Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.3, May : 2024

MUSIC GENRE CLASSIFICATION USING DEEP LEARNING

Shaik Rafi Professor, Department of Computer Science and Engineering, Narasaraopeta Engineering College, Narasaraopet, Andhra Pradesh, India shaikrafinrt@gmail.com,
K. Ravi Varma Student Department of Computer Science and Engineering, Narasaraopeta Engineering College, Narasaraopet, Andhra Pradesh, India 57varma@gmail.com,
M. Kodanda Rami Reddy Student Department of Computer Science and Engineering, Narasaraopeta Engineering College, Narasaraopet, Andhra Pradesh, India kodandaramireddymirla@gmail.com,
Shaik Madire Ashaduddeen Student Department of Computer Science and Engineering, Narasaraopeta Engineering College, Narasaraopet, Andhra Pradesh, India ashaduddeen2230@gmail.com

Abstract –
Music genre classification is a pivotal area of research within audio technology, holding immense importance for content organization and recommendation. Audio feature extraction and Music genre classification constitute a complete recognition system. Audio feature analysis and Music genre classification together form an integrated recognition system for comprehensive music genre identification and organization. This technology is frequently utilized to accurately detect and classify various types of music genres or characteristics present in audio signals, contributing significantly to the effective organization and recommendation of music content. Our experiment was conducted with the dataset from GTZAN that is taken from Kaggle repository. Convolutional neural networks (CNN) are employed to train our model, which is subsequently utilized for the classification of music genres in audio signals.

Keywords :
Music, Genre, Deep Learning, Support Vector Machine, CNN, CRNN

I. INTRODUCTION

Genre plays a crucial role in distinguishing between music pieces, though it can be influenced by personal biases. Despite diverse interpretations of genre on a global scale, the rise of digital platforms highlights the potential advantages of implementing automated music classification. This research delves into the world of automatic music genre classification, seeking to showcase the power of machine learning and deep learning techniques in accurately categorizing songs based on audio signals[1]. This automation has the potential to significantly decrease search times within the vast music databases commonly found on digital platforms[6].

Our study delves into the comparison of traditional machine-learning models, such as Support Vector Machines (SVM), and advanced deep-learning models like Convolutional Neural Networks (CNN) and Convolutional Recurrent Neural Networks (CRNN). By conducting this analysis, we aim to uncover the strengths and limitations of each method in the realm of music genre classification. Furthermore, we expand our investigation to evaluate the effectiveness of machine-learning classifiers using both three-second and thirty-second duration features[3].

This study adds valuable insight into the potential of these models and the importance of feature duration, further fueling the conversation on utilizing machine learning and deep learning to achieve more efficient and precise music genre classification[2]. This has the power to benefit all those within the ever-evolving world of digital music platforms.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.3, May : 2024

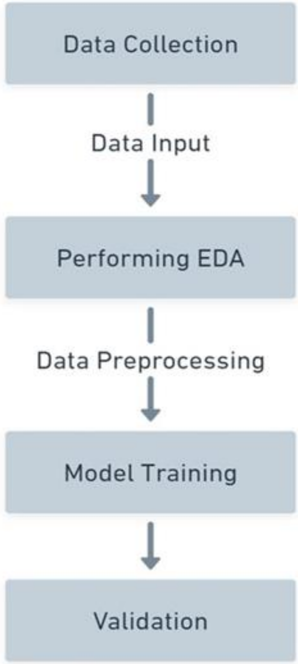

Fig-1 Steps involved in a Model

II. LITERATURRE SURVEY

The examination of music sort classification has earned critical consideration in inquire about circles, where different techniques have been investigated to progress exactness and effectiveness. A few essential considers have contributed to this field. Analysts have dug into different approaches to refine the classification handle, pointing to way better get it and categorize melodic sorts.Here are a few striking ponders within the domain of music sort classification.

