ISSN: 2776-0960 Volume 3, Issue 3 Mar., 2022

MUSIC GENRE CLASSIFICATION USING MACHINE AND DEEP LEARNING TECHNIQUES: A REVIEW

Peace Busola Falola
Department of Computer Science,
Faculty of Science, University of Ibadan, Ibadan, Nigeria
Email: pfalola8545@stu.ui.edu.ng

Emmanuel Oluwadunsin Alabi
Department of Computer Science,
Faculty of Science, University of Ibadan, Ibadan, Nigeria

Folashade Titilope Ogunajo
Department of Computer Science,
Faculty of Science, University of Ibadan, Ibadan, Nigeria

Oluwakemi Dunsin Fasae
Department of Computer Science,
Faculty of Science, University of Ibadan, Ibadan, Nigeria

Abstract

The ability to accurately classify music into their genres is the aim of a music genre classification system. This is broadly achieved by extracting good features from the songs and employing a good classification system. Various song libraries which are the datasets, input formats which are the features extracted from the datasets, machine learning techniques, deep learning techniques which have been employed in existing works have all delivered various degree of success. This paper therefore aims to review some of the machine learning and deep learning techniques that have been developed to accurately classify music into their genres. An evaluation of the different techniques reviewed in this paper in the context of their demonstrated results for music genre classification will also be examined.

Introduction

The easy access to internet in recent times has resulted in the rapid availability of large amount of music on various online music streaming platforms. The easy retrieval, organization and structuring of music online in other to improve user's

experience has been a major interest in the Music Information retrieval (MIR) community. One of the prominent and efficient ways to structure and easily retrieve music automatically is by genre classification [1].

For an automatic genre classification system, three steps are usually involved: 1) features such as timbre, spectro-temporal and statistical features are extracted from the audio signal [1]. Other features such as name of artist, cover album and many more are also extracted as features. 2) some techniques are then applied to select meaningful subset of the features [1, 2] or aggregate features [1,2,3] to improve the classification accuracy. This is mostly termed pre-processing. 3) a classifier based on machine or deep learning methods is then trained over the selected features from (2) to classify the input music automatically into their various genres.

Tzanetakis and Cook [4] were the first to introduce music genre recognition as a pattern recognition task. Since then, there have been applications of machine learning and deep learning techniques to music genre classification with different features being selected. The purpose of this paper is therefore to review existing systems for automatic music genre classification in respect to the selected features and classifier chosen to achieve an excellent classification result. The summary of the performance of the articles reviewed was also presented in Table 1.

Existing Music Genre Classification Systems with Machine and Deep Learning Techniques

Elbir and Aydin [5] implemented a music genre classifier and recommendation system based on signal processing and a CNN model named MusicRecNet as against existing systems that were only classification systems. The system built was also capable of checking plagiarism of songs.

MusicRecNet had three layers and each layer had a two-dimensional convolution, an activation function and a two-dimensional maximum pooling operation and a dropout operation. The dataset was GTZAN which contained 1000 songs and ten genres. Each genre had 100 different samples of songs of 25s duration each. Each song was further divided into six parts with a duration of 5s to make 6000 songs. Melspectrogram was generated from each of the 5s duration music and was saved as an image. The images generated served as the input and were applied to the MusicRecNet for training. After the training, the model was used for genre classification. Also, the dense 2 layer of the MusicRecNet was used as a feature vector of the test music samples which were further fed into various classifiers

such as MLP, Logistic regression, random forest, LDA, KNN and SVM for music genre classification, music similarity and music recommendation.

The GTZAN dataset used consisted of ten different genres, therefore, accuracy was the main performance metric that was used. Also, the average percentage of music similarity was also used as a metric for the quality of music recommendation. MusicRecNet as a standalone classifier gave a mean accuracy of 81.8% which performed better than the results of other studies. For the application of the dense 2 layer of MusicRecNet as feature vector to the test music samples, MusicRecNet with Support Vector Machine (SVM) gave an accuracy of 97.6% for the music genre classification, music similarity and music recommendation. The proposed MusicRecNet model showed improved performance in terms of music genre classification, music similarity and music recommendation as compared to previous studies.

