Classification of Music Based on Genre Using Hybrid Artificial Intelligence Models

Harshini J
School Of Computer Science and Engineering
Vellore Institute Of Technology(VIT) Chennai
Chennai, India
harshini.j2021@vitstudent.ac.in

Gerina Mary C
School Of Computer Science and Engineering
Vellore Institute Of Technology(VIT) Chennai
Chennai, India
gerinamary.c2021@vitstudent.ac.in

Abstract— **The aim of this project is to create an efficient automatic music classification system that will be driven by the need for customized song recommendations and a well-organized listening experience. Even though the genre is subjective, it's difficult to place different types of music under distinct categories. Our approach combines several feature extraction techniques such as spectral analysis, temporal patterns and timbral characteristics with advanced machine learning algorithms like deep neural networks. In order to improve classification accuracy, we propose a multi-modal data fusion where disparate information sources are merged together. The results obtained from our experiments show that our method achieves high classification accuracy rates, it is also robust to cultural influences and intra-genre variations. This particular project makes a contribution towards improving automated music classifications in music search systems.**

Keywords: music classification, automated, feature extraction, machine learning, genre boundaries, multi-modal data fusion

I. INTRODUCTION

The motivation for this study arises from the expanding digital music consumption landscape, with its huge and diverse music choices that pose challenges to users as well as to music service providers. Therefore, in this context, there is a growing need for more advanced automated music classification systems. They have the potential of providing even more personalized and accurate recommendations on music while facilitating smooth organization and retrieval of musical content leading to increased user satisfaction and engagement.

This research revolves around the complicated issue of putting music into categories or genres without mistakes. Although it seems easy to assign different genres to a song, genre classification is a complex task due to many variables. First, the definitions of genres are subjective and changeable hence; they differ from one culture, period and person to another drastically. Furthermore, musical compositions within the same genre vary greatly in terms of style, instruments used as well as themes employed. Consequently, automatic music genre classification is not a simple task because there are very many things that need to be considered while trying to make such classifications in addition to the fact that innovative ways and methodologies should be adopted for better results.

The overarching objectives of this study encompass two primary facets: firstly, to develop robust and adaptable classification models capable of effectively addressing the nuanced intricacies inherent in music genre classification, and secondly, to conduct a comprehensive comparative analysis of different methodologies to discern the most efficacious approach.

Within the available music classification methods, many different techniques have been proposed and explored. These methods usually involve extracting key features from music data, then applying machine learning algorithms for classification. Characteristic extraction techniques include a variety of approaches including, but not limited to, spectrum analysis, temporal patterns, timbral properties, rhythmic structures, and harmonic content, each of which provides unique insights into the underlying musical properties.

To achieve the objectives, four different models were carefully selected for music classification in this study: Convolutional Recurrent Neural Network (CRNN), Long Short-Term Memory (LSTM) with Support Vector Machine (SVM), Convolutional Neural Network (CNN) with k-Nearest Neighbors (KNN) and CNN with SVM. Each of these models represents a sophisticated amalgamation of cutting-edge techniques, tailored to address specific challenges and exploit different facets of music data.

The CRNN model is a combination of convolutional and recurrent neural network architectures that can effectively capture both spatial and temporal dependencies of sequential music data. The LSTM+SVM model uses the memory retention capacity of LSTM networks to model sequences, followed by the discrimination power of SVM in classification, which allows long-term temporal dependencies to be included in music sequences. On the other hand, the CNN + KNN model provides a hybrid approach that combines the characteristic learning capabilities of convolutional neural networks with k-nearest neighbor event-based inference that uses both local and global similarity measures. Finally, the CNN + SVM model uses the robust feature representation learned by CNNs, followed by the discriminative classification ability of SVM, resulting in a synergistic fusion of complementary methods.

