International Journal of Research in Engineering, Science and Management
Volume 4, Issue 7, July 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

Music Genre Classification using Machine Learning Techniques

K. Pushpalatha[1], U. S. Sagar[2*], Rashmi[3]
[1,3]*B.E. Student, Department of Computer Science and Engineering, Srinivas Institute of Technology, Mangaluru, India*
[2]*Asst. Professor, Department of Computer Science and Engineering, Srinivas Inst. of Technology, Mangaluru, India*

***Abstract*: Digital music processing is involved in many subjects, including music genre prediction. Machine learning techniques were used to classify music genres in this research. Deep neural networks have recently been shown to be successful in a variety of classification tasks, including the classification of music genres. In recent years, deep neural networks have been shown to be effective in many classification tasks, including music genre classification. In this paper, we proposed two ways to improve music genre classification with convolutional neural networks: 1) combining max- and average pooling to provide more statistical information to higher level neural networks; 2) using shortcut connections to skip one or more layers, a method inspired by residual learning method. The input of the KNN is simply the short time Fourier transforms of the audio signal. The output of the KNN is fed into another deep neural network to do classification. By comparing two different network topologies, our preliminary experimental results on the GTZAN data set show that the above two methods can effectively improve the classification accuracy, especially the second one.**

***Keywords*: KNN, Machine Learning, Classification, Deep neural network.**

1. Introduction

With day-by-day increasing internet penetration, huge amount of useful data is available at proximity to people. Although it seems that there is ease of access to data, but this exponentially increasing amount of data brings to table a new

problem – most of this chunk is unclassified. We will create a deep learning project to automatically distinguish various musical genres from audio files in this project. We'll categorise these audio files based on their low-level frequency and time domain characteristics. We'll need a dataset of audio tracks of identical lengths and frequency ranges for this project. The KNN classification dataset is the most commonly recommended dataset for music genre classification projects, and it was collected specifically for this purpose. With the rise in popularity of personal multimedia devices in recent years, a vast amount of music has become available on a variety of platforms. Humans are finding it difficult to structure and organise such a vast volume of music. One of the current methods for structuring music content is genre grouping. To allow automated structuring and organising of large music archives, an efficient and precise music genre classification system is urgently required. The musical genre is a kind of high-level mark. The standard phase of an automated genre classification system consists of three steps as a classification problem: 1) From the original audio signal, timbre, spectro-temporal, and statistical features are extracted; 2) To increase classification accuracy, several techniques are used to pick a meaningful subset of features or aggregate features.3) To automatically classify the input music into different genres, a machine learning-based classifier is trained over the selected features. Finding appropriate representations or features for the system is a critical component of its performance. Extraction of hand-crafted features from the original songs is a popular way to do this. This method necessitates technical know-how as well as engineering creativity. Neural net-works have become very successful in a variety of fields, including retrieving musical knowledge, thanks to the advancement of deep learning (MIR). With convolutional neural networks, we suggest two ways to increase the accuracy of music genre classification.:1) combining peak and average pooling to provide further statistical information to higher-level neural networks using a technique influenced by residual learning; 2) Bypassing one or even more layers with shortcut connections.

2. Literature Survey

Hareesh Bahuleyan's novel Haree (2018). Using Machine Learning techniques, the researchers developed a method for automatically identifying music in a user's collection by assigning tags to the songs in their collection. In order to accomplish their goals, it explores both Neural Networks and traditional Machine Learning algorithms. The first approach uses a Convolutional Neural Network that is trained from beginning to end using audio signal Spectrogram features (images).

To classify the music into its genres, the second approach employs a range of Machine Learning techniques. Random Forest, Logistic Regression, Gradient Boosting (XGB), and Support Vector Machines are some of the ML algorithms that are used (SVM). After examining both approaches separately, they discovered that the VGG-16 CNN model had the best accuracy. A VGG-16 CNN and XGB ensemble classifier was used to create the optimised model, which had an accuracy of

*Corresponding author: sagar.udupa@gmail.com

0.894.