Ghosal and Sarkar [6] proposed a novel approach for automatic music-genre classification system using a deep learning model with GTZAN as the dataset. The model leveraged on Convolutional Neural Nets (CNN) to extract spectrogram and the output was fed into Long Short-Term Memory (LSTM) sequence to sequence Autoencoders which collected key information about the temporal properties of the input sequence in its hidden state. The final hidden state of the LSTMencoder (LSTMenc) was then passed through some layers, the output of which was used to initialize the hidden state of the LSTM decoder (LSTMdec). The function of the LSTMdec was to reconstruct the input sequence based on the information contained in its initial hidden state. The network was trained to minimize the root mean squared error between the input sequence and the reconstruction. After a complete training, the activation of the fully connected encoded layer was used as representations of the audio sequence and was fed as input to Clustering Augmented Learning Method Classifier (CALM). Clustering Augmented Learning Method Classifier was based on the concept of simultaneous heterogenous clustering and classification to learn deep feature representations of the features obtained from LSTM autoencoder. To implement the Clustering Augmented Learning Method, a proposed approach of input augmentation and cluster centers were considered. For the input augmentation, a matrix of input data D and a set of cluster centers C was considered. C was kept at 10 since the dataset had 10 genres. Clustering was used to augment the input data for better learning. New set of features were added to the input example whether they belonged to the cluster or not to augment the input data. To distinguish input examples, an

additional index was added. A one-hot representation was then used to determine the cluster of the input examples. To determine the cluster centers, CALM consisted of a clustering model and a Feed-Forward Neural Net (FNN) having a softmax output to classify the music genres. For the clustering model, Random Forest classifier was used to determine the cluster centers. It was proposed that the input sample with the lowest error in predicting its cluster label is considered the center of that cluster in the subsequent iteration of the proposed approach. Therefore, the center would be the input sample which is the most fitting representative of that cluster. The clustering process would therefore aggregate the data having similar characteristics resulting in better learning by the FNN classification model.

There was a distance or dissimilarity measure between input examples and cluster centers. The clustering problem aimed to assign each input example to a cluster such that the total distance between the elements of a cluster and its center is minimized. A novel dissimilarity measure based on the weights of the trained FNN classifier was proposed which used the average of weights linked to each neuron of the input layer. The distance measured computes the distance between two examples based on how important the contribution of each feature to the resulting prediction. The resulting clusters therefore contain examples with similar potential to improve the classification results.

The performance of the model was evaluated using precision, recall and accuracy. Confusion matrix was plotted and the proposed model classified 80% of rock audio as rock correctly and labelled others mainly as country or blues. It incorrectly classified some country and a small fraction of blues and reggae as rock music.

Four traditional classification models were trained on the dataset as a baseline classifiers which were k-nearest neighbours, logistic regression, random forest, multilayer perceptrons and linear support vector machine, using Mel Frequency Cepstral Coefficients (MFCC) by flattening them into a 1-D array. Also, the features obtained from Convoutional Net and LSTM Autoencoder was stacked on a Logistic Regression classifier to test the performance of the CALM classifier.

CALM outperformed all the models with an accuracy of 95.4%.

Pelchat and Gelowitz [7] worked on 1880 songs with seven genres as the dataset. The duration of each song was three minutes. The dataset was processed by transforming the stereo channels into one mono channel, and SoundXchange command-line music application utility was used to convert the music data into a

spectrogram. The songs were further divided into 2.56 seconds to approximately make 132,000 labelled spectrogram snippets which were the inputs into the classifier. Convolutional neural network was the neural network used for classification with Rectified Linear Unit (ReLu) activation function which gave a test accuracy of 67%.