Sturm et al.[1] proposed a music sort classification strategy utilizing scanty representation, accomplishing 83.00% accuracy on the GTZAN dataset. He extricated MFCC highlights, built genre-specific word references, and sp track utilizing iotas from its coordinating word reference.This approach viably utilized the discriminative control of genre-specific timbral characteristics.

Benetos et al.[2] investigated music sort classification through non-negative tensor factorization (NTF), coming to 75.00% accuracy on GTZAN. They spoken to music recordings as include tensors and created a novel NTF calculation. Sort classification was accomplished by breaking down the tensor into components capturing genre-specific designs and employing a administered classifier. Whereas not the beat entertainer at the time, this approach illustrated the potential of NTF for class classification and cleared the way for encourage headways in this region.

Bahuleyan et al.[7] displayed a strategy utilizing Convolutional Neural Systems (CNNs) for sound classification on the Sound Set dataset, accomplishing 65.00% accuracy. This approach included preparing a profound CNN design on a enormous collection of labeled sound recordings, permitting it to memorize complex designs specifically from the sound information. Whereas not particularly centered on music sort classification, this work showcased the potential of CNNs for sound assignments and propelled encourage investigate into their application for music class recognizable proof. It's critical to note that Audio Set could be a broader dataset enveloping different sounds past music sorts, making coordinate comparison with genre- specific datasets like GTZAN less direct.

Choi et al.[3] proposed a music class classification strategy utilizing Convolutional Neural Systems (CNNs) on the Million Tune Dataset, accomplishing an precision of 75.00%. Their approach, named Navier Music CNN, included preparing a profound CNN design on spectrograms determined from music sound. This permitted the organize to naturally learn pertinent highlights for class distinguishing proof straightforwardly from the unearthly representations.

These examinations grandstand a assorted cluster of methods and techniques utilized in anticipating music class classifications. They emphasize the noteworthiness of consolidating different highlights, counting melodic components, craftsman measurements, and relevant variables, to improve the exactness of class expectations. By investigating a run of approaches, analysts point to create strong models that can viably categorize and distinguish melodic classes based on comprehensive highlights and characteristics.

III. PROPOSED SYSTEM

Our Model is Proposed based on certain criteria as follows.

- Dataset Analysis
- Preprocessing Techniques
- Model Creation
- Result and Analysis

A. Dataset Analysis

The dataset utilized in this venture was sourced from Kaggle, a stage facilitating differing datasets from the web. Particularly, the GTZAN dataset, famous for its part in music sort classification inquire about, shapes the establishment of our examination. The GTZAN dataset comprises 1000 melodic selections, each enduring thirty seconds. Thisfastidiously categorized dataset envelops 10 unmistakable classes, with 100 pieces distributed to each class. In expansion to the initial dataset, a partitioned ponder was conducted to improve preparing comprehensiveness. The dataset was reproduced and subdivided into 10,000 passages, each enduring three seconds. Whereas this methodology expanded the accessible preparing information, it presented a certain degree of lopsidedness among the classes. A few classes finished up containing slightly more or less than the required 1000 pieces.

The GTZAN dataset, accessible on Kaggle, gives important data for music sort classification assignments. It incorporates sort names for each melodic excerpt, allowing for administered machine learning approaches.

The estimate and composition of the dataset are vital for understanding its scope. The first GTZAN dataset comprises of 1000 rows, equitably dispersed over 10 classes. The amplified dataset, made by imitating and subdividing the initial, contains 10,000 rows[7]. and professionals working with the GTZAN dataset within the context of music class classification thinks about.

B. Preprocessing Techniques

Compelling pre-processing procedures are necessarily to planning input information for a music sort classification show, particularly given the significant measure of our dataset. We utilize a assorted set of pre-processing strategies to handle challenges such as exceptions and clamor inborn in music information. Different strategies, counting name encoding and coding strategies like Power Transformer, StandardAero, and Min-Max Scaler, are connected to make a comprehensive technique for dealing with both categorical and numerical highlights.