The purpose of this study is to delve into the complexities of each model and explore their strengths, weaknesses and performance characteristics across different datasets and scenarios through a thorough and careful comparative

analysis. By carefully evaluating the effectiveness and robustness of these methods, the study aims to provide valuable insights into the optimal approach to classify music into well-defined genres, push the boundaries of automated music classification systems, and advance the field.
Through this research, we aim to provide an overview of the performance of these models and their suitability for automated music classification tasks. By performing a comparative analysis, we try to find out the most efficient way to accurately classify music into predefined genres, thus contributing to the development of automated music classification systems.

II. LITERATURE SURVEY

[1] To meet the demands of the people, it is essential that music be categorized based on genres. The application of machine learning algorithms for music genre classification has attracted a lot of attention in the field of music. In this paper, the author has used CNN based approach for classification. The authors utilized mel-spectrograms as CNN input features, emphasising how well these visual depictions of audio signals can capture key elements of music. It also used different sound file characteristics stored in the Decision Tree, SVM, MLP, and ANN csv archives, and 91% of them were equivalent to the human understanding of the genre with the highest accuracy achievement.
[2] Mel frequency cepstral coefficient (MFCC) and feature extraction are two sophisticated techniques that the authors use to tackle the difficult job of automatic genre classification. With a focus on neural network topologies, the paper examines the subtleties of convolutional layers and training procedures through the prism of deep learning. The combination of different approaches is discussed, providing insights into the field of music classification. Through the perspective of convolutional neural networks (CNNs) and deep learning approaches, in particular, this work makes a noteworthy contribution to the rapidly developing field of music genre classification.
[3] One of the most common characteristics that sets apart musical compositions is genre. Although there are widely accepted categories of genre, responses to genre by individuals can be biassed. The authors cover a wide range of techniques. Convolutional Neural Networks (CNNs) and Mel Frequency Cepstral Coefficients (MFCC) as essential components of the deep learning space. They also discuss Support Vector Machines (SVMs) and other traditional machine learning techniques. Following a series of training iterations, they found that k-Nearest Neighbours (kNN) model yielded the highest accuracy of 92.69%, while CNN model yielded provided relatively low accuracy, attaining 72.40%.
[4] The authors use a hybrid neural network architecture, which mixes recurrent and convolutional layers, to present a novel method. This integration shows a sophisticated grasp of the intricate patterns present in many musical genres, indicating a thorough investigation of spatial and temporal aspects inside audio data. The research offers a complex model for music genre categorization by utilising cutting-edge methods including feature extraction and the hybridization of convolutional and recurrent networks. The 3-second audio clips that were taken from the GTZAN dataset were utilised by the author to train the model using a residual neural network (RNN). A 94% accuracy rate was attained when overlapping elements from several genres were taken into account. The investigation should shed light on how effective this architecture is, maybe leading to improvements over more traditional approaches. Through the lens of state-of-the-art neural network designs, the work is important in expanding our understanding of music genre recognition.
[5] Convolutional neural network (1D CNN) architectures are the authors' primary area of application. Set within the larger field of computer intelligence, the paper presents a specialised investigation of deep learning methods designed for sequential data in the form of music signals. The authors test several 1D CNN setups to determine how well they capture pertinent information for genre categorization since they specifically address the distinct temporal patterns present in music. The research is anticipated to provide important insights into the most advanced techniques for classifying musical genres by illuminating the design decisions and performance assessments of these structures. The understanding of how to use 1D CNNs to improve the precision and effectiveness of music genre classification systems is advanced by this work, which is especially pertinent.
[6] This paper provides with the comparison of the automatic results with human genre classifications on the same dataset. Automated music genre classification makes it possible to automatically arrange and structure huge music archives. It also offers a useful method for contrasting and assessing feature sets that aim to depict musical content. The study carefully assesses the performance of classification algorithms against the complex and frequently subjective assessments of human listeners by utilising Mel Frequency Cepstral Coefficients (MFCC) for feature extraction, a method praised for its effectiveness in capturing the fundamental characteristics of audio signals.
[7] Using deep learning techniques, the study tackles the current issues with music genre classification. The authors particularly point out the use of CNNs in conjunction with Support Vector Machine (SVM) classification, highlighting the mutually beneficial effects of these techniques. Phrases like "music features," "mel frequency cepstral constant," and "feature extraction" allude to a thorough investigation of sophisticated methods for separating out discriminating information from audio input. Further indicating a multimodal approach is the integration of deep learning and K-Nearest Neighbours (KNN). The training procedure, deep learning architecture, and support vector machine integration for ultimate classification are covered in the paper. With an accuracy level of roughly 97% for training and 74% for testing, the project will greatly improve and promote the categorization of musical genres.
[8] With a particular focus on the application of CNNs to automatically classify music genres, the paper explores the nexus between deep learning and music analysis. This study describes the architecture of the used CNN, the training procedure, and the use of Mel-spectrum to capture frequency properties. Moreover, the classification procedure must take temporal factors into account when time series analysis is used. To transform the source audio files into the appropriate Mel spectrums, Librosa is utilised. Next, the suggested CNN