Chillara, Kavitha, Neginhal, Haldia and Vidyullatha [8] worked on finding a better machine learning algorithm than the pre-existing models that predicts the genre of songs with the Free Music Archive (FMA) dataset

Few of the models were trained on the mel-spectrograms (2D representation of a signal) of the songs along with their audio features and few others were trained solely on the spectrograms of the songs. The audio features were extracted using a python library called librosa. The models trained with spectrogram alone were Convolutional Neural Network (CNN), Convolutional Recurrent Neural Network (CRNN) and Convolutional Neural Network plus Recurrent Neural Network (CNN-RNN) models. The feature based models that were trained with audio features also called low-level features are Logistic Regression and Simple Artificial Neural Network. The CNN model trained with only spectrogram gave the highest accuracy of 88.54%. They concluded that image based classification is better than feature based classification.

Buhuleyan [9] compared the performances of two classes of models. The model was evaluated on the Audio Set dataset. The first class of model which was a CNN (Convolutional Neural Model) model was trained to predict the genre label of an audio signal using spectrogram only. The second class of model utilized hand-crafted features both from the time and frequency domains. An ensemble combining the two classes of model was also employed. A CNN architecture known as VGG-16 was the first class of model used which was implemented in two ways. The VGG-16 was downloaded with pre-trained weights and the conv base was extracted. The output of the conv base was then sent to the new feed-forward neural network which in turn predicted the genre of the music. The two VGG-16 CNN implementations were: Transfer learning and fine tuning. For transfer learning, the weights in the conv base were kept fixed but the weights in the feed forward network were allowed to be tuned to predict the correct genre label. While for the fine tuning, the pre-trained weights of the VGG-16 was started but the model weights was tuned during training of the process. A baseline feed-forward neural network was also trained with spectrogram image as the input. A simple 2-layer neural network was trained to predict the genre of the audio signal.

ResearchJet Journal of Analysis and Inventions
https://researchjet.academiascience.org

The second class of model used time and frequency domain features as the inputs. These features were extracted from the raw audio files. These features were inputs into various classifiers for the music genre prediction. The classifiers used for the feature based model were Logistic Regression, Random Forest, Extreme Gradient Boosting (XGB) and Support Vector Machines. The third class of model which was an ensemble was a combination of VGG-16 CNN and XGB (Extreme Gradient Boosting).The evaluation metrics were accuracy, f-score and AUC (Area Under Curve). The VGG-16 CNN Fine Tuning model which used only spectrogram to predict the music genre turned out with the highest accuracy score of 64%, f-score of 61% and AUC of 88.9%. The highest accuracy gotten from the feature based engineering models was 59%, f-score of 55% and AUC of 86.5%. The ensemble classifiers gave an accuracy of 65%, f-score of 62% and AUC of 89.4%.

Vishnupriya and Meenakshi [10] worked on classifying Million Song Dataset (MSD) into different genres. The dataset had 1000 songs with 10 genres. There are two important stages in classification: feature extraction and classification [10]. Feature extracted is crucial to the efficiency of the classifier [11]. The feature vector extraction was done using the librosa package in python. The package is specifically used for audio analysis. The extracted feature vector was Mel-frequency Coefficient (MFCC). MFCC encode the timbral properties of the music signal by encoding the rough shape of the log-power spectrum on the Mel-frequency scale. Two types of feature vectors was obtained: Mel Spectrum with128 coefficients and another is MFCC with 13 coefficients. The feature vectors obtained was stored into a database. The database consisted of the MFCC obtained with 10 array size for the genre. The size for the two types of feature vector extracted were 599x128x2 for Mel Spec and 599x13x5 for MFCC. The data was then shuffled for a good form of generalization before it was fed into the neural network. 800 song features were taken for the training of the model while the remaining 200 were for testing. After training the CNN model, the learning accuracy for Mel Spec feature vector and MFCC feature vector were 76% and 47% respectively. Another insight in the research was that MFCC took less time for converging whereas Mel Spec was more time consuming for learning. It was concluded that the methodology employed is promising for the classification of huge database of songs into their respective genres.

