottou and Patrick Heffner. One of the early CNN models that did really well in accurately recognizing handwritten digits on the MNIST dataset, this one primarily designed to recognize handwritten characters in photographs.

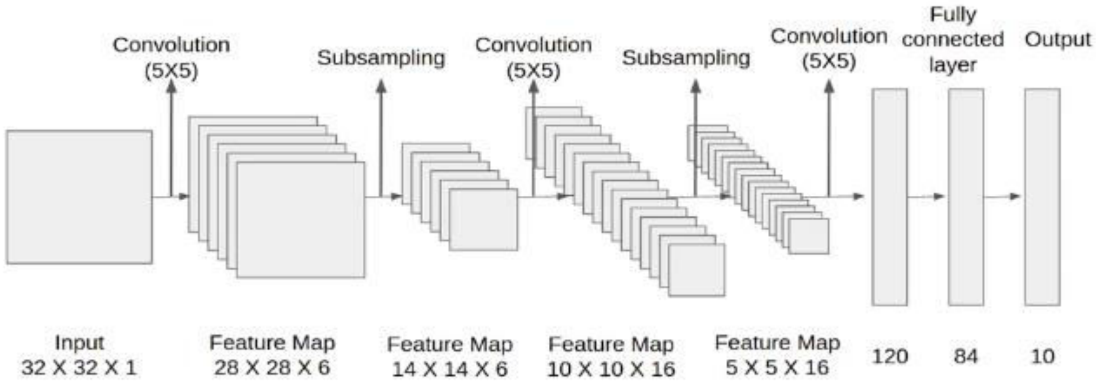


Fig 1: Lener-5 Architecture

LeNet-5's architecture consists of seven layers, including two convolutional layers, two pooling layers, and three fully linked layers. The first convolutional layer recognizes fundamental aspects in the input image, such as edges and corners, and the second layer combines these information to recognize more sophisticated patterns. Pooling layers reduce output spatial size, enhancing computational efficiency. The fully connected layers classify input images into ten digit groupings. LeNet-5's architectural design has become a cornerstone for subsequent CNN models, demonstrating the efficacy of convolutional neural networks in image recognition tasks. Originally developed for handwritten digit recognition, its adaptability extends to various computer vision applications, including facial recognition, object detection, and image segmentation. The versatility and foundational role of LeNet-5 underscore its significance in advancing the field of computer vision.

1.2 CNN-64

A Convolution Neural Network (CNN) architecture called CNN-64 as shown in the Figure 2 is employed for image categorization applications. The network's last layer, which is in charge of removing features from the input image, has 64 filters or channels, as indicated by the number "64" in the name.