model is trained using the converted Mel spectrum. The 10 classifiers' judgements are subject to majority voting, and the average accuracy on the GTZAN dataset is 84%. By demonstrating developments in the application of deep learning techniques to the field of music analysis, this work probably adds to our understanding of how well CNNs perform when it comes to music genre classification.

[9] This research focuses on the classification of musical genres using taxonomy. Their investigation into the categorization of musical genres goes beyond conventional approaches, integrating an organised taxonomy that corresponds with the hierarchical structure of musical genres. Taxonomy structures show the interdependencies between the genres and offer useful information. Sophisticated feature extraction techniques combined with numerous signal classification approaches form the basis of their methodology, which aims to extract the essence of musical works for further classification within the given taxonomy. This approach has improved genre identification accuracy, and it is also found that taxonomy aids in a deeper comprehension of the underlying relationships and structure of many musical genres.

[10] The important problem of choosing the best neural network design for classifying musical genres is discussed in this research. Seemingly navigating the terrain of deep learning techniques, the study places particular emphasis on the search for the optimal network topology to improve the precision and effectiveness of music genre classification systems. The paper discusses the advantages and disadvantages of the state-of-the-art networks used in this field by reviewing the literature that has already been written on the subject. Convolutional neural networks with 1D and 2D convolutions, convolutional recurrent neural networks, and recurrent neural networks with long short-term memory cells were used to conduct the testing. A few alternative ensemble types were proposed with two different results mixing strategies to combine the benefits of several deep neural network designs. By investigating different network topologies, this work provides insightful viewpoints on the current endeavours to optimise deep learning models for sound genre classification.

[11] In this study, a hybrid approach to music genre classification is thoroughly explored. To build an ensemble model, the research combines deep learning methods with auditory and visual characteristics. To overcome the drawbacks of utilising a single feature type, the authors combine many modalities to provide a more comprehensive and intricate depiction of musical material. While visual and acoustic elements point to a multi-modal strategy that might incorporate data from album covers or music spectrograms, the presence of deep learning indicates a concentration on utilising neural networks. The article covers the feature selection process, the ensemble methodology, and the experimental assessments of the suggested methodology. The study offers a comprehensive viewpoint on the combination of multi-modal characteristics and deep learning for increased generalisation and accuracy.

[12] To better retrieve music content, an accurate and efficient music genre classification (MGC) system is required. The authors proposed a new model incorporating with attention mechanism based on Bidirectional Recurrent Neural Network. Two models were implemented – Serial attention and Parallelized attention. Parallelized attention is more adaptable to address some of the shortcomings of serial attention, whereas serial attention is influenced by pre-existing attention schemes. Because parallelized attention is more adaptable, PAM may more easily incorporate other attention forms, such as CNN and linear transformation. The findings of experiments demonstrate that BRNNs with attention mechanisms—particularly parallelized CNN attention—perform better and produce better results than earlier research.