Tang, Chui, Yu, Zhiliang and Wong [12] examined the application of Long Short Term Memory (LSTM) model in music genre classification. The GTZAN dataset with ten music genre was used for this research which was pre-processed before

ResearchJet Journal of Analysis and Inventions
https://researchjet.academiascience.org

they could be used as input into LSTM model. The MFCC features were extracted from the songs using librosa, a python library. The dataset contained 420 audio tracks of 30 seconds each. Each 30 second track had 1293 frames and 13 MFCC features. Two different approaches were carried out to classify music into their genre. The first method was one single LSTM which was directly used to classify 6 different genres of music. The dataset contained 420 audio tracks of 30 seconds each. 120 was set for validation and 60 for testing. The batch size which defined the number of samples to be propagated through the network was set to 35. At 20 epochs, the test accuracy reached the maximum and the loss was minimized. A classification accuracy of 50% to 60% was achieved. The major limitation was the small data size which led to low accuracy and overfitting. Some genres such as metals were outstanding and easy to be recognized but some other genres were quite similar.

The second method was a hierarchical approach for 10-genre classification. A divide-and-conquer scheme was employed. A multi-step classifier involving 7 LSTM classifiers was used to achieve the 10-genre classification. The training sample was 120 and testing was 60 as in the first method. The 7 LSTM classifiers are as below:

- LSTM1: It classifies music into strong (hiphop, metal, pop, rock and reggae) and mild (jazz, disco, country, classic and blues) group.
- LSTM2a: It divides the music into Sub-strong1 (hiphop, metal and rock) and Sub-strong2 (pop and reggae) classes. During training, only music samples of hiphop, metal, rock, pop and reggae are involved.
- LSTM2b: It categorizes music into Sub-mild1(disco and country) and Sub-mild2 (jazz, classic and blues) groups. We used samples only from disco, country, jazz, classic and blues for training.
- LSTM3a: It classifies music into hiphop, metal and rock. Only music from hiphop, metal and rock classare involved.
- LSTM3b: It differentiates pop music from reggae
- LSTM3c: It differentiates disco music from country.
- LSTM3d: It recognizes jazz, classic and blues.

In the testing stage, the input music is first classified by LSTM1to determine if the song is strong or mild. The result from LSTM1 determines whether LSTM2a or LSTM2b will be applied. Finally, LSTM3a, LSTM3b, LSTM3c, LSTM3d was used to classify the music into the 10 genres according to the result obtained from the

previous level. The accuracy obtained for the multi-step classifier was 50% which was better than Convolutional Neural Network which gave an accuracy of 46.87% in Mlachmish [13] as discussed in the literature.

Oramas, Barbieri, Nieto and Serra [14,15] in other to overcome the limited way of classifying music items into broad genres using handcrafted audio features and assigning of a single label per item, worked on classifying musical items into the genres using three different modalities: audio, text and images. A new large-scale multimodal dataset called MuMu which contains information of roughly 31k albums classified into one or more 250 genre classes were used for this research. Each album contained cover image, text reviews and audio tracks and therefore the focus was mainly on album classification. The first modality was audio-based approach. Every music album contained series of audio tracks which are associated with different genres. To learn the genre of the album, three steps were involved. The steps were: (i) Feature vectors were tracked while trying to predict the genre labels of the album from every track in a deep neural network. (ii) The average of the tracked vector was gotten to obtain the album feature vectors. (iii) Album genres were furthered predicted from the album feature vectors in a shallow network where the input layer is directly connected to the output layer. CNN was used to learn the higher-level features from spectrograms.15 seconds long patch from each track was sampled to avoid the variability of the length of the songs which served as the fixed-sized input to the CNN. CNN with four convolutional layers and experiment with different number of filter, filter sizes and output configurations were used to learn the genre labels. The output configuration was LOGISTIC output also denoted as TIMBRE-MLP and COSINE output. A traditional approach based on audio descriptors presented from Million Songs Dataset (MSD) was used to classify the genre. CNN applied over audio spectrogram outperformed the traditional approaches based on handcrafted features. The TIMBRE-MLP approach achieved 0.792 of AUC as compared to 0.888 from the best CNN approach. It was observed that the COSINE regression approach achieved better AUC scores in most configurations and results are more diverse in terms of catalog coverage. The second modality was text-based approach. The albums in the dataset contained a variable number of customer reviews. All reviews from the same album are combined into a single text. The combined result was truncated at 1000 characters to balance the amount of text per album. A Vector Space Model approach (VSM) with tf-idf weighting which performs better when dealing with large texts than CNN was used to create a

