Decoding the Subjective Symphony: Machine Learning for Music Genre Classification Using Audio Features and Taxonomy Reduction

V. Ajitha, J. Nishitha, R. Naga Rishika, Md Sufiyan, K. Sai Umanth and K. Laxmi Narayana
Department of Computer Science and Engineering (AIML), Nalla Malla Reddy Engineering College, Hyderabad, India

Keywords: Music Genre Classification, Machine Learning, Audio Features, XGBoost, Neural Networks, K-Nearest Neighbours, Ensemble Models, Multi-Label Classification.

Abstract: This paper addresses the application of machine learning to categorize music genres using audio attributes from a dataset of 114,000 songs representing 125 genres. Addressing the difficulty of categorizing detailed, subjective audio data into several genres, the research aimed to predict both individual and multi-genre classifications while decreasing the goal taxonomy from 114 to 56 genres using hierarchical clustering. Through exploratory data analysis, the dataset's 15 features (11 numeric, 4 categorical) were pre-processed, non-audio genres removed, and data balanced. Models tested included a neural network, Boost, K-nearest neighbours, and an ensemble approach. Evaluated using top-3 categorical accuracy, a critical metric for recommendation systems in which neural network achieved the highest performance (73.74%), followed by the XG Boost (69.54%) and KNN (70.36%). Results revealed that genres with separate aural properties were labelled more accurately than those with overlapping traits, underscoring the subjective complexity of musical categorization. The study implies that while machine learning shows progress, genre subjectivity remains a basic hurdle. Future directions include refining ensemble techniques, adding lyrics, and exploring multi-label classification to enhance accuracy and nuance in music categorization.

1 INTRODUCTION

Music is part of human culture, transcending boundaries and various languages. From traditional folk music to modern complex electronic music, music is diverse and exists in various forms, typically categorized into various types. This categorization aids listeners but poses giant issues to music suppliers and researchers. As the volume of available music data continues to rise, understanding and automating genre classification has become an important subject of study within the discipline of machine learning. The rise of sites such as Spotify, which counts over 70 million music in its catalog, has further complicated genre classification responsibilities. Although the evaluation of the auditory qualities and accompanying metadata gives some insight, innate subjectivity in genre classification makes it a very hard undertaking. Historically, genres have been developed according to cultural and historical settings, which often lead to overlaps and ambiguities. For instance, a song may contain elements of various styles such as pop, jazz, or electronic, and it becomes difficult to categorize both for humans and computers. New machine learning technologies provide new methods to manage this diversity. By applying massive datasets and sophisticated computer approaches, researchers may now examine the tremendous possibilities for automated music genre classification. The issue comes not just in the algorithms themselves but in identifying the necessary features to train these models. Features identified by audio analysis, such as beat, tempo, and timbre, are vital in appreciating how varied sounds compose the heart of a musical genre. The confluence of auditory features with various machine learning methods gives a viable path for boosting categorization accuracy.

2 LITERATURE REVIEW

The study of music genres is an important element of music analysis and classification. Various research initiatives have been performed to study different parts of music genre classification, including the building of datasets, the invention of novel music acoustic encoders, and the application of deep

Ajitha, V., Nishitha, J., Rishika, R. N., Sufiyan, M., Umanth, K. S. and Narayana, K. L.
Decoding the Subjective Symphony: Machine Learning for Music Genre Classification Using Audio Features and Taxonomy Reduction.
DOI: 10.5220/0013920400004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 4*, pages 762-766
ISBN: 978-989-758-777-1