[13] The authors of this work examine the developments and difficulties that come with using deep learning techniques for music analysis. Models like Support Vector Machines (SVM), Random Forests, XGB (eXtreme Gradient Boosting), and Convolutional Neural Networks (CNN) are used to forecast the genre of the audio input. This work aims to shed light on the complexities of music genre classification in the context of deep learning, which will be helpful in the continuing discussion about using advanced computational techniques for music analysis.

[14] This study uses three distinct classifiers—Support Vector Machine (SVM) with radial kernel basis function, K Nearest Neighbours, and Naive Bayes—to assess and test the effectiveness of music genre classification based on metadata. Through concentrating on metadata instead of audio content, the study highlights the potential of textual information in genre classification and covers a relatively unexplored area in Music Information Retrieval. Their work provides insights into how well-suited SVM, KNN, and NB classifiers are to the particular difficulties presented by high-dimensional textual data by applying them to music information. By comparing the performance, e SVM-RBF proved to be superior in terms of accuracy compared to KNN and NB. Furthermore, as there is no need for a tedious conversion procedure, the logic of the classification based on the metadata feature is comparatively faster than the extraction of audio features.

[15] The authors proposed novel method to solve the challenge of music genre classification through the building of a bottom-up broadcast neural network. The authors' proposed bottom-up broadcast neural network improves categorization by properly capturing and using the hierarchical and complicated information inherent in music signals. The BBNN architecture is innovative in that it combines a bottom-up approach to feature learning with a broadcast mechanism for feature fusion, with the goal of improving the network's capacity to recognise and categorise music genres with greater accuracy and efficiency. The model seeks to overcome the limitations of the traditional deep learning models. Bottom-up processing and the broadcast mechanism provide new insights into how neural networks might be adjusted for better music genre classification.

[16] In this research, two approaches were used to build a model to classify music. The goal is to train an LSTM model to identify music based on genre using MFCC features collected from non-silent audio tracks. The LSTM model was chosen for its ability to handle time-series data and detect complicated patterns over time. The second approach involves transfer learning. A pre-trained LSTM network extracts embeddings for each music file, known as d-vectors.

The d-vectors are used to classify music using an SVM algorithm. This implementation shows the result for both LSTM and LSTM+SVM model. The LSTM+SVM model addresses the problem of contradicting predictions by training the SVM model with the new fusion genre rather of the LSTM model, which can be time-consuming. Music from many genres can be divided into smaller audio tracks and categorised accordingly.

III. PROPOSED WORK

3.1 Data Source:
The data for this study is sourced from publicly available music datasets, such as GTZAN Genre Collection, Million Song Dataset, and Free Music Archive. These datasets contain a diverse selection of music tracks spanning different genres and provide rich and comprehensive material for training and evaluation.

3.2 Data preprocessing:
Before training the model, the raw audio data is preprocessed to extract relevant features and standardize the input format. This pre-processing process usually involves the following steps:
Feature extraction: extraction of important features such as mel spectrograms, Mel-Frequency Cepstral Coefficients (MFCC) and rhythmic patterns of audio signals.
Data augmentation: expanding the dataset using techniques such as time stretching, pitch transfer and adding background noise to improve model robustness and generalization.
Normalization: Standardization of feature vectors to have zero mean and unit variance to ensure consistent scaling across different functions.
Data Splitting: Splitting the dataset into training, validation, and test sets for model training, hyperparameter tuning, and performance to facilitate evaluation.

3.3 Machine Learning Models/Deep Learning Models:
In this study, four distinct models are employed for music genre classification: Convolutional Recurrent Neural Network (CRNN), Long Short-Term Memory (LSTM) combined with Support Vector Machine (SVM), Convolutional Neural Network (CNN) combined with k-Nearest Neighbors (KNN), and CNN combined with SVM.