feature vector for each album. The vocabulary size was reduced to 10k to have a good balance of network complexity and accuracy. A second approach to semantically enrich the album texts using Babelfy (a state-of –the-art tool for entity linking) [16] was proposed. Babelfy associated a given textual fragment candidate to the most suitable entry in a reference knowledge base (KB). Babelfy maps words from a given text to Wikipedia [16]. In Wikipedia, categories are used to organize resources. All Wikipedia categories of entities identified by Babelfy in each document are added to the end of the text as new words. Then the VSM with tf-idf weighting was applied to semantically enrich texts and the vocabulary is also limited to 10k terms. Either word or categories were part of the vocabulary. From this representation, a feed forward network with two dense layers of 2048 neurons and a Rectified Linear Unit (ReLU) after each layer was trained to predict the genre labels using LOGOSTIC and COSINE configurations. The result for this approach showed that the semantic enrichment of texts yields better results in terms of AUC and diversity. The COSINE configuration slightly outperformed LOGISTIC in terms of AUC and greatly in terms of catalog coverage. The information gain of words in the different genres was also studied. The text-based results are slightly superior to the audio-based ones.

The third modality was image-based approach. The albums in the dataset had an associated cover art image. Deep Residual Networks (ResNets) was used for music genre classification from the images. ResNet is a common feed-forward CNN with residual learning which bypasses two or more convolution layers. A slightly modified version of the original ResNet was employed with scaling and aspect ratio augmentation obtained from [17]. The photometric distortions from [18] and weight decay were applied to all weights and biases. The network used composed of 101 layers (ResNet-101), initialized with pretrained parameters learned on ImageNet dataset. The final layer in the ResNet implementation had a logistic regression final layer with sigmoid activations and uses the binary cross entropy loss.

Genre classification for this approach had lower performance in terms of AUC and catalog coverage compared to other modalities. The COSINE configuration couldn't be used for this modality due to the use of an already pre-trained network with LOGISTIC configuration.

The fourth modality approach was the multimodal approach. All the three modalities were combined into a single model. An internal feature representation for every album after training them on genre classification task was obtained. The

last fully connected layer of each network became the feature vector for each respective modality. L2-regularization was applied to each of them and then concatenated into a single feature vector, which becomes the input to a simple Multi Layer Perceptron (MLP). The input layer is directly connected to the output layer. The output layer either had a LOGISTIC or COSINE configuration.

The result obtained showed that the combination of modalities outperformed the single modality approaches. Multimodal approaches that included text features improved the results. COSINE approaches had similar AUC than LOGISTIC approaches but a much better catalog coverage which was due to the spatial properties of the factor space.