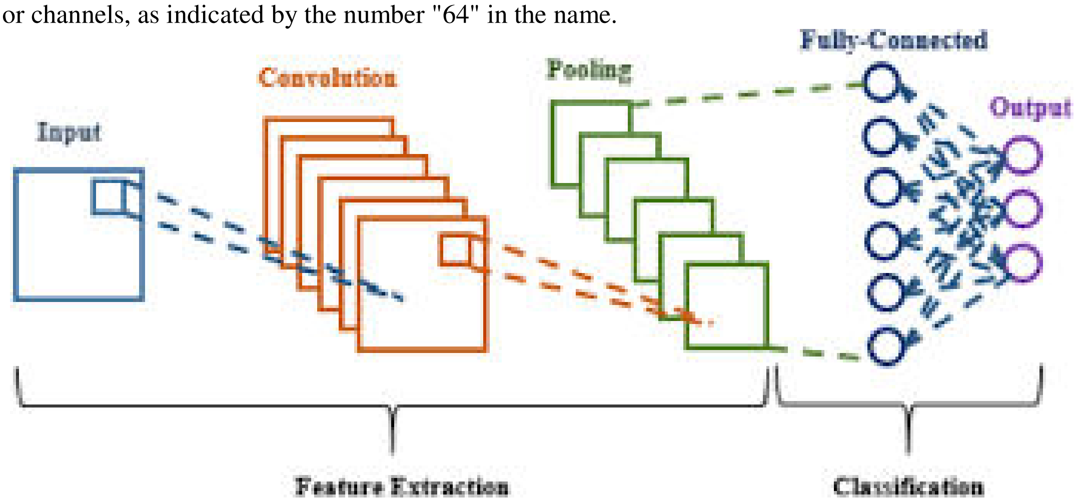


Fig 2: CNN-64 Architecture

A large dataset of tagged picture data is often used to train the CNN-64 deep learning model. The model gains the ability to recognize patterns & features in the photos that are typical of particular image classes throughout training, for as identifying various musical genres based on album covers. CNN-64's architecture consists of many convolutional layers, which are followed by pooling layers that minimize the spatial dimensionality of feature maps. These layers are followed by fully connected layers that provide the network's final output, which is a probability distribution over the various image classifications. CNN-64 has been used in various image classification tasks and has shown promising results, achieving high accuracy in recognizing image classes. Depending on the job and dataset, the model's architecture and hyperparameters may need to be changed.

V. Proposed System

As in Figure 3 the proposed technique is to construct a machine learning model specifically intended for exact music genre classification using Convolutional Neural Networks (CNN). Music categorization plays a vital role in Music Information Retrieval (MIR) for effectively organizing and recommending music to consumers.

To extract time-frequency information from audio recordings, the system will leverage the LeNet 5 and CNN-64 architectures. The CNN-based classifier will take input in the form of Mel-frequency cepstral coefficients (MFCCs) derived from the audio data. Training of the system will be conducted using a substantial dataset comprising audio files that represent diverse musical genres and styles.

The suggested system will be evaluated using a variety of performance criteria, such as accuracy, precision, recall, and F1-score. To determine its efficacy, a comparative analysis will be performed against other existing approaches for musical genre classification.

The proposed system holds significant applications in the realm of music. It can automate the organization of music collections based on genre, facilitate genre-specific music searches, and offer personalized music recommendations based on user preferences. Additionally, it can enhance user experience in music streaming services by employing automatic classification of music genres.