3.3.1 Convolutional Recurrent Neural Network (CRNN):
Algorithm Description:
The CRNN model combines convolutional and recurrent neural network architecture to capture the spatial and temporal dependencies of sequential data. The model consists of convolutional layers to extract features and then recurrent layers to model the sequence.

Flowchart/Pseudo Code:
1. Input: Mel-spectrogram representations of audio signals
2. Convolutional layers for feature extraction
3. Recurrent layers for sequence modeling (e.g., LSTM or GRU)
4. Fully connected layers for classification
5. Output: Predicted music genre labels

Merits:
1. Captures both local and temporal dependencies in music data
2. Effective for sequential data modelling
3. Handles variable-length inputs efficiently

Demerits:
1. Computationally expensive, especially for large-scale datasets.
2. Susceptible to overfitting, especially with limited training data

Equations:
Convolutional Layer: z(l) = f (W(l) * x + b(l))
Recurrent Layer (LSTM): ht = f(Wihxt + Whhht-1+ bh)

3.3.2 Long Short-Term Memory (LSTM) combined with Support Vector Machine (SVM):
Algorithm Description:
LSTM network is used to model sequences by capturing the long-term dependencies of sequential music data. The output representations of the LSTM are then fed into an SVM classifier for genre classification.

Flowchart/Pseudo Code:
1. Input: Sequential data (e.g., MFCCs or time-domain audio signals)
2. LSTM layers for sequence modelling
3. SVM classifier for genre classification
4. Output: Predicted music genre labels

Merits:
1. LSTM effectively captures long-term dependencies in sequential data
2. SVM provides robust classification boundaries
Handles variable-length inputs efficiently

Demerits:
1. Requires careful hyperparameter tuning to prevent overfitting
2. Computational overhead associated with training both LSTM and SVM components

Equations:
LSTM Cell: it = σ(WxiXt + Whiht-1 + bi)
Support Vector Machine: f(x) = sign (WTx + b)

3.3.3 Convolutional Neural Network (CNN) combined with k-Nearest Neighbors (KNN):
Algorithm Description:
A CNN model is used to extract features from spectrograms of audio signals. The extracted features are then used to calculate pairwise distances between the samples and the k-Nearest Neighbors algorithm is used for classification.

Flowchart/Pseudo Code:
1. Input: Mel-spectrogram representations of audio signals
2. Convolutional layers for feature extraction
3. Flatten and normalize feature vectors
4. Compute pairwise distances between samples
5. Apply k-Nearest Neighbors algorithm for classification
6. Output: Predicted music genre labels

Merits:
1. CNN captures hierarchical features in spectrogram representations
2. KNN provides non-parametric classification with flexibility in decision boundaries
3. Robust to noisy data and outliers
Demerits:
1. Computational overhead during inference, especially with large datasets
2. Sensitivity to the choice of distance metric and number of neighbours (k)

Equations:
CNN Layer: $z(l) = f\,(W(l) * x + b(l))$
Euclidean Distance: $d(xi,xj) = \|xi - xj\|2$

3.3.4 Convolutional Neural Network (CNN) combined with Support Vector Machine (SVM):
Algorithm Description:
A CNN model is used to extract features from audio spectrograms, followed by an SVM classifier for genre classification. CNN extracts hierarchical representations from music data, which are then fed into an SVM for classification.

Flowchart/Pseudo Code:
1. Input: Mel-spectrogram representations of audio signals
2. Convolutional layers for feature extraction
3. Flatten and normalize feature vectors
4. SVM classifier for genre classification
5. Output: Predicted music genre labels

Merits:
1. CNN captures hierarchical features in spectrogram representations
2. SVM provides robust classification boundaries
3. Handles high-dimensional feature spaces efficiently