Tzanetakis and colleagues (Tzanetakis et al., Tzanet (2002). They have looked at how to automatically classify audio signals into musical genres. They agree that these music genres are just human-made categorical labels used to group together pieces of music. Some general characteristics are used to classify them. All of these features are influenced by the instruments used, the rhythmic structures, and, most significantly, the harmonic nature of the music. To organise incredibly large online music collections, genre hierarchies are often used. They've proposed three different feature sets: timbral texture, rhythmic material, and pitch content. Researchers were able to examine proposed features in order to analyse their performance and relative significance by training statistical pattern recognition classifiers using real-world audio collections. In this paper, we look at both whole-file and real-time frame-based classification schemes. This model correctly classifies nearly 61 percent of ten music genres using the proposed feature sets.

Lu, L.et al., (2002). They presented their research on audio content segmentation and classification in content review for audio segmentation and classification. The type of audio or the speaker's identity are used to separate an audio stream into parts. Their plan is to build a robust model that can distinguish and segment audio signals into speech, music, ambient sound, and silence. This two-part grouping can be used for a number of purposes. The first move is to separate speech from non-speech discrimination. This paper introduces a new algorithm based on KNN (K-nearest-neighbor) and LSP-VQ (linear spectral pairs-vector quantization). After that, using a rule-based classification system, the non-speech class is divided into three categories: music, environmental sounds, and silence. They've used some unusual and innovative features, such as noise frame ratio and band periodicity, which are not only introduced but also thoroughly discussed. Also included and developed is a speaker segmentation algorithm. This is going on without any kind of oversight. It makes use of a new correlation analysis approach that combines quasi-GMM and LSP correlation analysis. The model will aid open-set speakers, online speaker modelling, and real-time segmentation without the need for any prior information.

Tom L. H. Li et. al., (2010). They've also created a speaker segmentation algorithm and used a convolutional neural network to extract musical pattern features automatically. This is a condition that is totally unsupervised. It uses a one-of-a-kind approach that combines quasi-GMM and LSP correlation analysis. Without any prior experience, the model can handle open-set speakers, online speaker modelling, and real-time segmentation. Audio clip-derived statistical spectral characteristics, rhythm, and pitch are less reliable, resulting in less accurate models. As a result, they took a different approach to CNN, focusing on musical data that is similar to image data and requires very little prior knowledge. The dataset in question was GTZAN. In total, there are ten genres with 100 audio clips in each. Each audio clip is 30 seconds long, with a 22050 Hz sampling rate and 16-bit resolution. When analysing musical patterns with the WEKA system, several classification models were taken into account. At first, the classifier's accuracy was 84 percent, but it gradually improved. The CNN features were more reliable and generated better results than the MFCC, chroma, and temp features. The accuracy of parallel computation on different combinations of genres can also be improved.

Convolutional Neural Networks Improved Music Genre Classification, Weibin Zhang (2016) proposed two methods for boosting music genre classification with convolutional neural networks: 1) combining peak- and average pooling to provide more statistical knowledge to higher level neural networks using a method influenced by residual learning; 2) using shortcut connections to bypass one or more. The CNN's production is fed into a classification deep neural network. Our preliminary findings are based on a comparison of two network typologies. Our preliminary experimental results on the GTZAN data set, which compare two different network typologies, show that the above two methods, especially the second, can effectively improve classification accuracy.

Andreas Rauber and Thomas Lidy. They presented "Evaluation of feature Extractors, Psycho-Acostic transformation for music genre classification," a study on the importance of psycho-acoustic transformations for effective audio feature estimation. Results show which parts of the algorithm for Rhythm Patterns function extraction are critical and which parts are problematic. We introduce two new function representations in this context: Statistical Spectrum Descriptors and Rhythm Histogram functions. A music genre classification task, involving three reference audio collections, is used to evaluate both the individual and combined feature sets. On the same data sets, the findings are compared to previously reported steps. Psycho-acoustic transformations significantly increase classification accuracy in all environments, according to experiments.

M. Julia Flores, Jaime Ramirez The paper focuses on emerging machine learning developments as they apply to the issue of music annotation. They have a music genre classification experiment that uses an Audio collection to compare various machine learning models. This paper adds to the domain of music genre classification (MGC) by offering an up-to-date analysis of recent research from a variety of perspectives. A deep fully-connected neural network was used to create the Audio set baseline audio event multi-label classifier. The reference metric for standardised audio event classification with Audioset provided by Google is a mean AP of 0.314 and an average AUC of 0.959, with the top performing class music having an AP of 0.896 and AUC of 0.951 and the worst performing class "Rattle" having an AP of 0.020 and AUC of 0.796. They focused on supervised learning in the context of music classification for the purpose of this article. Supervised learning builds a theory based about how each sample's inputs relate to the related predicted output in order to accurately predict the output of new inputs.