The basis behind utilizing these pre-processing strategies lies in their capacity to improve the vigor and execution of the music sort classification demonstrate. Name encoding guarantees that categorical highlights are fittingly changed over into numerical arrange, encouraging the compatibility of information with machine learning calculations. In the interim, procedures like PowerTransformer, StandardScaler, and MinMaxScaler play a vital part in normalizing and scaling numerical highlights, tending to issues related to changing scales and disseminations.

Particular consideration is coordinated towards refining key highlights such as 'chroma_stft,' 'spectral_centroid,' and 'tempo.' Methods like PowerTransformer with Yeo-Johnson, StandardScaler, and MinMaxScaler are deliberately connected to guarantee compelling normalization and scaling for these highlights. This fastidious approach points to relieve potential challenges like predisposition and overfitting, contributing to the by and large vigor of the show. The consolidation of a ColumnTransformer advance upgrades the pre-processing pipeline by adeptly isolating categorical and numerical highlights. This isolation empowers custom fitted pre-processing for each sort of include, optimizing the planning of information for consequent stages within the classification demonstrate.

In substance, the intensive and comprehensive pre- processing methodologies utilized in this ponder are adapted towards ensuring that the input information is well- conditioned, minimizing commotion and exceptions, and maximizing the model's capacity to memorize important designs from the music class dataset.

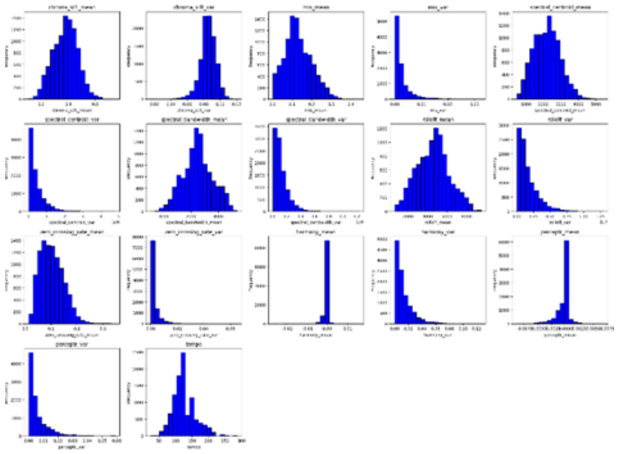

Fig-2 Distribution before Preprocessing

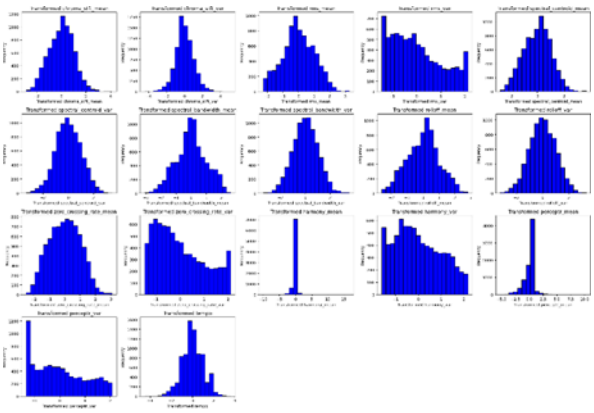

Fig-3 Distribution After Preprocessing

In Figure 2, the distribution of the dataset prior to preprocessing is visually represented, providing a snapshot of the initial data distribution. Notably, the distribution exhibits skewness, indicating an uneven spread across different classes or categories.

However, in Figure 3, we observe the distribution after the application of preprocessing techniques. The transformation is evident, with the skewed distribution evolving into a more normalized form. This shift towards a normal distribution is a significant achievement.

C. Model Creation

In the quest for precise music genre classification, three distinct models were developed: Support Vector Machine (SVM), Convolutional Neural Network (CNN), and Convolutional Recurrent Neural Network (CRNN). Each model was carefully designed to exploit specific strengths in handling diverse audio features. SVM, a traditional machine learning algorithm, excels in capturing intricate decision boundaries. The CNN model utilizes convolutional.

