International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-07, Issue-10, JAN 2022

Survey on Music Genre Classification using Machine Learning

[1]Yukta Padgaonkar, [2]Janhavi Gole, [3]Prof. Bhanu Tekwani

[1,2]Student, [3]Assistant Professor, Vidyalankar Institute of Technology, Mumbai, India,

[1]yukta.padgaonkar@vit.edu.in, [2]janhavi.gole@vit.edu.in, [3]bhanu.tekwani@vit.edu.in

Abstract - In today's world music plays a vital role in people's life. Music is a way for people to express themselves. Different people enjoy different types/genre of music. Certain music helps people to relax and improves their mood. Classifying music into different genres will help improving the overall listening experience of the user. Machine Learning can help to automate this tedious and time-consuming task. In this paper we shall see how several other Music Genre classification systems are already developed using machine learning that are able to predict the genre of the music file upto a certain accuracy with the help data of the dataset. We shall also learn about which machine learning algorithms, methodologies and datasets they have used for the models. Names of the methodologies include Convolutional Neural Network (CNN), Dense Neural Network (DNN), Simple Artificial Neural Network (ANN), Support Vector Machine, Random Forest, Gradient Boosting, Logistic Regression, K-Nearest Neighbours, Multilayer Perceptron, Ada Boost, Naive Bayes, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA). We will compare all these algorithms one by one considering parameters such as accuracy to see how well they are performing.

Keywords: accuracy, convolutional neural network, deep learning, machine learning algorithms, music genre classification, precision.

I. INTRODUCTION

Music has become an integral part of our lives. Music enhances people's lives, influences their mind, their expressive ability, and a whole lot of other qualities. People when they are at their happiest listen to music at the same time people when they are at their worst listen to music too. People have always found music compelling in their lives, whether it is for enjoyment they derive by listening to it, the emotional response music generates or for the pleasure of performing or creating music. People listen to different types/ genre of music depending on their mood and their music taste.

In the olden days, the only means to listen to music was radio or music CDs. But now plenty of music is available online. Many platforms are rising day by day. As the amount of music on the digital platforms goes on increasing it becomes important to classify music into different genres. Maintaining metadata like genre of the music is very difficult. Also manually classifying the music into different genres is a time-consuming task. Genre boundaries are not very rigid and keep on evolving which makes music classification even more difficult. Machine learning techniques can be used to automate the process of classifying music into different genres. Machine learning is a subset of Artificial Intelligent which uses data and algorithms to train machines to imitate the way humans learn and take better decisions. In order to train the machine, we need to extract different and appropriate features from the dataset of music files. In order to do that we need a sizeable dataset. Extracting the appropriate features to train the machine learning model is an important task and can help to improve the overall accuracy of the model. Different Machine Learning algorithms can be used and their performance can be measured using appropriate parameters.

In this paper we have examined various papers from different authors and studied the techniques and methods they have used.

II. LITERATURE SURVEY

Nirmal Vekariya, Hemang Vyas, Nirav Dedhiya in [1] proposed two approaches for music genre classification, Machine learning techniques and Deep Neural Network. They pre-process the data and then perform feature extraction. Machine learning algorithms like KNN, decision tree, Random Forest and SVM algorithms have been used. They have used Linear, Poly, and Radial Basis Function (RBF) kernels for SVM, but got the best accuracy with Linear kernel. They have also analysed the top 25 features that play a vital role in classifying the music files into different genres. In order to evaluate the models, they have used training and testing accuracy and confusion matrix to see which model performs well on which genres.

Snigdha Chillara, Kavitha A S, Shwetha A Neginhal, Shreya Haldia, Vidyullatha K in [2] explore feature-based

 DOI : 10.35291/2454-9150.2022.0030

models like Logistic Regression and Simple Artificial Neural Network and spectrogram-based models like Convolutional Neural Network (CNN), Convolutional Recurrent Neural Network (CRNN), CNN-RNN Parallel Model. The classification models are trained over Free Music Archive (FMA) dataset. They used spectrogram images for CNN model and Features extracted for Logistic regression and Artificial Neural Networks.

Rajeeva Shreedhara Bhat, Rohit B. R., Mamatha K. R. in [3] have used GTZAN Genre Collection dataset. They extracted features using the librosa module in python. All the tracks from the training dataset are pre-processed and feature vectors are extracted which form a feature vector database. A Mel Spectrum of each track is obtained. They have used the Convolutional Neural Network model which is trained using the feature vector generated. They used 80% data for training and 20% for testing.