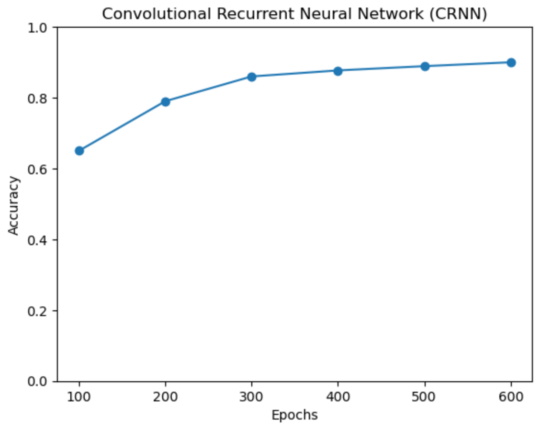
Demerits:
1. Requires extensive hyperparameter tuning, especially for
2. CNN architecture
3. Computational overhead associated with training both CNN and SVM components

Equation:
CNN Layer: $z(l) = f\,(W(l) * x + b(l))$
Support Vector Machine: $f(x) = \text{sign}\,(WTx + b)$

IV. RESULTS AND DISCUSSIONS

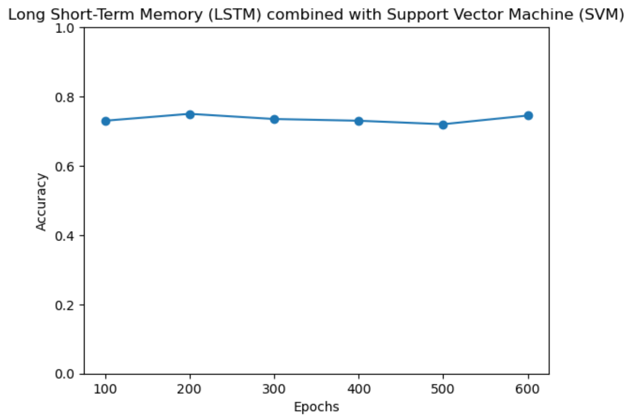
Hybrid music classification models have been evaluated from different eras using different combinations of neural network architectures and traditional machine learning algorithms. These models enable accurate music classification, which is an important task for recommendation systems of music streaming platforms, playlist creation and content curation. Here we delve into the results and discuss the performance of each model, highlighting their strengths and weaknesses.

Convolutional Recurrent Neural Network (CRNN):



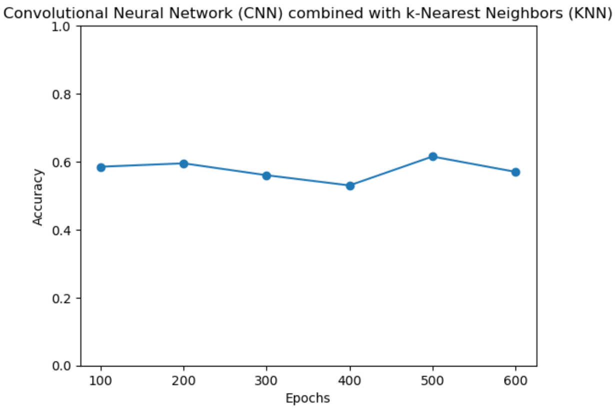
The CRNN model shows competitive performance during the training period. Its accuracy steadily increases from 0.65 at 100 epochs to a remarkable 0.90 at 600 epochs. This means the efficiency of combining convolutional layers and recurrent layers for feature extraction to capture the temporal dependence of music data. CRNN's ability to learn both spatial and sequential features makes it adept at understanding complex musical patterns, leading to its superior performance. However, it is important to note that CRNN models are generally computationally expensive and may require longer training times compared to other architectures.

