Music Genre Classification using Machine Learning

Jeffery Ho Kin Pou
School of CSE
REVA University
Bengaluru, India
r18cs164@cit.reva.edu.in

H Keshav Rao
School of CSE
REVA University
Bengaluru, India
r18cs156@cit.reva.edu.in

Geetansh Bhambhani
School of CSE
REVA University
Bengaluru, India
r18cs145@cit.reva.edu.in

Joel Joseph
School of CSE
REVA University
Bengaluru, India
r18cs165@cit.reva.edu.in

Suriya Prakash J
Asst. Prof., School of CSE
REVA University
Bengaluru, India
suriya.prakashj@reva.edu.in

***Abstract*— A 'music genre' is a classification system, that distinguishes parts of a music line or the entire music into some music form or music style. We are able to categorize music into various genres in several ways, like religious music and pop music, secular music and classical music. The amount of data available to us is increasing rapidly, making it infeasible for manual curation. In this work, we apply simple and basic machine learning algorithms namely Logistic Regression, K-Nearest Neighbor, Random Forest, Support Vector Machine and Artificial Neural Network, along with dimensionality reduction techniques namely, PCA, KPCA and LDA, on the GTZAN dataset. Further, we compared their accuracies and found that the combination model of KNN with PCA provides the highest accuracy of 77.41% among the compared models.**

Keywords*— *Music genre; Machine learning; Classification algorithms; KNN; Principal component analysis; GTZAN dataset.

I. INTRODUCTION

A genre is a crucial feature of any tune or audio that can advise or recommend users to their preferred choice of music. Most music lovers and fanatics usually create playlists established on distinct genres and forms of music, leading to probable applications such as management and playlist recommendation systems. Further, we see that the increment of music databases, nowadays which are online, and interactive applications for users opened up prospective, effective and automatic tools that are being used for classification of music, thereby becoming an essential issue.

The inescapable utilization of the World Wide Web and Internet has achieved tremendous swaps in the music business also causing a wide range of progress. Instances of these turn of events incorporate the far and wide utilization of online music tuning in, music copyright issues, grouping and classifying audio types and several other music proposals. As with CDs and DVDs and music tapes coming to an end as the main wellspring of music, the rise of web has a consequent and remarkable command over the progression of data, i.e., retrieval from the web to each and every being on the planet associated to this system. People now listen to audio and music any time they want and from wherever they want, thanks to the expansion of music broadcast platforms, and they can access thousands of songs through numerous listening platforms. The outcome is a massive collection and assortment of tunes, sound documents, stacked up indiscriminately in different files and folders, making it difficult for any person to manage and keep up with the genre of every single tune and orchestrate them similarly. Segregating and classifying sound tracks and tunes by labeling them to the fitting kind is the most legitimate approach to make in order to oversee and control sound records and audio files that are in such massive numbers.

Genre classification can be considered a way to categorize music by sighting it as to a common or similar custom or some different accord. An important note about this is that there is no straightforward definition of what a genre should sound like. There is no correct or incorrect way in saying that a song should be classified a certain way, rather it is a person's opinion of what a song makes the person feel and how they relate to it. Nevertheless, there should be some sort of regularity and uniformity when it comes to what a genre sounds like.

Computerized classification labeling that utilizes AI based models have unlocked additional opportunities in this dynamic area of research with favorable outcomes. Various models using neural networks [7] have been implemented along with combinations of other models, to predict the genre with greater accuracies. Likewise complex models such as Double Weighted KNN [6], hybrid models of LSTM and SVM [3] have also been experimented with, in order to further the study in this area.

In this study, music genre classification system is performed based on various machine learning techniques. We aim to perform music genre classification on the GTZAN dataset, by using simple and basic classification algorithms: Logistic Regression and K-Nearest Neighbor and Random Forest and Support Vector Machine and Artificial Neural Networks. Before actually performing the classification, we also make use of dimensionality reduction by using and comparing 3 different techniques: Principal Component Analysis and Linear Discriminant Analysis and Kernel Principal Component Principal Analysis.