Amelie Anglade, Rafael Ramirez, and Simon Dixon are among the artists who have contributed to this collection (2009). Harmony rules derived from automatic chord transcriptions were used for genre classification. They demonstrated in this paper that our genre classification system,

which is focused on harmony and first-order logic and has previously been tested on symbolic data, can learn classification models directly from audio data with classification accuracy well above chance level. Even when working with audio details, we can use a chord transcription algorithm to follow a high level representation of harmony. As a result of this high-level representation of harmony based on first-order logic, classification models that are human-readable, i.e. transparent, are possible. We improved clarity by implementing a new harmony representation scheme based on the western harmony representation, which defines chords in terms of degrees and chord categories. Not only is this representation musically more meaningful than the previous one we used, but it also yielded better classification results and enabled us to construct classification models faster.

In Alessandro L. Koerich, Carlos N. Silla Jr and Celso A. A. Kaestner (2008). Automatic Music Genre Classification Using a Machine Learning Approach. This paper presents a novel method for automatically classifying music genres. In the proposed solution, which is based on time and space degradation schemes, many feature vectors and a pattern recognition ensemble approach are being used. Regardless of the fact that categorising music genres is a multi-class task, we solve it by combining the results of a collection of binary classifiers (space decomposition). Decomposition of the music samples is also done using the time segments extracted from the start, center, and ending portions of the original music signal(time-decomposition). To get the final classification from the collection of individual outcomes, a hybrid technique is used. Some of the really common machine learning algorithms include Nave-Bayes, Decision Trees, k Nearest-Neighbors, Support Vector Machines, and Multi-Layer Perceptron Neural Nets. The Latin Music Database, which contains 3,160 music parts split into ten musical genres, was used in the experiments. As per the results of the experiments, the proposed ensemble approach outperforms global and individual segment classifiers in the vast majority of cases. In addition, using the genetic algorithm model, some feature selection experiments were carried out. They show that the most significant features for classification differ depending on where they came from in the music signal.

Aadam Saleem and Anshuman Goel, as well as Sarfaraz Masood (2017). Neural Network-Based Music Genre Classification (MGC). The aim of this project is to automate the manual classification of music genres in each album. This work allows for real-time categorization of songs, and the proposed parallel architecture can be deployed on a multi-processing computer. As a result, Echonest libraries are used to gather features like beats, tempo, energy, loudness, speechiness, valence, danceability, acousticness, DWT, and so on, which are then fed into a Parallel Multi-Layer Perceptron Network for song genre classification. Though classifying the songs for two distinct well-known genres of Indian Hindi songs, Sufi and Classical, the proposed scheme had an accuracy of about 85 percent. In recent years, neural networks (NNs) have been used to classify music genres with varying degrees of performance. The output of song libraries, machine learning algorithms, input formats, and the types of NNs used has all been mixed. This article examines a few of the machine learning techniques employed in this area. It also contains research on music genre classification(MGC). In this analysis, images of spectrograms generated from time slices of songs are fed into a NN to classify the songs into their corresponding musical genres.

3. Methodology

This section explains how we built this music genre classifier, which consists of three main measures. -

1. Data Preparation
2. Extraction and Selection of Features
3. Categorization

1) Data Preparation

Before diving right into feature analysis and classification, one needs to have an appropriate format of analyzable data. The GTZAN data-set has data in the ".au" format.

2) Extraction and selection of Features

Any audio signal is made up of a variety of features. We must, however, remove the characteristics that are pertinent to the problem we are attempting to solve. Feature extraction is the procedure for extracting features in order to use them in research. The Mel frequency of a signal MFCCs are a small set of features (usually 10–20) that describe the overall shape of a spectral envelope in a concise manner. It is focused on the features of human vocalisations. Feature extraction is a dimensionality reduction technique that breaks down a large volume of raw data into manageable chunks. These large data sets often have a large number of variables. The large number of variables in these large data sets necessitates a lot of computational resources to process.

3) Categorization

Finally, our main goal is to identify the information provided. Audio dataset based on genre was accomplished by the classifier method been there.