learning algorithms. Roberts et. al., 2020 produced a dataset for time-scale modification (TSM) with subjective quality descriptors, spanning a wide range of music genres. This dataset serves as a great resource for obtaining objective measurements of quality in TSM. Zhao et. al., 2020 proposed Music Oder, a universal music-acoustic encoder based on transformers, which outperformed existing models in music genre categorization and auto-tagging tasks. These breakthroughs in dataset development and acoustic representation learning lead to the improvement of music genre classification systems. Cuesta et. al., 2020 focused on multiple F0 estimation in voice ensembles using convolutional neural networks, demonstrating the usefulness of CNNs in varied circumstances and data configurations. Lerch, 2020 examined audio content analysis in the context of music information retrieval systems, stressing music genre classification as one of the primary applications of audio content analysis. The study by Ozan, 2021 uses convolutional recurrent neural networks (CRNN) for audio segment categorization in contact centre records, drawing parallels to music genre classification problems' et. al., 2021 extended the usage of deep learning to electronic dance music (EDM) subgenre categorization, including tempo-related feature representations for better classification accuracy. Muñoz-Romero et. al., 2021 studied nonnegative orthogonal partial least squares (OPLS) for supervised design of filter banks, with applications in texture and music genre categorization. These studies demonstrate the diverse methodologies and techniques utilized in music genre classification research. Zhao et. al., 2022 created a self-supervised pre-training technique with Swim Transformer for music categorization, underlining the importance of learning meaningful music representations from unlabelled data. Chak et. al., 2022 presented the use of Generalized Morse Wavelets (GMWs) in the Scattering Transform Network (STN) for music genre categorization, proving the superiority of this approach over conventional methods. These works illustrate the continual study of fresh strategies to boost music genre classification accuracy. Ian, 2022 focuses on optimizing musician impact models and assessing musical qualities across different genres, highlighting the necessity of recognizing the influence of performers in music categorization tasks. Liu et. al., 2022 discussed open set recognition for music genre classification and presented an algorithmic approach towards the segmentation of known as well as unknown genre classes. Heo et. al., 2022 proposed a framework for hierarchical feature extraction and aggregation in the classification of music genres so that short and long-term musical features are captured appropriately.

3 METHODOLOGY

The music genre categorization experiment utilized a dataset of 114,000 tracks collected from the Spotify API, which comprised audio attributes and information spanning 125 categories. The dataset was partitioned into a training set of about 91,200 tracks and a test set consisting of about 22,800 tracks to evaluate model performance. The independent variables comprised numerous audio parameters such as danceability, energy, loudness, acoustics, pace, and categorization features like key and time signature, whereas the dependent variable was the track genre. The dataset is realistic and interesting due to its significant size, wide genre representation, and rich audio features that mirror real-world music categorization issues. This intricacy is further highlighted by the overlap between genres, which often share similar auditory traits. This poses a major challenge to effective classification in the context of machine learning. These characteristics not only increase the model's applicability to real-world scenarios in music streaming services, but they also provide an interesting exploration of the intricacy of music, illuminating the challenges associated with genre classification and the effectiveness of algorithms in this regard.

Algorithm: Music Genre Classification Using Neural Networks

Data Preparation

1. Load Dataset: Import dataset D comprisinging audio features X and genre labels y.
2. Clean Dataset:
    - Remove duplicate tracks to maintain uniqueness,eliminate non-sound-based genres (e.g., language categories) and extraneous attributes (e.g., track IDs).
    - Implement One-hot encoding for categorical features (e.g., key, time signature).
    - Convert boolean features (e.g., explicit content) to binary values (0/1).
3. Split Data: Divide D into training (Xtrain,ytrain) and test sets (Xtest,ytest) using test split ratio τ.
4. Normalize Features: Apply *StandardScaler* to numerical features in Xtrain and Xtest.
5. Consolidate Genres:
    - Perform hierarchical clustering on ytrain and ytest using Ward's method and Euclidean distance.

- Merge genres into g broader categories via consolidation threshold θ (e.g., reduce 125 genres to 56).

Model Architecture

1. Build Neural Network:
   - Input Layer: Neurons = feature dimensionality of X.
   - Hidden Layers: Sequential layers with 256, 128, 64, and 32 neurons (ReLU activation).
   - Output Layer: Softmax activation for genre probabilities.
2. Compile Model:
   - Optimizer: Adam.
   - Loss Function: Sparse categorical cross-entropy.
   - Metrics: Accuracy and top-k categorical accuracy.