II. LITERATURE SURVEY

In the very beginning, music categorization and genre classification were done manually by humans. People who studied this field, would first listen to the audio or music and then categorize them into a particular genre based on past similar music form or style. But since the boom of big data, the amount of data available to us is increasing rapidly, making it infeasible for manual curation. Despite previous studies and research on classifying music genres using machine learning models, there is obviously still room for exploring and creating complex models.

The most influential work using machine learning technique for classifying music genres was done by Tzanetakis and Cook [1] . Archit Rathore and Margaux Dorido [2] proposed several classification models for the GTZAN dataset. The GTZAN dataset is considered as a standard for genre classification. The authors obtained the best accuracy by using SVM with Polynomial kernel.

Prasenjeet Fulzele et al. [3] proposed a model using combination of LSTM and SVM. This paper proposes a model that is implemented in 2 sections. First, properly training both the classifiers individually. Second, combining both the models by using the sum rule, i.e., to calculate the separate posterior probabilities of both models and then adding them to retrieve a combined final result. Ahmet Elbir et al. [4] extracted a total of six features through digital signal processing and training a CNN model. Five different machine learning algorithms were used: KNN, SVM, Naïve Bayes, Decision Tree and Random Forest classifier. Gabriel Gessle et al. [5] performed analysis using a comparison between CNN and LSTM. In this work, two datasets were used namely GTZAN and FMA datasets. Both the datasets were divided into training and validation sets. Using both CNN and LSTM on both the datasets, results were obtained. In 2019, Meimei Wu and Xingli Liu [6], proposed a KNN algorithm which was the doubled weighted version of the simple KNN algorithm on the GTZAN dataset.

Nikki Pelchat et al.[7] used artificial neural networks for genre classification. In their paper, they collected songs, converted them into Short-time segments and then represented the time segments in the form of spectrogram images. These images were further given as input to the CNN model. Their CNN model had multiple convolutional layers and a connected layer. They implemented SoftMax function for genre probability.

A web-based analysis was proposed by Jamie Ramirez Castillo and M. Julia Flores [8]. This paper presents an application that fetches music and songs from online platform like YouTube and segregates them into belonging genres. Their approach uses different machine learning classifiers namely, Naïve Bayes, and Recurrent Neural Networks, Feed-forward and SVM. These models were trained as multi-class classification algorithm structure. Akash Poornasingh and Dylan Dhoray [9] used PCA to accumulate the timbral qualities in order to classify the music to their genres.

III. DATA

GTZAN was initially proposed by G. Tzanetakis and is still one of the most popular music record dataset used for genre classification. The data files were cumulated in around 2000-2001, from numerous sources like CDs, DVDs, radios and microphone recordings. It contains 1000 music track records which are 30-seconds in length, with 22050 Hz sampling frequency and are 16 bits. There are a total of 9991 rows (records) with 59 columns (features/attributes). Genres in the GTZAN are blues, classical, disco, etc. All of these genres have 100 music track record. Each audio track is studied as a .wav format.

A. *Preprocessing*

Different preprocessing methods are applied on the dataset prior to preparing our model. This preprocessing on the dataset is done to make the information more viable with the models and to make our picked dataset heartier.

Features in most datasets follow different scaling. As an outcome, it makes it difficult for the classifying model to converge faster and the time for computation of training increases and it consequently results are not favorable. To overcome this issue, we implement standardization to bring down all the features to a standard scale without changing any variability in the range of the values. We do this by rescaling all the attributes such that their mean value is ‘0’ and the variance value is ‘1’.

Standardization:

$$z = \frac{x-\mu}{\sigma}$$

with mean:

$$\mu = \frac{1}{N}\sum_{i=1}^{N}(x_i)$$

and standard deviation:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}$$

Fig. 1 Formula for Standardization, mean and variance

We use StandardScaler() method from sklearn.preprocessing to achieve this task. In this way, to accomplish most extreme efficiency, we need to preprocess the information before actually classifying the dataset.