We applied two most famous classifiers and found the best which helped us solve our problem. The two classifiers were,

1. KNNC (K-nearest neighbor classifier).
2. SVM (Support Vector Machine).

KNNC: k-Means is a term used to describe a group of people who Clustering is an unsupervised clustering algorithm, while KNN is a supervised classification algorithm. We ran the KNN algorithm several times with various K values to find the K that reduces the number of errors while maintaining the algorithm's ability to make accurate predictions when given data it hasn't seen before.

A few things to keep in mind:

1. As K approaches zero, our predictions become less stable. Consider the situation where K=1 and the question point is surrounded by several reds and one green (I'm thinking of the top left corner of the coloured plot above), but the green is the only one who is close to it. The query point should be red, but since K=1, KNN predicts it will be green.

2. However, as we raise the value of K, our predictions become more stable, and therefore more likely to be accurate, due to majority voting/averaging (up to a certain point). The number of errors slowly starts to rise. At this point, we know

we've shifted K's value too far.

3. In cases where we take a majority vote among brands, we usually make K an odd number to have a tiebreaker (e.g. choosing the mode in a classification problem).

SVM:

The formal description of an SVM (Support Vector Machine) is a discriminative classifier with a separating hyperplane. To put it another way, the algorithm produces an optimal hyper plane that categorises new examples given labelled training data (supervised learning). In order to differentiate the two types of data points, we must select one of many hyper planes. Our goal is to find the plane with the greatest margin, or the greatest distance between data points from both classes. By increasing the margin difference, it becomes easier to distinguish possible data points.

The position and orientation of the hyper plane are affected by support vectors, which are data points that are closer to the hyper plane. These help vectors are used to optimize the margin of the classifier. The path of the hyper plane would be changed if the support vectors were removed. These are the factors that will influence the development of our SVM. They have two key advantages over newer algorithms such as neural networks: they are faster and perform better with fewer samples (in the thousands). This makes the algorithm ideal for music genre classification problems, where a dataset of just a few thousand labelled samples is typically available.

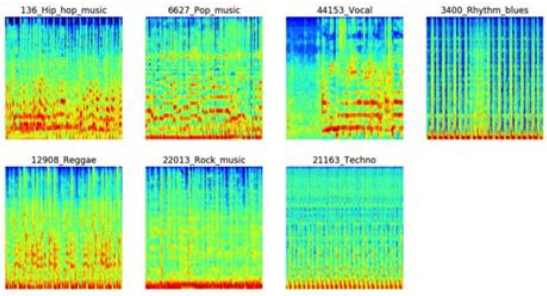

Fig. 1. Spectrograms for audio signal

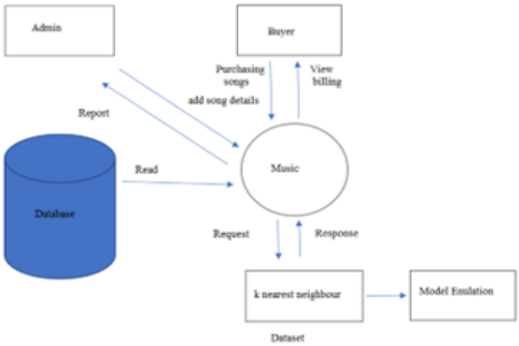

Fig. 2. Architectural design

4. Implementation

In short, the supervised machine learning algorithm K-nearest neighbors is a technique which classifies a data point by calculating the distance between k labelled data points. The amount of votes it gets from the k-nearest data points is used to calculate it. The user can get details about different songs using this restriction scheme. To use our framework, a user must first create an account. The server administrator will be in charge of adding, removing, and updating the song dataset. The system's administrator is in charge of overseeing its operation, after which the user can send the different songs to the system which he would want to know.

We make use of a protocol that is followed step-by-step from recording songs to giving the final data of the song asked by the user. Each step has its different use in our system.