1) SVM

We utilized the Support Vector Machines (SVM) algorithm as the cornerstone of our machine-learning model, implementing it through the Scikit-Learn library. To fine- tune the SVM model, an exhaustive search for optimal hyperparameters was conducted, resulting in the selection of the following values:

'C': 42

'gamma': 1.52

'kernel': 'poly'

These specific hyperparameter choices were determined to deliver optimal performance within our experimental setup. In the subsequent sections, we will delve into comprehensive descriptions of our SVM models, covering architectural details, hyperparameter tuning specifics, and intricacies of the training process.

2) CNN

The Convolutional Neural Network (CNN) architecture in this research was constructed using Keras. The CNN built here has an input layer and five convolutional blocks, with each convolutional block consisting of the following:

- Dense Layers:

Four dense layers with 512, 256, 128, and 64 units respectively Each followed by ReLU activation and 0.2 dropout

- Training Configuration:

Optimizer: Adam Training Epochs: 200

3) CRNN

The Convolutional Recurrent Neural Network (CRNN) architecture in this research was constructed using Keras. The CRNN model built here has an input layer and five repeating blocks, with each CRNN block consisting of the following:

- Convolutional Layers:

Conv1D layer with 64 filters, 3 kernel size, and ReLU activation MaxPooling1D layer with a pool size of 2

- Dense Layers:

Four dense layers with 512, 256, 128, and 64 units respectively Each followed by ReLU activation and 0.2 dropout

- Compilation and Training Configuration: Optimizer: Adam Training Epochs: 200

IV. RESULT AND ANALYSIS

In this section, we delve into the outcomes and insights gained from the three distinct models employed in this research: Support Vector Machine (SVM), Convolutional Neural Network (CNN), and Convolutional Recurrent Neural Network (CRNN). Each model brings its unique approach to the task of classifying music into ten GTZAN genres. The ensuing analysis sheds light on their respective performances, revealing patterns, strengths, and areas for improvement.

The below figure-4 presents the confusion matrix attained when classifying music into ten GTZAN genres using Convolutional Neural Network model of this research. The model seems to perform well in general, with high correct classifications for most genres.

However for Classes Country (Class 2) and Jazz (Class 5) have relatively lower correct classifications compared to other genres.

The model often confuses Jazz (Class 5) with Blues (Class 0) and Classical (Class 1)

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.3, May : 2024

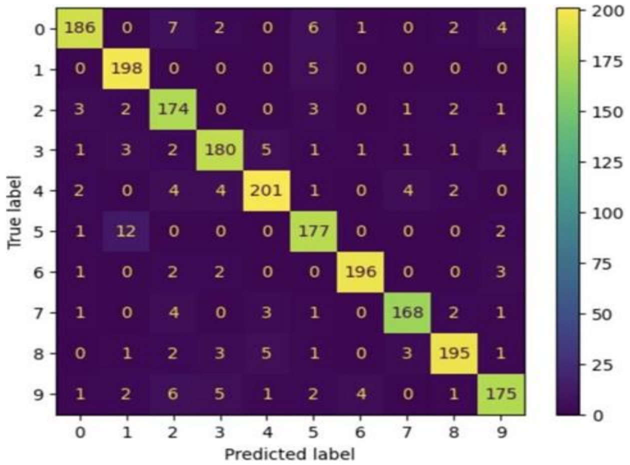

Fig-4 Confusion Matrix of CNN

The below figure-5 presents the confusion matrix attained when classifying music into ten GTZAN genres using Convolutional Recurrent Neural Network model of this research. However the model misclassifies some of genres like metal (Class 6) ,Country (Class 2) ,Jazz (Class 5)