IV. Methodology

In this study, we perform the genre classification on the GTZAN dataset, by using five basic classification algorithms: Logistic Regression, Random Forest, KNN, SVM and Artificial Neural Networks. Before the actual classification, we will also perform dimensionality reduction by using three different techniques: PCA, K-PCA and LDA. Each of the classification algorithm is used along with all the three dimensionality reduction techniques and as a result we obtain a total of 15 different combinations of classification models.

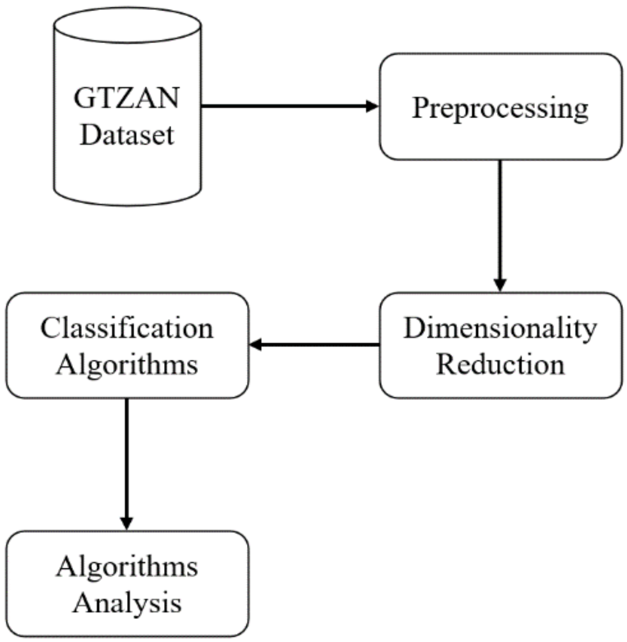

Fig. 2 System Flow Diagram

A. *Machine Learning Classification Algorithms*

As mentioned before, in this part we make use of five different learning algorithms for classification.

- **Logistic Regression:** [14] a process of demonstrating or illustrating the likelihood of distinct outcomes when we provide an input. Often logistic regression works as a binary outcome; anything that takes two values such as 0/1, true/false, etc. Though, this algorithm is mostly used for two class classification, it can also be used for datasets with multiple classes.
- **Random Forest:** [15] Another machine learning algorithms for classification and regression is Random Forest, which is a supervised learning technique. It works by basically building trees on distinct samples and takes their bulk vote for classification and average value in case of regression.
- **K-Nearest Neighbor:** a supervised learning method [6][16][17], also used for classifying data. The input to this algorithm is a variable number of closest training points in a dataset. The output is a label integration or belonging. A data point is categorized based on a majority vote of its

neighboring data points, with the object being assigned to the label which is nearest to or most common among its ‘k’ nearest datapoints.

- **Support Vector Machine:** another supervised classification algorithm [3][18], widely used for classification. The objective of SVM algorithm is to calculate a hyperplane in some n-dimensional space which classifies the training examples clearly in a distinct manner. This hyperplane is constructed by optimization of the boundary surfaces of the labels being separated.
- **Artificial Neural Networks:** ANN or simply Neural Networks [7][19] are a series of algorithms that attempt to emulate the human mind and track down the relationship between the sets of data. An Artificial Neural Network has many layers. Each of these layers perform a specific task, and as the intricacy of the model builds up, the quantity of layers consequently increases. This is a reason for it to be also called as multi-layer perceptron.

B. *Dimentionality Reduction Techniques*

As the model has 59 features, there is obviously a need for the application of dimensionality reduction on the dataset. Some features contribute highly towards the dataset, whereas other features only have very little effect on the entire dataset. We need to capture those features that contribute the highest information to the dataset when we train and test the classification models.