Zhang, Lei ,Xu and Xing [1] worked on improving music genre classification with convolutional neural network. Improving the music genre classification was achieved in two different ways. The first method was by combining the max-pooling and average-pooling layers of the convolutional neural network architecture used in this work(called nnet1) in other to provide more statistical information to higher level neural networks. The convolutional neural network contained 10 layers including the input layer and the softmax output layer. The Rectified linear units (ReLUs) were used as the activation function in all convolutional and dense layers of the architecture except for the top layer where the softmax function was applied. The second way was using shortcut connections to skip one or more layers in the Convolutional Neural Network (CNN). The first convolutional layer was connected to the output of the third convolutional layer. This method was inspired by the residual learning method [19]. The GTZAN dataset was used. 3 seconds was extracted from every song in the dataset with 50% overlap. This showed an improvement in the classification accuracy. The Short Time Fourier Transform (STFT) magnitude spectrum (Short Time Fourier Transform Spectrogram) of each song served as the input into the architecture for the two methods. The Short Time Fourier Transform (STFT) usually represents the timbre texture of music. For the two methods used, Adadelta [20] was used as the optimizer with default learning 1.0. Categorical cross-entropy was the loss function. The dropout technique with 0.2 dropout rate was used to alleviate the overfitting problem.

For the second method; the residual method called nnet2, the output of the first convolutional layer had 256 feature maps and each map of the output of the third convolutional layer was a 119 dimensional vector. Before the wise adding operation component, that is before adding the first convolutional neural network

to the residual network (the shortcut), zero padding was used to make sure that the two vectors are of the same dimension. Mini batches of 50 samples were used and were shuffled after each epoch. The evaluation metric was accuracy. The nnet1 with the combination of max and average pooling gave an accuracy of 84.8%. Also, the neural network was used with max pooling and average pooling separately without combining them. The nnet1 with max pooling gave an accuracy of 79.9% while the nnet1 with average pooling gave an accuracy of 84.4% and the combination of the max and average pooling layers gave an accuracy of 84.8%

The nnet2 with the combination of max and average pooling gave an accuracy of 87.4%. Nnet2 with max pooling alone gave an accuracy of 85% while the nnet2 with average pooling gave an accuracy of 87.4%.

The result of nnet1 and nnet2 was compared to some existing works.

In [21], the features were mel-spectrum, SFM, SCF and an accuracy of 83.9% was obtained. In [22], the features were FFT (aggregation) and an accuracy of 83% was obtained. In Multilayer invariant representation [23], the features were STFT with log representation and an accuracy of 82% was achieved. The nnet2 result outperformed the listed previous results.

Falola and Akinola [24] worked on improving music genre classification with content based features and 1D Convolutional Neural Network model for an excellent accuracy. As a result of much focus on spectrogram feature for music genre classification over the years, the paper focused on another type of feature embedded in songs for classification. 1D Convolutional Neural Network (1D CNN) was considered as the classifier because of its benefits. Generally, 1D CNN works well for analysis of a time-series of sensor data, analysis of signal data over a fixed-length period such as audio recording and also for natural language processing [23, 24].

A new dataset consisting of Nigerian traditional songs was used for the research. A duration of 30s was extracted from each of the songs, after which seven content based features were extracted from the songs which served as input into the classifier. The dataset was further pre-processed using feature scaling and feature encoding methods for a better classification. After the pre-processing of the data, it was served into the 1D Convolutional Neural Network using different hyper parameters for an excellent classification result. Evaluation metrics used were accuracy, precision, recall and f1-score. 1D CNN gave an excellent accuracy result of 92.5%, precision score of 92.7%, recall of 92.5% and f1 score of 92.5%.

ResearchJet Journal of Analysis and Inventions
https://reserchjet.academiascience.org

Conclusion

With various features and different machine learning and deep learning techniques being used for genre classification, it's been observed that deep learning techniques are excellent for genre classification as compared to machine learning techniques. Ensemble method which involves combining two or more classifiers delivered excellent results for genre classification as observed from some of the works reviewed.

As features extracted play a vital role in classification, audio features such as spectrogram feature which are basically extracted from the signal of the music is one of the best features that gave excellent results. Content based features have also contributed to an excellent accuracy result when used with an appropriate classifier. Combination of two or more features as input into the classifiers in some of the works was also a major contribution to getting excellent classification results. More work still needs to be done on text based features such as name of artist, year of production, name of producer and image based features such as cover art image or album image for accurate music genre classification.