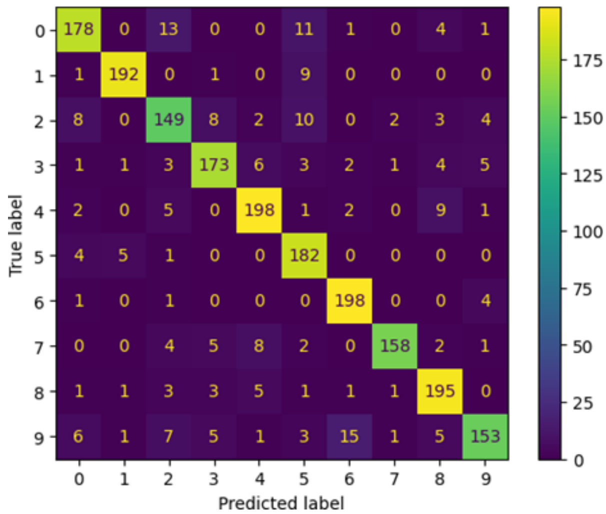

Fig 5 Confusion Matrix of CRNN

The Label Mapping: blues: 0,classical: 1,country: 2, disco: 3, hiphop: 4,jazz: 5,metal: 6, pop: 7, reggae: 8, rock: 9 Accuracy serves as a widely employed metric for assessing the efficacy of a machine learning algorithm. It quantifies the ratio of accurately classified instances relative to the entirety of instances within the test dataset.

TABLE I. ACCURACY OF EXISTING SYSTEM[4]

Classifier	Epochs	Accuracy
CNN (3-Sec Features)	50	72.4%
CNN (Spectrograms)	120	66.5%
CNN (30-Sec Features)	30	53.5%

TABLE II. ACCURACY OF DIFFERENT ALGORITHMS FOR 30_SEC_FEATURES

Classifier	Epochs	Accuracy	Training Time(S)
Support Vector Machines	-	73.5%	1
Convolutional Neural Network (CNN)	200	75.2%	27.5
Convolutional Recurrent Neural Network (CRNN)	200	72.0%	95.4

TABLE III. ACCURACY OF DIFFERENT ALGORITHMS FOR 3_SEC_FEATURES

Classifier	Epochs	Accuracy	Training Time(S)
Support Vector Machines	-	91.00%	2
Convolutional Neural Network (CNN)	200	93.58%	369
Convolutional Recurrent Neural Network (CRNN)	200	88.89%	841

In comparing the performance of our proposed models with those presented in the research paper titled "Music Genre Classification: A review of Deep-Learning and Traditional Machine-Learning Approaches" (published on April 21, 2021)[4], notable distinctions emerge across various temporal feature lengths. For the 30-second feature duration, our Support Vector Machine (SVM) model achieved an accuracy of 73.5%, slightly trailing behind the counterpart in the referenced paper, which reported 75.4%. Interestingly, our Convolutional Neural Network (CNN) model outperformed the alternative, exhibiting an accuracy of 75.2% compared to the referenced paper's CNN accuracy of 53.50%. This discrepancy underscores the efficacy of our CNN architecture in capturing temporal patterns for longer feature durations.

Shifting focus to the 3-second feature duration, our SVM model exhibited a substantial accuracy advantage with 91.0%, surpassing the accuracy of 80.8% reported in the referenced paper. Moreover, our CNN model demonstrated a significant improvement, achieving an accuracy of 93.5%, in contrast to the 72.40% accuracy recorded in the referenced paper. These results suggest that our models excel in capturing short-term temporal features, showcasing superior performance in both SVM and CNN architectures. These findings contribute valuable insights to the field, emphasizing the significance of our proposed approaches in achieving heightened accuracy for temporal feature extraction.

V. CONCLUSION AND FUTURE SCOPE

This investigation into music genre classification using SVM, CNN, and CRNN models has yielded valuable insights. The robust SVM demonstrated commendable accuracy at 91%, while the deep learning models outperformed, with CNN achieving an impressive 93% and CRNN reaching 78%. This underscores the significance of capturing both spatial and temporal data, with CNN showcasing notable feature extraction capabilities. The study reaffirms the dominance of deep learning, particularly CNNs, in this field, highlighting the critical roles of meticulous feature engineering, model tuning, and data curation for success. Looking ahead, the future of music genre classification lies in the exploration of optimal combinations of traditional and deep learning techniques to leverage their complementary strengths.