We do this dimensionality reduction by using three techniques:

- **Principal Component Analysis:** PCA[10][11] is an unsupervised statistical procedure[9] that sums up the data content in an enormous information table to a more modest arrangement that can be all the more effectively imagined and analyzed. In simple terms, it will transform a substantial set of variables into a more compact and smaller set, but still containing most of the information from the original set. Hence, reducing the number of attributes/features of the dataset, while preserving as much information as possible. It constructs the principal components in a way that the first component holds the largest variance in the dataset. The next component is then constructed in the same way, but on a condition that it should not be correlated with the first one and should account for the next highest variance. After computing all principal components, it further calculates the covariance matrix and estimates the values of eigen vectors. Finally, we can rank the eigen vectors in order of their eigen values, highest to lowest to get the principal components in order of their significance towards the dataset.
- **Linear Discriminant Analysis:** LDA (or Normal Discriminant Function or Discriminant Function Analysis) is a supervised technique which is widely used for dimensionality reduction, in most of the classification problems. It works by modelling differences in groups or by segregating two or more classes. LDA projects the higher dimension space into a lower dimension space. LDA is likewise firmly related to PCA in that both search for the combinations in a linear fashion of the variables best explaining the data.
- **Kernel Principal Component Analysis:** KPCA is similar[13] to and an extension of the PCA technique and which is a non-linear dimensionality reduction technique. KPCA works by using the kernel method. KPCA technique relies on the assumption that many datasets, which are not linearly separable, can be made so by projecting them onto some other higher dimensional space. The additional dimensions are simply basic arithmetic operations, which are performed on the

unchanged original dimensions. Therefore, we project our dataset onto some higher dimensional space and hence they become linearly separable. Finally, we can perform reduction by applying PCA on this new dataset.

V. EXPERIMENTAL RESULTS

After creating all 15 different combination classification models, the following results were obtained. In this section we compare all the results and evaluate the accuracy of each model. The accuracy values are provided in Table. 1.

Table. 1 Classification results by accuracy

Accuracy (%)			
Classification Algorithm	**Dimensionality Reduction Technique**		
	PCA	***LDA***	***KPCA***
Logistic Regression	53.28	53.32	67.56
Random Forest	77.41	75.14	76.97
KNN	69.20	69.96	72.07
Support Vector Machine	69.40	69.26	74.17
Artificial Neural Network	73.50	65.29	75.17

From the Table. 1, we see that the accuracy achieved by the KNN classification algorithm used along with Principal Component Analysis (PCA) was highest. The value of ‘k’, i.e., number of nearest neighbors used was 3. This value was changed and the model was evaluated by giving values as 5 and 7, but the accuracy value was decreased. A possible explanation could be that, since more data points were considered, the distribution would have affected the calculation of distances between the data points. For PCA, the optimal number of principal components was 10. We constructed a scree plot to analyze and estimate this number.