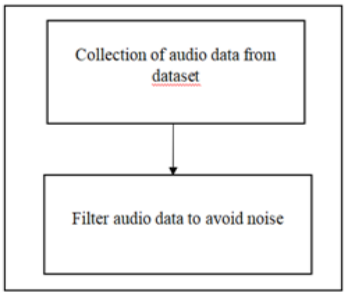

Fig. 3. Steps in pre-processing

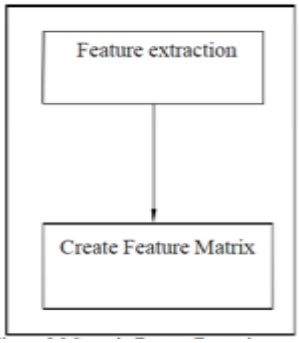

Fig. 4. Steps in feature extraction

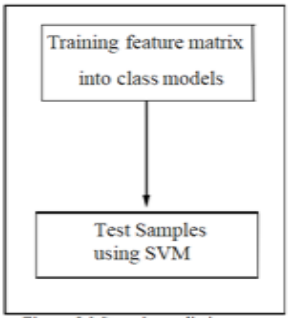

Fig. 5. Steps in prediction

1. Pre- Processing is where we convert the data based on our need. Unwanted noise, blank spaces are removed in this step.
2. The most commonly used features for describing the spectrum of an audio recording in a concise and insightful manner are feature extraction and after which this data is stored in a matrix format for further use. For this purpose, we make us of Librosa package of python. This contains many APIs that will help in fetching the features.
3. A function is an observable property or attribute of an observed phenomenon. Feature extraction begins with a features which are intended to be descriptive and non-

redundant array of calculated data and constructs, easing the learning and generalisation steps and, in some cases, resulting in better interpretations made by humans. SVM is used to analyse samples at a later date.

A. Support Vector Machine (SVM)

Pattern classification is normally dealt with by a support vector machine. Linear and non-linear patterns are two types of patterns. Linear patterns are patterns that are easily distinguishable or can be easily distinguished in low dimensions, whereas nonlinear patterns are patterns that are difficult to distinguish or cannot be easily separated, necessitating further investigation They've been exploited to make it simple to divide them. SVM's key concept is to construct an optimal hyper plane for linearly separable patterns that can be used for classification. The most suitable hyperplane for classifying patterns is the one that maximizes hyper plane's margin, i.e. the distance between the hyper plane and each pattern's closest point SVM's main goal is to optimize the margin so that it can correctly classify the given patterns; the greater the margin size, the more accurately it can classify the patterns.

- Step 1: Imports
- Step 2: Create a function to locate neighbours and calculate the distance between feature vectors.
- Step 3: Determine who your immediate neighbours are.
- Step 4: Make a model evaluation function.
- Step 5: Extract features from the dataset and save them as a binary.dat file called “my.dat.”
- Step 6: On the dataset, there is a train and test break.
- Step 7: Make a forecast with KNN and see how accurate it is on real-world results.

Prediction is a phase where we test a single audio file against our trained dataset. Input audio file undergoes pre-processing and feature extraction process. Weight file is created. Compare model weights with input audio features. Obtained output (song name) is displayed in web application with few additional features of song.

5. Results

The k-nearest neighbour (kNN) classification algorithm is a non-parametric classification that ballots the "k" nearest training points to a given test point and analyses the test point depending on their majority of votes. The kNN algorithm works well enough with data that is expressed by a series of dense clusters, each represents a distinct label. This algorithm, however, is easy to implement, it is vulnerable to inputs with a large number of dimensions. This section represents the result of our project. We have included few images of our system which are obtained using machine learning techniques like pre-processing, feature extraction and SVM.