Training

1. Train Model: Fit the model M on Xtrain and ytrain for E epochs. Validate performance on Xtest and ytest.

Prediction

1. Generate Predictions: For each test sample xi ∈Xtest:
   - Compute softmax probabilities pi=M(xi).
   - Rank genres by descending probability.
2. Interpret Results: Assign the genre with the highest probability (e.g., "Pop" with probability p=0.65).

Evaluation

1. Compute Metrics:
   - Top-1 Accuracy: Proportion of exact matches between predicted and true genres.
   - Top-k Accuracy: Proportion of true genres in the top k predictions (e.g., k=3).
2. Store Results: Save metrics in Meval.

Results

The trials encompassed four distinct machine learning models: Neural Network, XGBoost, KNN Classifier, and Ensemble Model. In the classification of music genres, numerous performance measures were gathered to evaluate the efficacy of different machine learning models in predicting musical genres based on aural characteristics.

3.1 Top-K Categorical Accuracy

The Neural network model proved to be highly effective for multi-class classification, as demonstrated by its higher overall performance, notably in the Top-3 ranking. In both metrics, the XGBoost performed somewhat worse than neural network model, requiring a lot of processing power despite having a good capacity. Despite having a lower Top-1 accuracy, the KNN Classifier was competitive in terms of Top-3 accuracy and substantially faster in terms of training time. KNN

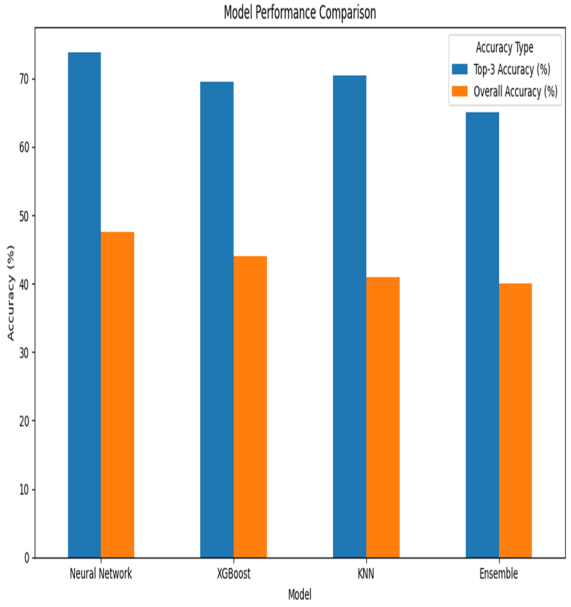

Figure 1: Top-1 Accuracy Shows the Percentage of Times the Model Predicted the Exact Genre Correctly, With Neural Network Leading at 47.47%. Top-3 Accuracy Reflects How Often the True Genre Appeared Within the Top Three Predicted, With Neural Network Again on Top at 73.74%.

Classifier scored better than the Ensemble model but lagged in both Top-1 and Top-3 accuracy compared ZZ to XGBoost and Neural Networks. Figure 1 shows the Top-1 Accuracy shows the percentage of times the model predicted the exact genre correctly, with Neural network leading at 47.47%. Top-3 Accuracy reflects how often the true genre appeared within the top three predicted, with Neural network again on top at 73.74%. However, the KNN model's relative speed and efficiency make it a desirable alternative in less resource-intensive settings. The Ensemble Model fared poorer than the other models, most likely as a result of including simpler models (like SVC and Logistic Regression) that had problems with the dataset's complexity. Ensemble Model incorporated numerous classifiers but resulted in the least favorable performance, both in Top-1 and Top-3 accuracy. This shows that the simpler models contained in the ensemble may not be consistent with the complexity necessary for music genre classification, resulting to lower overall predictions.