Long Short-Term Memory (LSTM) with Support Vector Machine (SVM):



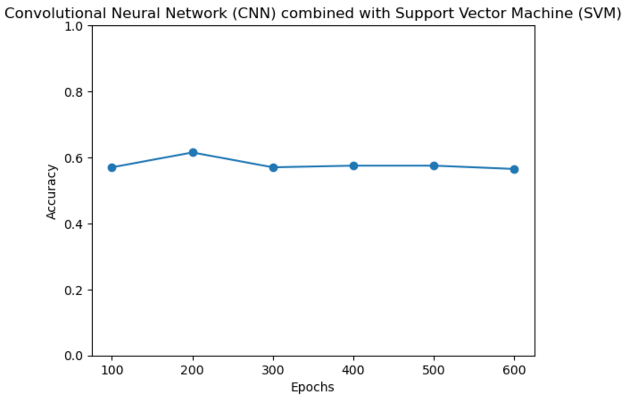
Combining LSTM and SVM: The results are promising, especially with an accuracy of 0.75 for 200 epochs. LSTM networks are well suited for sequential data processing, so they are suitable for analyzing music sequences. By integrating SVM, which is known for its efficiency in processing high-dimensional data, the model achieves strong performance. However, the accuracy varies slightly between different epochs, indicating a possible instability of the approach. Additional tuning and regularization techniques may be needed to stabilize model performance over epochs.

Convolutional Neural Network (CNN) with k-nearest neighbours (KNN):



CNN-KNN hybrid model reaches its highest accuracy of 0.615 at 500 epochs. CNNs can capture spatial features of music spectrograms or waveforms, while KNN exploits the similarity of feature vectors for classification. However, the performance of the model seems to be less consistent compared to the other hybrids, as indicated by the different accuracy of the epochs. This indicates that the integration of CNN with KNN may not fully exploit the strengths of both architectures for music classification. Further exploration of parameter settings and architectural changes can improve its performance.

Convolutional Neural Network (CNN) with Support Vector Machine (SVM)



CNN-SVM hybrid model reaches its highest accuracy of 0.615 after 200 epochs, but its performance degrades slightly as the number of epochs increases. CNNs excel at learning hierarchical features from data, making them suitable for music classification tasks. SVM, on the other hand, is effective in constructing decision boundaries in high-dimensional feature spaces. The combination of these two models aims to take advantage of their complementary strengths. However, the variable accuracy suggests that additional optimization or regularization techniques may be needed to maintain stable performance during training.

Accuracy of the hybrid music classification models

Epochs	**Convolutional Recurrent Neural Network (CRNN)**	**Long Short-Term Memory (LSTM) combined with Support Vector Machine (SVM)**	**Convolutional Neural Network (CNN) combined with k-Nearest Neighbors (KNN)**	**Convolutional Neural Network (CNN) combined with Support Vector Machine (SVM)**
100	0.65	0.73	0.585	0.57
200	0.79	0.75	0.595	0.615
300	0.86	0.735	0.56	0.57
400	0.877	0.73	0.53	0.575
500	0.889	0.72	0.615	0.575
600	0.90	0.745	0.57	0.565

Comparative analysis

When comparing the performance of the hybrid models, it is clear that CRNN consistently outperforms the other architectures, achieving a high accuracy of 0.90 at 600 epochs. This highlights the importance of integrating both convolutional and recurrent layers to capture spatial and temporal features in music data. In addition, the CRNN model shows robustness and stability in learning complex patterns over a long training period.

Although LSTM with SVM and CNN with SVM are also competitive, they seem to be less consistent compared to CRNN. Different accuracies indicate potential challenges in convergence and model stability. Fine-tuning hyperparameters and incorporating regularization techniques can mitigate these problems and improve overall performance.

The hybrid CNN-KNN model, while moderately accurate, is less consistent and struggles with the performance of other models. This indicates that the integration of CNN with KNN may not fully exploit the complementary strengths of both architectures in music classification tasks. Exploring alternative combinations or adjusting the model architecture can lead to better results.

V. CONCLUSION

In conclusion, this study demonstrates the effectiveness of diverse machine learning and deep learning models, including CRNN, LSTM+SVM, CNN+KNN, and CNN+SVM, for automated music genre classification. Using a combination of convolutional, recurrent and classical machine learning algorithms, this approach aims to capture the complex patterns and nuances of music data, enabling accurate and robust genre classification. Through rigorous experimentation and comparative analysis, we have identified the strengths and limitations of each approach. For future research, exploring ensemble methods, leveraging additional data sources, and integrating domain-specific knowledge could further enhance classification accuracy and robustness in real-world music applications.