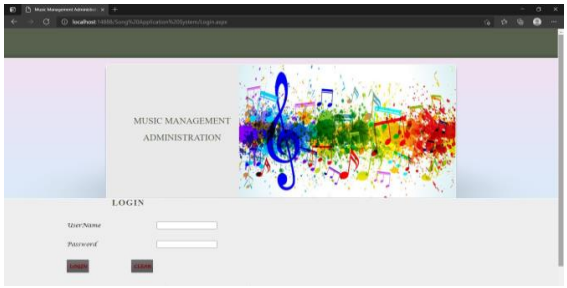

Fig. 6. Login page

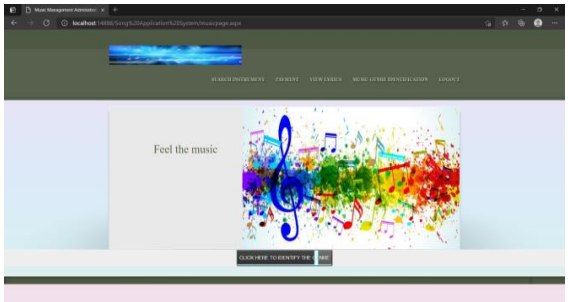

Fig. 7. Dashboard

Table 1
Analysis table

Name of the Paper	Methodology/ Algorithm	Advantages	Disadvantages
[1]. Music Genre Classification using Machine Learning Techniques	Logistic regression, Random forest	It proposes a method for automatically classifying music by assigning tags to each song.	The study's data set consisted of YouTube video audio clips, which are infamous for being loud. Futures research would aid in the development of methods for pre-processing noisy data before loading it into a machine learning model, in order to increase performance.
[2] The audio signal is used to classify music genres.	STFT algorithm, Mel-frequency Cepstral coefficients [MFCC]	It investigated how audio signals are categorised into a music genre hierarchy automatically	The proposed feature set is useless for query-by-example retrieval of musical signal similarities and audio thumb-nailing.
[3]. Auditory segmentation and classification using content analysis.	KNN[k-nearest algorithm] Linear Spectral Pair vector quantization.	This analysis of audio content processing for classification and segmentation divides an audio stream into segments based on audio type or speaker recognition.	There is no effective way to use audio content analysis to aid in video content analysis and indexing.
[4] Automated Musical Pattern Feature Extraction with Convolutional Neural Networks	CNN [Convolutional Neural Network]	Create the most accurate music genre classification model possible and extract musical patterns from the audio file.	The model is insufficiently robust to apply the training results to previously unknown musical data.
[5] Multifaceted Analysis and Experimentation with Audio in Machine Learning for Music Genres	NB classifier, DNNs, linear SVM and RNN are all examples of decision trees.	When using Audioset data, the performance of various models and music genres varies greatly.	Some audio characteristics derived from the raw audio signals were left out.

6. Analysis

The advantages and drawbacks of the current systems and approaches are described in tabular form below. In comparison to the current system, the proposed system's methodology/ algorithm is more effective.

7. Conclusion

FM a small is a dataset from the Free Music Collection. It is used to investigate music genre classification. We suggested a simple solution to the classification problem and compared it to a number of further complex, reliable models. We contrasted the models as well. according to the type of data they were providing. The models were fed with spectrogram images for CNN models and audio features stored in a csv for Logistic Regression and ANN models. Simple ANN was known to be the optimum feature-based classifier among Logistic Regression and ANN models, with an accuracy level of 64%. The CNN model was found to be the best spectrogram-based model among the CNN, CRNN, and CNN-RNN parallel models, with an accuracy of 88.5 percent. If the dataset is expanded, the CRNN and CNN-RNN models should perform well. Overall, it appears that image-based classification outperforms feature-based classification.

References

[1] Ossama Abdel-Hamid, Abdel-Rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu. 2014. Convolutional neural networks for speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 22(10):1533–1545.

[2] Loris Nanni, Yandre MG Costa, Alessandra Lumini, Seung Ryul Baek, and Moo Young Kim for music, genre that combines visual and audio elements classification. Expert Systems with Applications 45:108–117, 2016.

[3] Thomas Lidy and Alexander Schindler. Parallel convolutional neural networks for music genre and mood classification. MIREX2016, 2016.

[4] Chathuranga, Y. M, & Jayaratne, K. L. Automatic Music Genre Classification of Audio Signals with Machine Learning Approaches. GSTF International Journal of Computing, 3(2), 2013.

[5] George Tzanetakis and Perry Cook. 2002. Musical Audio signal genre grouping IEEE Transactions on speech and audio processing 10(5):293–302.

[6] Hagen Soltau, Tanja Schultz, Martin Westphal, and Alex Waibel. 1998. Recognition of music types. InAcoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on. IEEE, volume 2, pages 1137–1140.

[7] Nitish Srivastava, Geoffrey Hinton, Ale Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15(1):1929–1958.

[8] Nicolas Scaringella and Giorgio Zoia. 2005. On the audio signal recognition systems, simulation of time details for automated genre ISMIR (Institute for Scientific and Medical Research), pp. 666–671.

[9] Dan-Ning Jiang, Lie Lu, Hong-Jiang Zhang, Jian-Hua Tao, and Lian-Hong Cai. 2002. Music type classification by spectral contrast feature. In Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on. IEEE, volume 1, pp. 113– 116.

[10] Chanwoo Kim and Richard M Stern. 2012. Power normalized Cepstral coefficients (pncc) for robust speech recognition. In Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, pp. 4101–4104.