3.2 Overall Accuracy

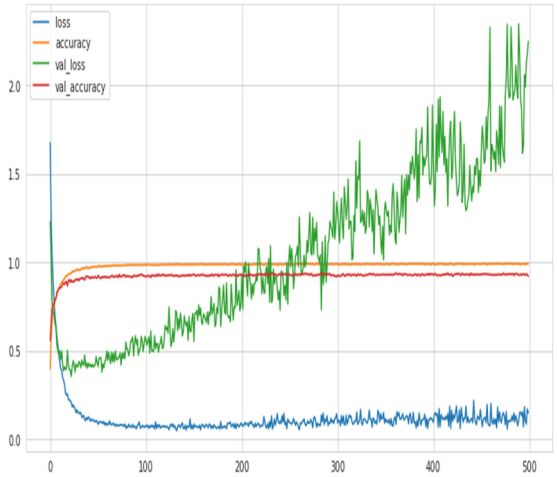

Figure 2: CNN Max. Accuracy 0.92.

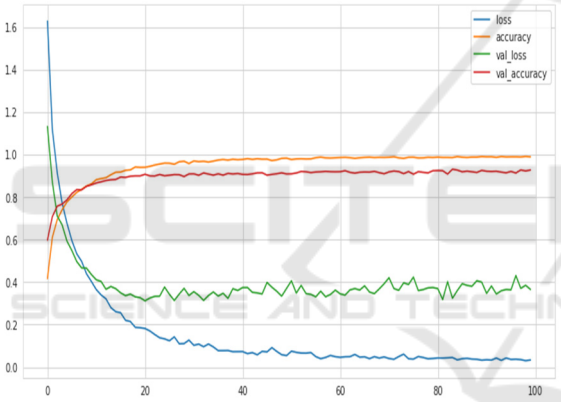

Figure 3: XGBoostmax. Accuracy 0.91.

Convolutional Neural Network gives the greatest balanced and consistent performance across all permutations. Its training and validation metrics indicate smooth convergence, with validation loss closely resembling the training loss, indicating high generalization capabilities. The figure 2 , 3,4,5 shows the accracy of mdels XGBoost has great accuracy but exhibits problematic behaviour in its validation measures. While reaching 95-98% training accuracy, the highly changing and growing validation loss shows the model is overfitting to the training data rather than learning generalizable patterns. This shows a need for stronger regularization strategies, either through dropout layers, reduced model complexity, or adopting early halting around epoch 200. While its overall accuracy is significantly lower than the CNN at roughly 90-95%, it maintains constant validation metrics during training. The stability in validation loss and very low variance makes it a dependable choice, especially when speedy deployment is favoured above reaching the exact best accuracy.

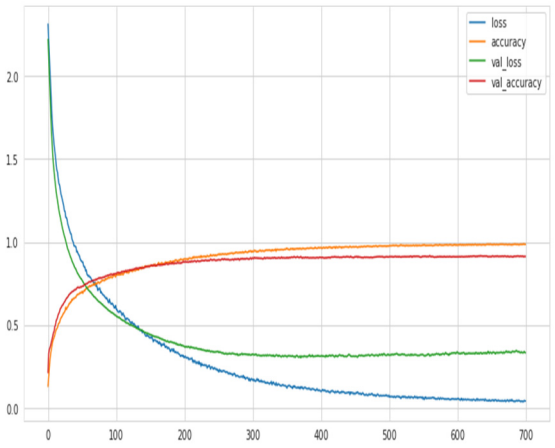

Figure 4: KNN Validation Accuracy 0.93.