Seethal V., Dr A Vijayakumar did a comparison between g k-nearest neighbor (k-NN) and Support Vector Machine (SVM) classifier, the dataset used is GTZAN dataset [4]. They have extracted Pitch and MFCC features from the sound records and used to train the models. KNN model obtained an accuracy of 75% whereas SVM obtained an accuracy of 65%.

Anirudh Ghildiyal, Komal Singh, Sachin Sharma have used five models namely CNN, ANN, SVM, Decision Trees and MLP [5]. They have trained and tested the models on GTZAN dataset. For CNN, they created a mel spectrogram for each audio file and they have split it into 64 images which increases the data set by 64 times, which they have used to train the CNN model. For ANN, SVM MLP and Decision Trees they used the features extracted from the audio files which were stored in csv files.

Hareesh Bahuleyan used Audio Set, which is a large-scale human annotated database of sounds, where the dataset was created by extracting 10-second sound clips from a total of 2.1 million YouTube videos [6]. To improve the Signal-to-Noise Ratio (SNR) of the signal, a pre-emphasis filter,was applied to the original audio signal. He proposed two different approaches to solve the problem, one was using the audio signal and treating it as an image. Then CNN based image classifier, VGG-16 was trained on the images to predict the music genre solely based on this spectrogram. To avoid over-fitting in CNN, two strategies namely L2-Regularization and dropout were adopted. The second approach involved extracting time and frequency domain features from the audio signals and training traditional machine learning classifiers like Logistic Regression, Random Forest, Gradient Boosting (XGB), Support Vector Machines (SVM) on the extracted features. To evaluate the performance of the model metric like accuracy, f-score and AUC were used.

Ndiatenda Ndou, Ritesh Ajoodha, Ashwini Jadhav they compared the traditional machine learning models and deep learning techniques [7]. They also study the how the traditional machine learning models perform with the 3 second duration features and the thirty second duration features. They have put the features for music genre classification into four categories which were Magnitude-based features, Tempo-based features, Pitch-based features, Chordal progression features. Information gain ranking algorithm was used to select the features that played a role in identifying getting the features that contributed the most to the correct classification. They conducted the research in three phases namely phase A, phase B and phase C. Phase A and B provides a machine learning approach. In phase A, a total of 459 features were selected while in phase B, 54 features were selected and 57 features in phase C. Phase C works with deep-learning approach and machine-learning algorithms with larger dataset but of low duration (3 seconds each). The machine learning algorithm used in this paper are k-Nearest Neighbours, Linear Logistic Regression, Multilayer Perceptron, Random Forests trees, and Support Vector Machines. From the confusion matrix that was plotted for Linear regression model, they found that Rock genre was the most misclassified genre. In phase C for deep learning technique, Convolutional Neural Network was trained on spectrograms, three-seconds features, and thirty-seconds features. They also found that implementing CNN with spectrograms provides high accuracy as the number of epochs go on increasing.

Nilesh M. Patil and Dr. Milind U. Nemade in [8] have put forward an automated system for music genre classification. They used the GTZAN dataset from MARSYAS website. They convert the audio from .au format to .wav format, then they choose five features namely MFCC, spectral centroid, zero crossing rate, Chroma frequencies, spectral roll-off to classify the music genres. They perform feature extraction on the five features to create a single feature vector of length 28. Then they train different classifiers on the feature vector. The classifiers used were K Nearest Neighbours, Linear Kernel SVM, and Polynomial Kernel SVM. They use accuracy, recall and precision to measure the performance of these classifiers.

Nikki Pelchat and Craig M Gelowitz in [9] used a music library which had 1880 music files which included seven genres. They implemented a convolutional neural network which was similar to the literature that they had surveyed with some changes. They also divided the spectrograms of the music files into 13200 spectogram slices so that each spectrogram was of 128x128 pixel, which increased the size of the dataset. They initailly trained CNN using 27 genres on a dataset with less songs but later limited the number of genres to seven and increased the size of the dataset.