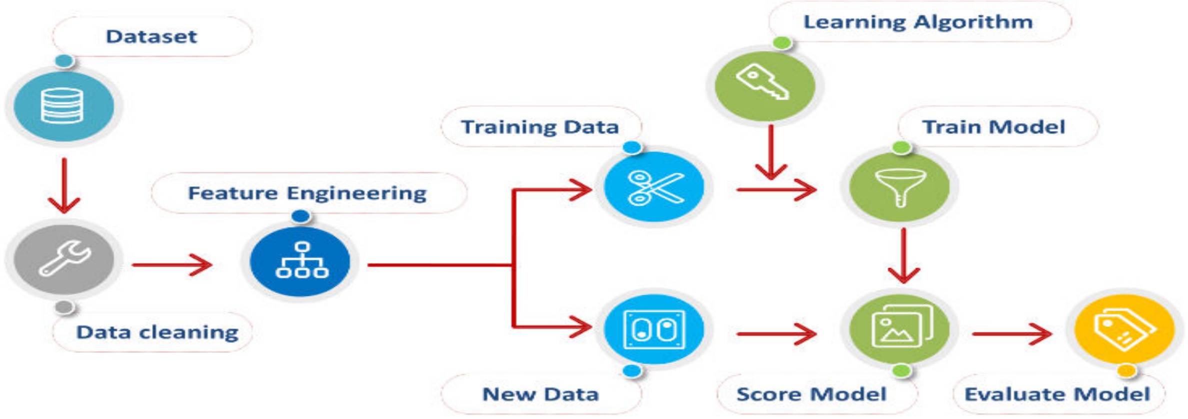
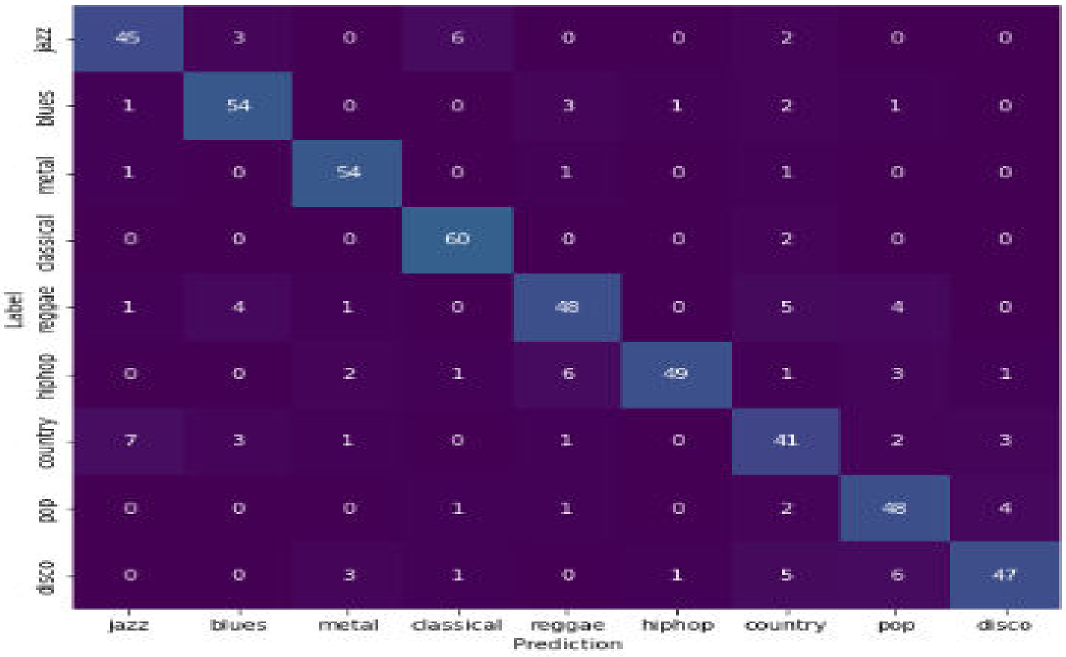


Fig 3: Proposed System Architecture

VI. Result Analysis

LeNet-5– Classification Report

The classification report displays the trained LeNet-5 classification model's accuracy, recall, F1 score, and support. Figure 4 shows all of the parameters from the Decision Tree categorization report.



	jazz	blues	metal	classical	reggae	hiphop	country	pop	disco
jazz	45	3	0	6	0	0	2	0	0
blues	1	54	0	0	3	1	2	1	0
metal	1	0	54	0	1	0	1	0	0
classical	0	0	0	60	0	0	2	0	0
reggae	1	4	1	0	48	0	5	4	0
hiphop	0	0	2	1	6	49	1	3	1
country	7	3	1	0	1	0	41	2	3
pop	0	0	0	1	1	0	2	48	4
disco	0	0	3	1	0	1	5	6	47

Fig 4: Confusion report for LeNet-5

LeNet-5 is a convolution neural network, or CNN, architecture designed specifically for the task of image recognition. An approach of assessing classification algorithm's performance is the confusion matrix.It depicts the model's share of accurate and faulty predictions for each class. Based on the confusion matrix, the LeNet-5 model has an accuracy of 82.59% in this scenario. This shows that the model correctly predicted the class for 82.59% of the evaluated items. The model's loss value of 0.6758 indicates how well it reduces prediction mistakes during training. Better performance is indicated by a lower loss value. For picture identification tasks, an accuracy of 82.59% is a respectable result overall.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202007 |

ROC Graph

A ROC (Receiver Operating Characteristic) graph as shown in Figure 5 visually depicts the performance of a binary classification model by displaying the trade-off between true positive rate (TPR) and false positive rate (FPR) as the classification threshold changes. The FPR represents the percentage of negative samples that were incorrectly labeled as positive, whereas the TPR represents the percentage of actual positive samples that were correctly classified as positive. The ROC chart compares TPR and FPR for various threshold values, generating a curve that displays the model's performance throughout the whole range of thresholds.