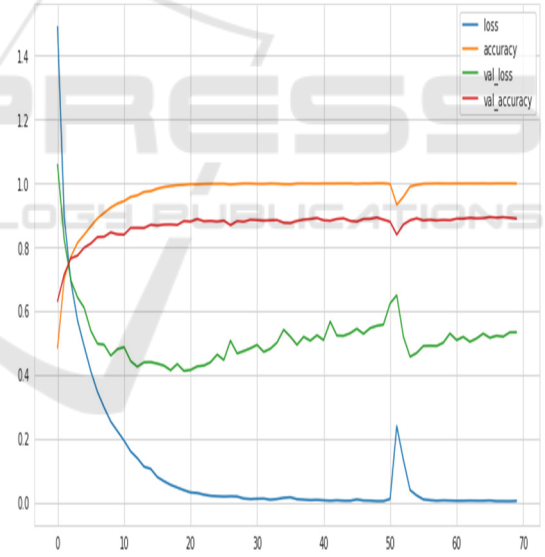

Figure 5: Ensemble Validation Accuracy 0.89.

The ensemble technique gives an interesting middle ground, displaying smooth learning progression and robustness to overfitting. There's a large performance increase at epoch 50, presumably owing to learning rate tweaks, but the model maintains steady accuracy around 85-90% with stable validation measures. This shows that merging numerous models helps offset individual model faults while maintaining consistent performance. Given these results, CNN looks to be the most trustworthy

standalone solution for genre categorization, delivering the optimum mix between performance and computational efficiency.

3.3 Genre-Wise Accuracy

Table 1: Accuracy Distribution Through Genre.

Model	Best Predicted Genres (Accuracy)	Worst Predicted Genres (Accuracy)
Neural Network	Grindcore (86%), Study (80%), Comedy (78%)	Emo (5%), Blues - R&B (11%), Techno (18%)
XGBoost	Grindcore (87%), Sleep (85%), Study (81%)	Techno (13%), Emo (16%), Blues - R&B (16%)
KNN	Comedy (77%), Grindcore (75%), Sleep (70%)	Techno (10%), Trip-Hop (8%), Emo (8%)

This data reveals which music genres were simpler or more difficult to categorize. Genre-specific accuracy figures highlighted trends in performance, allowing discussions about genres that algorithms find particularly challenging. The table 1 shows the Accuracy Distribution Through Genre. For instance, genres characterized by distinct sonic features tended to be categorized more accurately, whereas those with broader or less defined characteristics saw lower accuracy scores.

4 CONCLUSIONS

This investigation evaluated four machine learning models for the classification of music genres: Neural Network, XGBoost, K-Nearest Neighbors Classifier, and Ensemble Model. Each model displayed distinct strengths and limitations. The Neural Network demonstrated the highest precision in Top-1 (47.47%) and Top-3 (73.74%) predictions, excelling in multi-class tasks despite its demand for substantial computational resources. XGBoost showed a competitive accuracy level; however, it showed signs of overfitting, which highlights the need for applying regularization methods. The KNN Classifier showed an effective balance between computational cost and Top-3 accuracy, with a rate of 68.5%, thus making it suitable for resource-constrained environments. The Ensemble Model showed poor performance, possibly because it is based on basic classifiers that do not fully capture the complexities of the dataset. Future studies should explore the creation of hybrid architectures that work towards minimizing overfitting and taking advantage of heterogeneous feature representations. The incorporation of multimodal data with sophisticated regularization methods can improve accuracy in domains with uncertainty.

REFERENCES

Cuesta, Helena, et al. "Multiple F0 Estimation in Vocal Ensembles Using Convolutional Neural Networks." ArXiv.org, 2020, arxiv.org/abs/2009.04172.

Roberts, Timothy, and Kuldip K. Paliwal. "A Time-Scale Modification Dataset with Subjective Quality Labels." The Journal of the Acoustical Society of America, vol. 148, no. 1, 1 July 2020, pp. 201210,https://doi.org/10.1121/10.0001567

Zhao, Yilun, and Jia Guo. "MusiCoder: A Universal Music-Acoustic Encoder Based on Transformer." Lecture Notes in Computer Science, 1 Jan. 2021, pp. 417429, arxiv.org/abs/2008.00781,https://doi.org/10.1007/978-3-030-67832-6_34.