Generally, Convolutional Neural Network (CNN), a deep learning technique has contributed massively to excellent music genre classification. Other machine and deep learning models can still be worked upon for accurate music genre classification.

Table 1 Performance of Some Reviewed Articles of Music Genre Classification

Author(s)	Dataset	Features Extracted	Classifier/ Method	Success Rate			
				Accuracy	F-Score	AUC	
Elbir and Aydin	GTZAN	Melspectrogram	CNN (MusicRecNet)	81.8%			
			MusicRecNet + SVM	97.6%			
Ghosal and Sarkar	GTZAN	Spectrogram	LSTM + CALM	95.4%			
Pelchat and Gelowitz	Not indicated	Spectrogram	CNN	67%			
Chillara, Kavitha, Neginhal, Haldia and Vidyullatha	FMA	Spectrogram	CNN	88.54%			
			CRNN	53.5%			
			CNN-RNN	56.4%			
		Time and frequency domain features	Logistic Regression	60.892%			
			Simple Artificial neural Network	64.0625%			
Buhuleyan	Audio set	Spectrogram	VGG-16 CNN Transfer Learning	63%	61%	89.1%	
			VGG-16 CNN Fine Tuing	64%	61%	88.9%	
			Feed-forward NN baseline	43%	33%	75.9%	
		Time and frequency domain	Logistic Regression	53%	47%	82.2%	
			Random Forest	54%	48%	84%	
			Support Vector Machines	57%	52%	85.6%	
			Extreme Gradient Boosting	59%	55%	86.5%	
Vishnupriya and Meenakshi	MSD	Mel Spectrum	CNN	76%			
		MFCC	CNN	47%			
Tang, Chui, Yu, Zhiliang and Wong	GTZAN	MFCC	One LSTM	50% to 60%			
			Seven LSTMs	50%			
Oramas, Barbieri, Nieto and Serra	MuMu	Spectrogram (Audio based)	CNN with Timbre MLP output configuration			93.6%	
		Customers review (Text based)	VSM +SEM			91.7%	
		Cover Art Image (Image based)	ResNet			74.3%	
Zhang, Lei ,Xu and Xing	GTZAN	Short Time Fourier Transform Spectrogram	nnet 1 (CNN); max pooling layer	79.9%			
			nnet 1 (CNN); average pooling layer	84.4%			
			nnet 1 (CNN); max + average pooling layer	84.8%			
			nnet 2 (CNN); max pooling layer	85.0%			
			nnet 2(CNN); average pooling layer	81.9%			
			nnet 2(CNN); max + average pooling layer	87.4%			
Falola and Akinola	Nigerian Traditional Songs	Content based features	1D CNN	92.5%	92.5%	Precision: 92.7%	Recall: 92.5%

References

1. Weibin Zhang, Wenkang Lei , Xiangmin Xu and Xiaofeng Xing (2016). Improved Music Genre Classification with Convolutional Neural Networks. Interspeech 2016. http://dx.doi.org/10.21437/Interspeech.2016-1236. pp. 3304-3308
2. N. Auguin, S. Huang, and P. Fung, "Identification of live or studio versions of a song via supervised learning," in Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific. IEEE, 2013, pp. 1–4.
3. J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. K´egl, "Aggregate features and adaboost for music classification," Machine learning, vol. 65, no. 2-3, pp. 473–484, 2006.
4. George Tzanetakis and Perry Cook (2002).Musical Genre Classification of Audio Signals, IEEE Transactions on speech and audio processing, VOL. 10, NO. 5, pp. 293-302
5. A. Elbir and N. Aydin (2020). Music Genre Classification and Music Recommendation By Using Deep Learning (2020). Electronics Letters. Vol. 56, No. 12, pp. 627–629
6. Soumya Suvra Ghosal and Indranil Sarkar (2020). Novel Approach to Music Genre Classification using Clustering Augmented Learning Method (CALM).Proceedings of the AAAI 2020 Spring Symposium on Combining Machine Learning and Knowledge Engineering in Practice (AAAI-MAKE 2020).Vol. 2600
7. Nikki Pelchat, Craig M Gelowitz. Neural Network Music Genre Classification. 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)
8. Snigdha Chillara, Kavitha A S, Shwetha A Neginhal, Shreya Haldia and Vidyullatha K S. (2019). Music Genre Classification using Machine Learning Algorithms: A comparison. International Research Journal of Engineering and Technology (IRJET). Volume: 06 Issue: 05, pp. 851-858
9. Hareesh Bahuleyan (2018). Music Genre Classification using Machine Learning Techniques.arXIV:1804.01149v1[cs.sd], https://www.researchgate.net/publication/324218667. Accessed in October 2020
10. Vishnupriya S, K.Meenakshi (2018). Automatic Music Genre Classification usingConvolution Neural Network. 2018 International Conference on