In an ideal scenario, a classifier would correctly classify all positive samples and avoid labeling any negative samples as positive, resulting in a TPR of 1 and an FPR of 0. The area under the receiver operating curve (AUC) is a widely used summary statistic for measuring classifier efficacy. A larger AUC implies a more effective classifier.Unlike other metrics such as accuracy, precision, and recall, the ROC graph provides insights into algorithm performance across various threshold values. This characteristic makes it a valuable tool for assessing the effectiveness of binary categorization algorithms, offering a comprehensive overview of their performance and aiding in the selection of suitable models for specific applications.

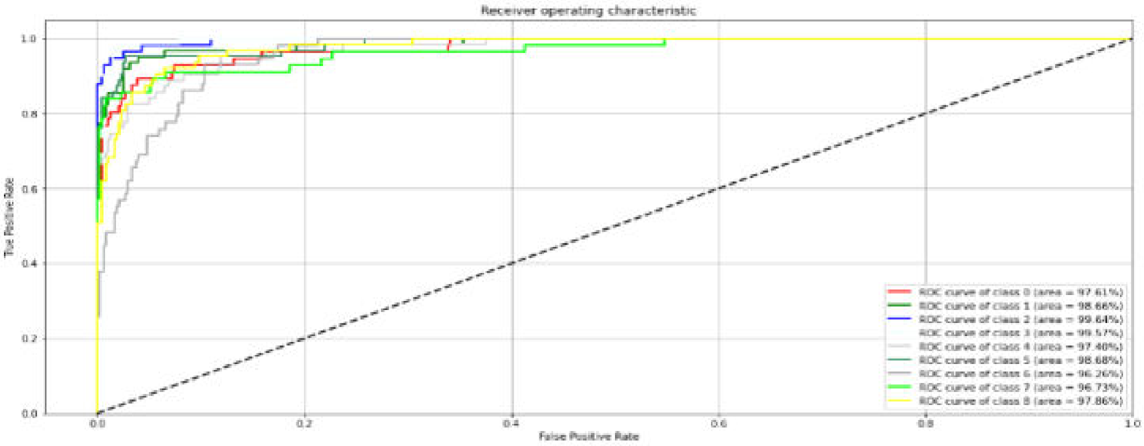


Fig 5: ROC curve

CNN-64 – Confusion matrix

A confusion matrix helps visualize the effects of a categorization effort by providing an orderly organization of the numerous forecasts and discoveries.