Archit Rathore, Margaux Dorido in [10] have chosen five features namely Mel-Frequency Cepstral Coefficients, Spectral Centroid, Zero Crossing Rate, Chroma Frequencies, Spectral Roll-off to classify the audio clips and these features

 DOI : 10.35291/2454-9150.2022.0030

were appended to give a 28 length feature vector. They have used different multi-class classifiers and an ensemble of the classifiers to get the result. The classifiers used are K Nearest Neighbours, Linear Kernel SVM, Radial Basis Function (RBF), Kernel SVM, Polynomial Kernel SVM, Sigmoid Kernel SVM, Decision Tree, Random Forest, Ada Boost, Naives Bayes, Linear Discriminant Analysis (LDA) classifier, Quadratic Discriminant Analysis (QDA) classifier, Logistic Regression. The ensemble classifier that they used the prediction of each classifier and ran a majority voting heuristic to obtain the optimal genre label for the given input.

III. FINDINGS AND ANALYSIS

Nirmal Vekariya, Hemang Vyas, Nirav Dedhiya in [1] found that the Deep Neural Network outperforms the Machine learning algorithms with highest training and validation accuracy which was 80% and 71%. They discovered that the genres like disco and blue were difficult to classify and classical and pop were easy to classify. They have concentrated on Chroma-based features since these features acted as a convenient measure for the observation of music audio by humans.

Snigdha Chillara, Kavitha A S, Shwetha A Neginhal, Shreya Haldia, Vidyullatha K. in [2] found that in feature-based models out of Logistic Regression and Simple Artificial Neural Network, Simple Artificial Neural Network proved to be the best with a test accuracy of 64%. In image-based model, CNN model proved to be the best with an accuracy of 88.5%. In general, Image-based models performed well as compared to feature-based models was their observation.

Rajeeva Shreedhara Bhat, Rohit B. R., Mamatha K. R. achieved a training accuracy of about 98% and validation accuracy of 73% using the Convolutional Neural Network model [3]. From the difference in the training and testing accuracy it is evitable that the model is overfitted.

Seethal V., Dr A Vijayakumar in [4] found that SVM can identify only a limited set of patterns. SVM model predicted rock genre with 94% accuracy whereas for blues it was 80%. KNN model predicted blues with 66% accuracy. KNN model worked well for hiphop, it gave an accuracy of 75%, whereas SVM gave an accuracy of 55%. It can be seen that for some genres KNN works well while for the others SVM works better.

Anirudh Ghildiyal, Komal Singh, Sachin Sharma in [5] found that the CNN model gave the best results with an accuracy of 91%. For Decision tree and ANN they obtained 74.3 and 70 percent accuracies respectively. For SVM and MLP got 68.9 and 68.7 percent respectively. They also did a comparative analysis with the other models in the literature that they had reviewed. They also pointed that since the AUC value for each class that is classified is good, which means their model is able to distinguish between the different genres

Hareesh Bahuleyan in [6] discovered that the CNN based deep learning model outperformed the traditional machine learning models. Gradient boosting algorithm performs the best among the traditional machine learning models. They have also demonstrated that ensembling the CNN and XGBoost model performed better than CNN and XGB which is evident from all the metric values used to evaluate the performance. They also discovered that the features that contribute the most during prediction are Mel Frequency Cepstral Coefficients (MFCC) features. They observed that with only the top 10 features, the XGB model performance is surprisingly good. They also compared the time and frequency domain features and concluded that frequency domain features are better than time domain features when it comes to modelling audio for machine learning tasks.

Ndiatenda Ndou, Ritesh Ajoodha, Ashwini Jadhav in Phase A, found that Linear Logistic Regression provided the highest accuracy of 81% with training time of 25.25 seconds [7]. Random Forest and Support Vector Machines showed accuracy of 75.70% and 75.40% respectively. Naíve Bayes showed the lowest accuracy in this phase which was 53.20%. k-Nearest Neighbour took the least training time of 0.0100s with an accuracy of 72.80%. In Phase B, Support Vector Machine provided the highest accuracy of 80.80% while Multilayer Perceptron, Logistic Regression, Random Forests, k-Nearest Neighbour, Naíve Bayes gave an accuracy of 77.30%, 75.80%, 72.40%, 69.70% and 54.50% respectively. In Phase C, for the larger dataset of 3 second audio files, k-Nearest Neighbours gave the highest accuracy of 92.69% followed by Multilayer Perceptron with an accuracy of 81.73%. For deep learning techniques, CNN trained on 3 second feature obtained the highest accuracy of 72.40%, this was attributed to the fact that the 3 second feature dataset was larger and thus provided increased training data. This was followed by CNN trained on Spectogram with accuracy of 66.50%. CNN model trained on 30 second features had the accuracy of 53.50 which was the least among the CNN models.