Computer Communication and Informatics (ICCCI -2017), Coimbatore, INDIA.IEEE

11. Abdulhamit Subasi (2019). Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques (pp. 193-275). Science Direct Publications
12. Chun Pui Tang, Ka Long Chui, Ying Kin Yu, Zeng Zhiliang and Kin Hong Wong (2018). Music genre classification using a hierarchical long short term memory (LSTM) model," Proc. SPIE 10828, Third International Workshop on Pattern Recognition, 108281B (26 July 2018); doi: 10.1117/12.2501763
13. Mlachmish. Music genre classification with CNN. https://github.com/mlachmish/MusicGenreClassification/blob/master/README.md. Accessed: 16- April- 2018.
14. Sergio Oramas, Francesco Barbieri, Oriol Nieto and Xavier Serra (2018). Multimodal Deep Learning for Music Genre Classification. Transactions of the International Society for Music Information Retrieval, 1(1), pp. 4-21. DOI: https://doi.org/10.5334/tismir.10
15. Sergio Oramas, Francesco Barbieri, Oriol Nieto and Xavier Serra (2017). Multi-Label Music Genre Classification From Audio, Text and Images Using Deep Features. 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017. arXiv:1707.04916v1 [cs.IR] 16 Jul 2017
16. Andrea Moro, Alessandro Raganato, and Roberto Navigli (2014). Entity Linking meets Word Sense Disambiguation: A Unified Approach. Transactions of the Association for Computational Linguistics, 2:231–244, 2014.
17. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich (2015). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9, 2015.
18. Andrew G Howard. Some improvements on deep convolutional neural network based image classification. arXiv preprint arXiv:1312.5402, 2013.
19. P. Zhang, X. Zheng, W. Zhang, S. Li, S. Qian, W. He, S. Zhang and Z. Wang (2015). “A deep neural network for modeling music,” in Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. ACM, 2015, pp. 379–386.
20. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” arXiv preprint arXiv:1512.03385, 2015.
21. M. D. Zeiler, “Adadelta: an adaptive learning rate method,” arXiv preprint arXiv:1212.5701, 2012.

22. S. Sigtia and S. Dixon (2014). "Improved music feature learning with deep neural networks", in Acoustics, Speech and Signal Processing(ICASSP), 2014 IEEE International Conference on. IEEE, 2014, pp. 6959–6963.
23. C. Zhang, G. Evangelopoulos, S. Voinea, L. Rosasco, and T. Poggio (2014) . "A deep representation for invariance and music classification," in Acoustics, Speech and Signal Processing (ICASSP), 2014. IEEE International Conference on. IEEE, 2014, pp. 6984–6988.
24. Falola, P. B., & Akinola, S. O. (2021). Music Genre Classification Using 1D Convolution Neural Network. International Journal of Human Computing Studies, 3(6), pp. 3-21. https://doi.org/10.31149/ijhcs.v3i6.2108
25. Nils (2018). Introduction to 1D Convolutional Neural Network, blog.goodaudience.com/

ResearchJet Journal of Analysis and Inventions
https://reserchjet.academiascience.org