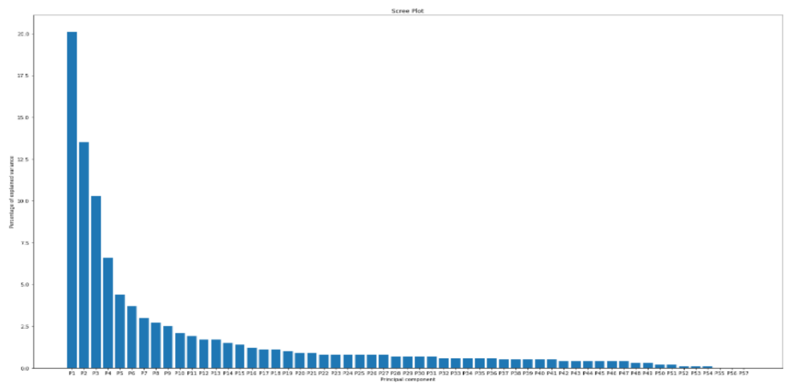

Fig. 3 Scree plot to estimate the optimum number of components for PCA

VI. Conclusion

This study and paper aim to categorize music or music clips based on their genre or form, using simple machine learning algorithms and few dimensionality reduction techniques. Our study has been conducted over two sections. Firstly, reducing the number of features using dimensionality reduction techniques: PCA, LDA and KPCA. Next, we classify the reduced dataset using 5 different classification algorithms: Logistic Regression, Random Forest, KNN, SVM and Artificial Neural Network. According to the result summarized in the previous section, KNN used along with PCA achieved a better accuracy than other models. Future works can include research and study relating to higher dimensional extraction of dataset features and several other complex deep learning models.

References

[1] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293-302, 2002.

[2] Archit Rathore and Margaux Dorido, Music Genre Classification, *Indian Institute of Technology, Kanpur, India*, Semantic Scholar, 2015.

[3] Prasenjeet Fulzele, Rajat Singh, Naman Kaushik, Kavita Pandey, A Hybrid Model for Music Genre Classification Using LSTM and SVM, *Proceedings of 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, India*, 2-4 August, 2018.

[4] Ahmet Elbir, Hilmi Bilal Çam, Mehmet Emre İyican, Berkay Öztürk, Nizamettin Aydın, Music Genre Classification and Recommendation by Using Machine Learning Techniques, *2018 Innovations in Intelligent Systems and Applications Conference (ASYU), Adana, Turkey*, 2018.

[5] Gabriel Gessle and Simon Akesson, A comparative analysis of CNN and LSTM for music genre classification, *KTH Royal Institute of Technology, Stockholm, Sweden*, Semantic Scholar – 2019.

[6] Meimei Wu and Xingli Liu, A Double Weighted KNN Algorithm and Its Application in the Music Genre Classification, *6th International Conference on Dependable Systems and Their Applications*, 2019.

[7] Nikki Pelchat and Craig M. Gelowitz, Neural Network Music Genre Classification, *Canadian Journal of Electrical and Computer Engineering*, Vol. 43, No. 3 – 2020.

[8] Jaime Ramirez Castillo and M. Julia Flores, Web-Based Music Genre Classification for Timeline Song Visualization and Analysis, *IEEE Journal Article* Vol. 9 – 2021.

[9] Akash Poornasingh (Member, IEEE) and Dylan Dhoray (Member, IEEE), Similarity Analysis of Modern Genre Music Based on Billboard Hits, IEEE Journal Article Vol. 9 – 2021.

[10] Nazila Panahi, Mahkrokh G. Shayesteh, Sara Mihandoost and Behrooz Zali Varghahan, Recognition of different datasets using PCA, LDA, and various classifiers, *5th International Conference on Application of Information and Communication Technologies, Baku, Azerbaijan, 2011.*

[11] https://en.wikipedia.org/wiki/Principal_component_analysis#:~:text=Principal%20component%20analysis%20(PCA)%20is,components%20and%20ignoring%20the%20rest.

[12] https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c?gi=b93822ea471b

[13] https://towardsdatascience.com/kernel-pca-vs-pca-vs-ica-in-tensorflow-sklearn-60e17eb15a64

[14] https://www.sciencedirect.com/topics/computer-science/logistic-regression#:~:text=Logistic%20regression%20is%20a%20process,%2Fno%2C%20and%20so%20on.

[15] https://www.analyticsvidhya.com/blog/2021/06/understanding-randomforest/#:~:text=Random%20forest%20is%20a%20Supervised,average%20in%20case%20of%20regression.

[16] https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning

[17] https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[18] https://www.geeksforgeeks.org/support-vector-machine-algorithm/#:~:text=Support%20Vector%20Machine(SVM)%20is,distinctly%20classifies%20the%20data%20points.

[19] https://www.sciencedirect.com/topics/earth-and-planetary-sciences/artificial-neural-network#:~:text=Artificial%20neural%20network%20(ANN)%20is,on%20their%20predefined%20activation%20functions.