Nilesh M. Patil and Dr. Milind U. Nemade in [8] found that polynomial kernel SVM performed better than K-NN and Linear Kernel SVM giving an accuracy of 78%. KNN and Linear Kernel SVM gave an accuracy of 64% and 60% respectively. Also, it can bee seen that Poly SVM kernel was able to predict most genres with good accuracy except Rock and Blues.

Nikki Pelchat and Craig M Gelowitz saw that the implementation of the initial smaller dataset with 27 genre was an overfit as the accuracy for training and testing data showed a huge difference [9]. As the number of genres were narrowed down and the size of dataset was increased it was seen that the testing accuracy improved by a good amount.

Archit Rathore, Margaux Dorido in found that any single classifier did not classify all the ten genres in the GTZAN dataset well [10]. They found that SVM with polynomial

 DOI : 10.35291/2454-9150.2022.0030

kernel worked well for most genres except blues and rock. They a classifier are very efficient for some specific genres. They used the SVM classifier with polynomial kernel to classify 6 genres they found that the accuracy was 82%. However, as the number of genres increased, the accuracy kept on decreasing.

After analyzing, we can state that the neural network model, CNN performed well and yields better results than the other ML models.

IV. CONCLUSION

This paper has reviewed some of the research work done in music genre classification using Machine Learning. The research involved various supervised machine learning algorithm and Deep learning techniques like Neural Network and Convolutional Neural Networks. For supervised machine learning algorithms, features were extracted from the audio files and for Deep learning techniques spectrogram images were created from the audio files so that they could be fed to the model. Out of the various machine learning techniques used in the different papers we can see that CNN gave a better result than the supervised machine learning algorithms used in most of the papers. KNN and SVM also performed well amongst the traditional machine learning algorithms. The precision or the correctness also depends on the size of the dataset available, hyperparameters and the features extracted. As the size of dataset increases, we can see an increase in the accuracy with which music genre is predicted.

REFERENCES

[1] Nirmal Vekariya, Hemang Vyas, Nirav Dedhiya. Music Information Retrieval And Genre Classification Using Machine Learning Techniques And Deep Learning. International Research Journal of Engineering and Technology (IRJET), Volume 07 Issue 07, July 2020.

[2] Snigdha Chillara, Kavitha A. S., Shwetha A. Neginhal, Shreya Haldia, Vidyullatha K. S., Music Genre Classification using Machine Learning Algorithms: A comparison, International Research Journal of Engineering and Technology IRJET), Volume 06 Issue 05, May 2019.

[3] Rajeeva Shreedhara Bhat, Rohit B. R., Mamatha K. R., Music Genre Classification, SSRG International Journal of Communication and Media Science (SSRG-IJCMS), Volume 7 Issue 1, April 2020.

[4] Seethal V, Dr. A. Vijayakumar, Music Genre Classification using Machine Learning, International Journal of Trend in Scientific Research and Development (IJTSRD), Volume 5 Issue 4, May-June 2021.

[5] Anirudh Ghildiyal, Komal Singh, Sachin Sharma, Music Genre Classification using Machine Learning ICECA-2020, Fourth International Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE Xplore Part Number: CFP20J88-ART; ISBN: 978-1-7281-6387-1, 2020.

[6] Hareesh Bahuleyan, Music Genre Classification using Machine Learning Techniques, University of Waterloo, 2018.

[7] Ndiatenda Ndou, Ritesh Ajoodha, Ashwini Jadhav, Music Genre Classification: A Review of Deep-Learning and Traditional Machine-Learning Approaches, IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2021.

[8] Nilesh M. Patil, Dr. Milind U., Music Genre Classification Using MFCC, K-NN and SVM Classifier, International Journal of Computer Engineering In Research Trends (IJCERT), Volume 4, Issue 2, February 2017.

[9] Nikki Pelchat, Craig M Gelowitz, Neural Network Music Genre Classification, IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), 2019.

[10] Archit Rathore, Margaux Dorido, Music Genre Classification, Indian Institute of Technology, Kanpur Department of Computer Science and Engineering, https://cse.iitk.ac.in/users/cs365/2015/_submissions/archit/report.pdf

 DOI : 10.35291/2454-9150.2022.0030