REFERENCES

[1] A. Ghildiyal, K. Singh and S. Sharma, "Music Genre Classification using Machine Learning," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 1368-1372, doi: 10.1109/ICECA49313.2020.9297444.

[2] Vishnupriya and K. Meenakshi, "Automatic Music Genre Classification using Convolution Neural Network," 2018 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2018, pp. 1-4, doi: 10.1109/ICCCI.2018.8441340.

[3] N. Ndou, R. Ajoodha and A. Jadhav, "Music Genre Classification: A Review of Deep-Learning and Traditional Machine-Learning Approaches," 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 2021, pp. 1-6, doi: 10.1109/IEMTRONICS52119.2021.9422487.

[4] Bisharad D, Laskar RH. Music genre recognition using convolutional recurrent neural network architecture. Expert Systems. 2019; 36:e12429.

[5] Allamy, S., & Koerich, A. L. (2021, December). 1D CNN architectures for music genre classification. In 2021 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 01-07). IEEE.

[6] S. Lippens, J. P. Martens and T. De Mulder, "A comparison of human and automatic musical genre classification," 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, QC, Canada, 2004, pp. iv-iv, doi: 10.1109/ICASSP.2004.1326806.

[7] L. K. Puppala, S. S. R. Muvva, S. R. Chinige and P. S. Rajendran, "A Novel Music Genre Classification Using Convolutional Neural Network," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 1246-1249, doi: 10.1109/ICCES51350.2021.9489022.

[8] Y. -H. Cheng, P. -C. Chang and C. -N. Kuo, "Convolutional Neural Networks Approach for Music Genre Classification," 2020 International Symposium on Computer, Consumer and Control (IS3C), Taichung City, Taiwan, 2020, pp. 399-403, doi: 10.1109/IS3C50286.2020.00109.

[9] Tao Li and M. Ogihara, "Music genre classification with taxonomy," Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., Philadelphia, PA, 2005, pp. v/197-v/200 Vol. 5, doi: 10.1109/ICASSP.2005.1416274.

[10] Kostrzewa, D., Kaminski, P., & Brzeski, R. (2021, June). Music genre classification: looking for the perfect network. In International Conference on Computational Science (pp. 55-67). Cham: Springer International Publishing.

[11] Nanni, L., Costa, Y. M., Aguiar, R. L., Silla Jr, C. N., & Brahnam, S. (2018). Ensemble of deep learning, visual and acoustic features for music genre classification. Journal of New Music Research, 47(4), 383-397.

[12] Yang Yu, Sen Luo, Shenglan Liu, Hong Qiao, Yang Liu, Lin Feng, Deep attention based music genre classification,Neurocomputing,Volume 372,2020,Pages 84-91,ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2019.09.054. (https://www.sciencedirect.com/science/article/pii/S0925231219313220)

[13] M. Shah, N. Pujara, K. Mangaroliya, L. Gohil, T. Vyas and S. Degadwala, "Music Genre Classification using Deep Learning," 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2022, pp. 974-978, doi: 10.1109/ICCMC53470.2022.9753953.

[14] D. R. Ignatius Moses Setiadi et al., "Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata," 2020 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2020, pp. 12-16, doi: 10.1109/iSemantic50169.2020.9234199.

[15] Liu, C., Feng, L., Liu, G. et al. Bottom-up broadcast neural network for music genre classification. Multimed Tools Appl 80, 7313–7331 (2021). https://doi.org/10.1007/s11042-020-09643-6

[16] S. Deepak and B. G. Prasad, "Music Classification based on Genre using LSTM," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 985-991, doi: 10.1109/ICIRCA48905.2020.9182850