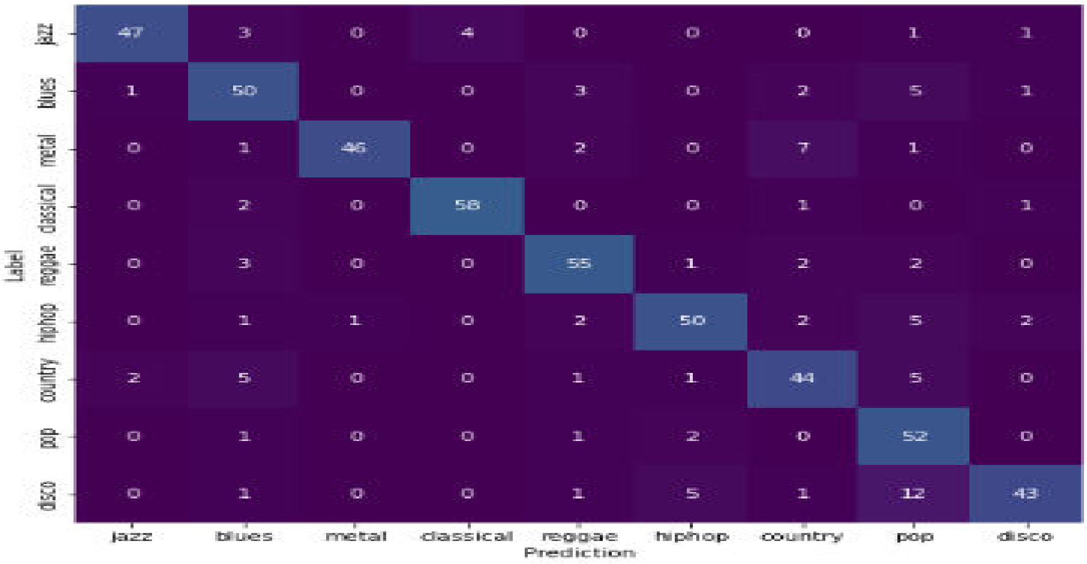


Fig 6: Confusion matrix for CNN-64

.Figure 6, also referred to as a confusion matrix, is used to assess how well a categorization model is working. Convolution neural networks, or CNN-64s for short, are deep learning models frequently employed in image recognition applications. The accuracy of a model is determined by the percentage of correctly classified samples in a dataset. In this case, the CNN-64 model predicts a class of 82.41% of the samples in the data set with an accuracy of 82.41%. The loss of a model tells us something about how well it performed during training. Diminished loss value indicates improved model performance. It is evident that the system is not yet optimized because the CNN-64 model's loss in this case is 1.0028.

ROC Graph

A Receiver Operating Characteristic (ROC) graph as in Figure 7 visualizes the performance of a binary classification model. It illustrates the relationship between the true positive rate (TPR) and the false positive rate (FPR) when the classification threshold changes. The TPR denotes the proportion of actual positive samples properly detected by the model, whereas the FPR represents the fraction of negative samples incorrectly categorized as positive.

The ROC graph, which plots the TPR against the FPR for various threshold values, provides a complete picture of the model's performance at all feasible thresholds. An ideal classifier has a TPR of 1 and an FPR of 0, indicating faultless identification of positive samples and no false positives.

The area under the ROC curve (AUC) is a typical statistic for comparing the efficiency of different classifiers. Higher AUC indicates greater performance because it evaluates the classifier's ability to discriminate between positive and negative data.The ROC graph is an invaluable tool for evaluating binary classification models. Unlike other evaluation metrics such as accuracy, precision, and recall, which focus on a single threshold, the ROC graph examines the model's performance across the entire range of threshold values. This broader perspective enhances the understanding of how the model's classification outcomes may vary depending on the threshold chosen.