REFERENCES

[1]L. Sturm, “On music sort classification by means of compressive sampling,” in 2013 IEEE Universal Conference on Mixed media and Expo(ICME), 2013, pp. 1–6.
[2]E. Benetos and C. Kotropoulos, “A tensor-based approach for programmed music sort classification,” in 2008 16th European Flag Preparing Conference, 2008, pp. 1–4.
[3]K. Choi, G. Fazekas, and M. Sandler, “Explaining profound convolutional neural systems on music classification,” 2016.
[4]Music Genre Classification: A review of Deep-Learning and Traditional Machine-Learning Approaches. (2021, April 21).
IEEE Conference Publication | IEEE Xplore.https://ieeexplore.ieee.org/abstract/document/9422487
[5]J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kégl, “Aggregate highlights and adaboost for music classification,” Machine Learning, vol. 65,
pp. 473–484, 12 2006
[6]I. Fujinaga, “Adaptive optical music recognition,” Ph.D. paper, McGill College, CAN, 1997,

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.3, May : 2024
aAINQ29937.
[7]H.Bahuleyan, “Music sort classification utilizing machine learning techniques,” 2018.
[8]D. S. Lau and R. Ajoodha, “Music class classification:A comparative consider between deep-learning and conventional machine learning approaches,” in 6th Worldwide Congress on Data and Communication Innovation (6th ICICT). Springer, 2021, pp. 1–8.
[9]J. H. Lee and J. S. Downie, “Survey of music data needs, employments, and looking for practices: preparatory findings.” in ISMIR, vol. 2004. Citeseer, 2004, p. 5th.
[10] A. Lerch, An Presentation to Sound Substance Examination: Applications in Flag Handling and Music Informatics. Wiley Online Library, 10 2012.
[11] T. Li, M. Ogihara, and Q. Li, “A comparative think about on content- based music class classification,” in Procedures of the 26th Yearly Universal ACM SIGIR Conference on Inquire about and Improvement in Informaion Recovery, ser. SIGIR '03. Unused York, NY, USA: Affiliation for Computing Apparatus, 2003,p.282– [Online].Available:https://doi.org/10.1145/860435.860487
[12] T. Lidy, A. Rauber, A. Pertusa, and J. Iñesta, “Combining sound and typical descriptors for music classification from audio,” 2007.
[13] C. McKay, R. Fiebrink, D. McEnnis, B. Li, and I. Fujinaga, “Ace: A system for optimizing music classification,” in ISMIR, 2005.
[14] C. McKay and I. Fujinaga, “Musical class classification: Is it worth seeking after and how can it be improved?” in ISMIR, 2006.
[15] T. Nkambule and R. Ajoodha, “Classification of music by class utilizing probabilistic graphical models and profound learning models,” in 6th Worldwide Congress on Data and Communication Innovation (6th ICICT). Springer, 2021, pp. 1–6.
[16] R. Ajoodha, R. Klein, and B. Rosman, “Single-labelled music class classification utilizing content-based features,” in 2015 Design Acknowledgment Affiliation of South Africa and Mechanical autonomy and Mechatronics Universal Conference (PRASA- RobMech), 2015, pp. 66–71.
[17] R. Ajoodha, R. Klein, and M. Jakovljevic, “Using measurable models and developmental calculations in algorithmic music composition,” in Reference book of Data Science and Innovation, Third Version. IGI Worldwide, 2015, pp. 6050–6062.
[18] A. Olteanu. Gtzan dataset - music sort classification. [Online]. Available: https://www.kaggle.com/andradaolteanu/gtzan-dataset- music-genre- classification.
[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Pretten hofer, R. Weiss,
V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay, “Scikit- learn: Machine learning in python,” 2018