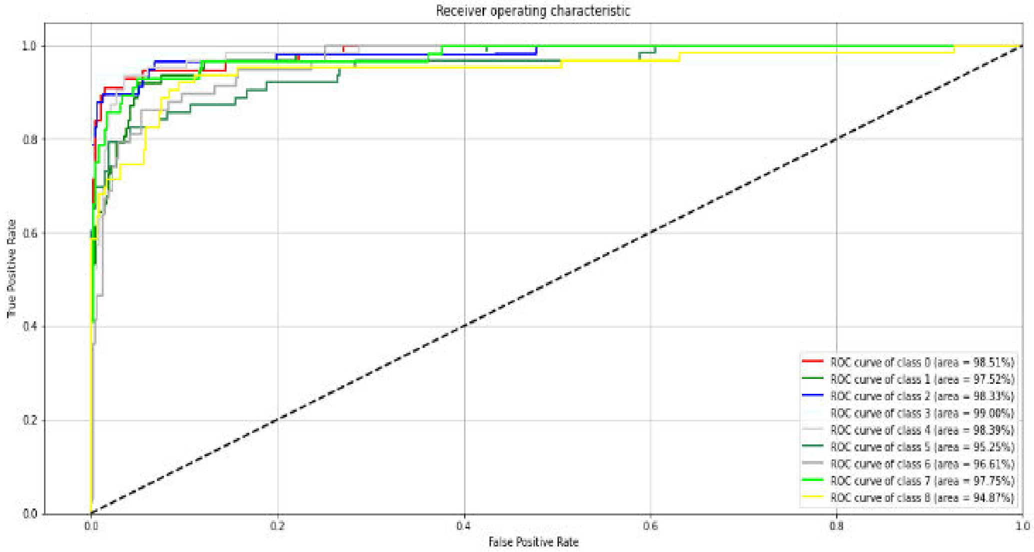


Fig 7: ROC curve

VII. CONCLUSION

The evaluation of machine learning models is an important task, and accuracy is one of the crucial metrics for assessing the performance of such models. In the case of classifying music genres, LeNet-5 and CNN-64 are two deep learning models that have been used to achieve this objective. Upon evaluating these models, it was observed that LeNet-5 achieved an accuracy of 82.59%, while CNN-64 had an accuracy of 82.41%. Both models achieved reasonably similar accuracies, indicating that they have the potential to be effective for music genre classification. The area under the ROC curve (AUC) is a typical statistic for comparing the efficiency of different classifiers. Higher AUC indicates greater performance because it evaluates the classifier's ability to discriminate between positive and negative data.The performance of the model can also be learned through other evaluation criteria including precision, recall, and F1-score. Therefore, it is necessary to analyze these metrics to gain a comprehensive understanding of the models' performance. Although LeNet-5 outperformed CNN-64 slightly in terms of accuracy, the difference between the two models is relatively small. Therefore, further analysis and testing are necessary to determine which model is more effective for music genre classification accurately. This analysis may involve comparing the models' performance on different datasets or using other evaluation metrics to obtain a more in-depth understanding of their performance. In conclusion, LeNet-5 and CNN-64 achieved promising results in classifying music genres using machine learning. While accuracy is an important metric for evaluating these models, it is crucial to consider other evaluation metrics to obtain a comprehensive understanding of their performance. For the best methodology for reliably classifying music genres, more research & testing may be required.

REFERENCES

1. George Tzanetakis and Perry Cook. 2002. Musical genre classification of audio signals. IEEE Transactions on speech and audio processing 10(5):293– 302.
2. Nicolas Scaringella and Giorgio Zoia. 2005. On the modeling of time information for automatic genre recognition systems in audio signals. In ISMIR. Pages 666–671.
3. Hagen Soltau, Tanja Schultz, Martin Westphal, and Alex Waibel. 1998. Recognition of music types. In Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on. IEEE, volume 2, pages 1137–1140
4. Michael I Mandel and Dan Ellis. 2005. Song-level features and support vector machines for music classification. In ISMIR. volume 2005, pages 594–599.
5. Thomas Lidy and Andreas Rauber. 2005. Evaluation of feature extractors and psycho-acoustictransformations for music genre classification. In ISMIR.pages 34–41.
6. E Zwicker and H Fastl. 1999. Psychoacoustics facts and models
7. George Tzanetakis and Perry Cook. 2002. Musical genre classification of audio signals. IEEE Transactions on speech and audio processing 10(5):293– 302.
8. Loris Nanni, Yandre MG Costa, Alessandra Lumini, Moo Young Kim, and Seung Ryul Baek. 2016. Combining visual and acoustic features for music genre classification. Expert Systems with Applications 45:108–117.
9. Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu. 2014. Convolutional neural networks for speech recognition. IEEE/ACM Transactions on audio, speech, and language processing 22(10):1533–1545.
10. Jort F Gemmeke, Daniel PW Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio set: An ontology and human-labeled dataset for audio events. In Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on. IEEE, pages 776–780.
11. Aaron Van Den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, NalKalchbrenner, Andrew Senior, and KorayKavukcuoglu. 2016. Wavenet: A generative model for raw audio. arXiv preprint arXiv:1609.03499 .
12. LonceWyse. 2017. Audio spectrogram representations for processing with convolutional neural networks. arXiv preprint arXiv:1706.09559 .
13. Tom LH Li, Antoni B Chan, and A Chun. 2010. Automatic musical pattern feature extraction using convolutional neural network. In Proc. Int. Conf. Data Mining and Applications.
14. Thomas Lidy and Alexander Schindler. 2016. Parallel convolutional neural networks for music genre and mood classification. MIREX2016.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING