NASHVILLE TO NEURAL NETWORKS: A STUDY OF SUBGENRE RECOGNITION IN COUNTRY MUSIC

An Undergraduate Research Scholars Thesis

by

MATTHEW CHANG

Submitted to the LAUNCH: Undergraduate Research office at
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Faculty Research Advisors: Dr. Tracy Hammond
Dr. Paul Taele

May 2023

Major: Computer Science

RESEARCH COMPLIANCE CERTIFICATION

Research activities involving the use of human subjects, vertebrate animals, and/or biohazards must be reviewed and approved by the appropriate Texas A&M University regulatory research committee (i.e., IRB, IACUC, IBC) before the activity can commence. This requirement applies to activities conducted at Texas A&M and to activities conducted at non-Texas A&M facilities or institutions. In both cases, students are responsible for working with the relevant Texas A&M research compliance program to ensure and document that all Texas A&M compliance obligations are met before the study begins.

I, Matthew Chang, certify that all research compliance requirements related to this Undergraduate Research Scholars thesis have been addressed with my Faculty Research Advisors prior to the collection of any data used in this final thesis submission.

This project did not require approval from the Texas A&M University Research Compliance & Biosafety office.

TAMU IRB #: 2022-1542 Exemption Date: 02/15/2023

TABLE OF CONTENTS

ABSTRACT

Nashville to Neural Networks: A Study of Subgenre Recognition in Country Music

Matthew Chang
Department of Computer Science and Engineering
Texas A&M University

Faculty Research Advisor: Dr. Tracy Hammond
Department of Computer Science and Engineering
Texas A&M University

Faculty Research Advisor: Dr. Paul Taele
Department of Computer Science and Engineering
Texas A&M University

With the growing popularity of music streaming services such as Spotify and Apple Music, music genre recognition has become an increasingly important task. In recent years, music services have gone beyond the basics and have started catering to users with song recommendations and auto-generated playlists. These recommendations and playlists help to better engage listeners by introducing them to tracks based on genre, mood, or user listening activity. While the metadata of a song gives useful information about a track such as artist and duration, the true genre of a song is characterized by the musical features contained in audio.

As with other mainstream musical genres, country music is divided into multiple subgenres. Originating in the early twentieth century in the Southern United States, country music has been shaped by many influences and produced many unique subgenres including bluegrass, country rock, traditional country, and country pop. As a whole, each subgenre has its own unique sound and style, but the distinctions between them are not easily identifiable to people who are unfamiliar with country music. The creative side of music makes subgenre classification especially complicated,

with loosely defined boundaries and even individual songs belonging to more than one subgenre. While the exact boundaries between the subgenres can be blurry, they still remain important to user preference. A quantitative approach to identifying subgenres within country music can further improve user engagement when it comes to country music recommendations.

In this research, I create a machine learning model that distinguishes between the subgenres of country music given a short audio clip. I use sample audio and genre label data from Spotify to determine the classification boundaries between the subgenres. Then, I evaluate the accuracy of my model and compare its performance using different audio features and audio processing techniques. I also compare its performance to previous research done on music genre recognition. Through this process, I identify key distinguishing features within the audio that are responsible for differentiating country music's subgenres. These features can be used by music services and listeners alike to quantify listener country music subgenre preference, helping listeners find music they enjoy and driving up engagement with the music industry.

ACKNOWLEDGMENTS

Contributors

I would first and foremost like to thank God for the opportunity to pursue an education at Texas A&M University. I would also like to thank my faculty advisors, Dr. Tracy Hammond and Dr. Paul Taele, for their guidance and support over the course of this research. They, along with the other staff from the Department of Computer Science and my fellow peers in the department have been developmental in my learning journey.

Finally, thanks to my family, friends, and buddies in the Corps of Cadets for their encouragement, support, and involvement in helping me become the man I am today.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

Undergraduate research was supported by Dr. Tracy Hammond and Dr. Paul Taele at Texas A&M University.

1. INTRODUCTION

1.1 Motivation

Music genres help to classify music into similar artistic styles and content. Attempting to accurately classify songs under these genres is known as music genre recognition (MGR), and it helps the music industry to define target audiences and match consumers with music they like [1].

Music genre recognition is a well-studied area of research and streaming services such as Spotify and Apple Music deploy music genre recognition techniques for creating auto-generated playlists and music recommendations [2]. For example, a Spotify user who tends to listen to music labeled as classic rock is more likely to enjoy recommended songs that fall under classic rock than an alternative genre. Since they have discovered music that they like through Spotify's services, they are likely to keep coming back for more. A wrong classification of genre would have the opposite effect on this feedback loop.

Both the user and the music industry benefit from this interaction. For the user, they discover music they enjoy. For the industry, they benefit from the increased use of streaming services, attention to artists, referrals to friends, concert attendance, and so on. All of these things lead to increased revenue across the board.

A key issue to be solved in music genre recognition is determining what features of music differentiate the genre of one music audio clip from another. This differentiation is commonly done through music data preprocessing and the use of machine learning models. While models such as KNN haven proven to have 90% classification accuracy [3] on the GTZAN music audio dataset, differentiation between subgenres requires examining audio features at a more detailed level.

Much research has been done on using machine learning methods to differentiate between common, well-known genres, for example, differentiation between rock, classical, and country, however, there is limited research that has been done on subgenre classification, particularly within the country music genre.

Nevertheless, country music is widespread in North America and plays an influential role in

American culture. Since its origins in Appalachia in the early twentieth century, country music and its subgenres have defined a unique culture, especially in the Southern United States. Dolly Parton, Willie Nelson, and George Strait “the King of Country” are just a few of the major artists who have shaped the country music industry. The influence of country music transcends the boundaries of country fans; pop music artists like Taylor Swift and Sam Hunt have been heavily influenced by the country music style, and artists such as Lil Nas X have taken a new spin on music, combining elements of both hip-hop and country music.

For country music producers, the authenticity of country music is important to keep country music fans engaged. For the last century, people have "complained that contemporary country is ‘selling out’ because it sounds like pop music” [4]. This variability of style between country music subgenres causes some listeners to prefer one subgenre of country music over another. For example, in College Station, Texas, the typical country music enthusiast prefers traditional country over pop country. Dyck [4] would argue that this preference is rooted in holding onto outdated ideals rooted in racism. While this may be true in certain scenarios, it is clear that there exist cultural, geographic, and musical factors that play into the different subgenres of country music [5], and naturally, different musical styles will lead to different subgenre preferences. In this paper, we will boil down subgenre differentiation to the science of sound. There is a lack of existing research on automated country music subgenre recognition by audio, and this paper serves to fill that void. The knowledge and understanding of the musical differences between the subgenres of country music are essential in a day and age when the sustainability of music producers and streaming services depends on engaging their intended audience.

1.2 Research Proposal

In this research, I use multiple machine learning methods that have been utilized on other music datasets and compare their performance on a country music genre data retrieved from the Spotify API. The Spotify dataset consists of 1000 30-second audio clips and subgenre labels falling under the country genre. These audio clips are transformed and run through multiple machine learning models namely KNN, neural network, and random forest. The models are finetuned and evaluated for their accuracy.

This research attempts to answer the question of if differentiation between subgenres in country music is possible at all, or if the subgenres are too musically similar to be effectively differentiated. If subgenres can be differentiated, we strive to learn what audio features differentiate the subgenres of country music and how we can best select features and tune hyperparameters to get the most accurate models.

2. RELATED WORK

2.1 Machine Learning Classifiers

There is a multitude of research that has been done on music genre recognition (MGR). Most of the research applies machine learning classifiers to differentiate between music genres. One of the earliest research papers on musical genre classification using machine learning was from George Tzanektakis in 2002 [6]. Tzanektakis used a Gaussian mixture model classifier and K-nearest neighbor classifier to categorize the music audio by genre. He found that he could achieve a 61% classification accuracy between the 10 genres which was comparable to human performance for genre classification.

In addition to providing extensive research on the science of MGR, Tzanektakis also created the GTZAN dataset, consisting of 10 genres with 100 audio files each. It was created using audio collections from CDs, the radio, and the Web. Due to its accessibility, it has become the most-used public dataset for MGR. Although it is the MGR standard for training models, it is important to note that it is not completely perfect. Some 7.2% of the excerpts come from the same recording, and 10.6% of the dataset is mislabeled, potentially leading to artificial inflation in the accuracy of systems that are split across training and test data, e.g., k-nearest neighbors, or artificial decrease in performance of systems that build models on data, e.g., Gaussian mixture models and boosted trees [7].

Since the creation of the GTZAN dataset, several other studies have achieved higher accuracy models on the same data using different data transformation techniques and machine learning classifiers [3, 8, 9, 10, 11]. The highest accuracy performance was an SVM which achieved an accuracy of 97.20% [10]. Another notable performance was with the KNN model and MFCC features was able to achieve a classification accuracy of 90% [3]. Other machine learning classifiers, namely decision tree and random forest also seemed to perform decently at 65% and 86% respectively. In another study using multiple classifiers on the GTZAN dataset, KNN proved to have the highest accuracy at 77% with naive bayes coming in second at 73% and decision tree in last at 59% [8]. Another third notable study focused more on neural networks and was able to achieve an accuracy of 88.1% by using a multilayer perceptron combined with data filtering methods [9].

In addition to different feature selection, because different research projects have run their models on other well-known datasets besides the GTZAN such as the Free Music Archive [12], it makes it difficult to directly compare performances between different classifiers between different studies. However, it seems that some ML classifiers that have consistently proved to be effective for audio classification include:

- K-Nearest Neighbors (KNN) [6, 8, 13]
- Neural Network [9, 10, 14]
- Random Forest [3, 15]
- Support Vector Machine (SVM) [13, 14, 15, 16, 17]

2.2 Audio Feature Selection

2.2.1 Mel-Frequency cepstral coefficient (MFCC)

Mel-Frequency cepstral coefficient (MFCC) data is one of the primary ways of training machine learning models on audio data since ML algorithms can't take in audio data directly. They have been used to effectively differentiate many kinds of audio including audio scenes [18] and music emotion [19].

MFCC is useful because it is lower-level and directly calculated from the raw audio data. Low-level MFCC data is significantly more informative to the machine learning classifiers than high-level Spotify-provided features (e.g., tempo, time signature, and energy) [3]. Using the same machine learning algorithm, MFCC features perform almost twice as well as Spotify features. The drawback is that the sound signal must undergo transformations to arrive at MFCC values and no longer provide meaningful insights to a user whereas the Spotify features are relatable from a human listener's perspective.

One study shows that MFFC coefficient-related features have the highest impact on classification for both genre and subgenre according to the features they used [20]. Another study found that Fast Fourier Transform features (FFT) were even more important than MFCC for classification, and a hybrid mix of features such as adding pitch and beat-related features contributed to better classification as well [21].

2.2.2 Spectrogram

Another common feature to differentiate audio data is called a spectrogram. A spectrogram is a visual graph that represents the frequencies that exist over the duration of an audio clip. Music audio is converted to a visual format which is then processed by a deep-learning convolutional neural network. This essentially turns an audio data recognition problem into an image recognition one. It has been used in a variety of sound research including video game genre recognition [22] and instrument recognition [23], where they achieved a 73.7% accuracy. However, when it comes to music genre classification, [24] found that the spectrogram method proved to not be as accurate as traditional ensemble methods. However, he noted that this was possibly due to GPU limitations on the computations required for CNNs and DNNs. Therefore, we will stick to MFCC and lower-level audio features for this study.

2.3 Subgenre Recognition

Compared to music genre recognition, subgenre recognition research is available but not nearly as common. In 2011, Tsatsishvili performed a study on the automatic subgenre classification of heavy metal music [25]. On a dataset representing seven different subgenres, he was able to achieve a maximum performance of 45.7% accuracy using the AdaBoost algorithm on features selected by their correlation. In another study [20] in 2016, they performed automatic classification on a dataset with both genre and subgenre labels. When they ran models on their dataset, their results showed that subgenre classification only reached 31% and 16% for neural network and KNN classifiers respectively, compared to 50% and 39% for genre classification. When comparing genre recognition studies to subgenre recognition studies, it appears that subgenre recognition generally yields lower classification accuracy. This makes sense due to the variability of the features within a genre being much less than between genres. We can expect that the country music classifier will have similar performances to other attempts at automatic subgenre classification.

3. METHODS

3.1 Dataset Creation

A completely new dataset was created in order to train the machine-learning models for this project. This dataset consisted of exactly 1000 songs from four different subgenres of country music. Each song is represented by 40 MFCCs and one subgenre classification label.

3.1.1 Subdivision of Country Music

The division of country music, or any kind of music, into subgenres can get complicated quickly. The differences between subgenres can be unclear with overlap. For the sake of simplicity, I have divided country music into four subgenres: bluegrass, country pop, country rock, and traditional. These four subgenres are the most prevalent of the country music subgenres and cover virtually all country music songs. Technically, there can be songs that contain elements of multiple of these subgenres, but again, for simplicity, I approached the problem as if a song only has one correct subgenre (the single subgenre that the song best represents). The paragraphs below detail the qualitative distinguishing characteristics of each subgenre based on Carney's analysis [5] and general attributes of the songs from the Spotify data. It is important to know the similarities and differences between these subgenres to have a better understanding of the results of the machine learning models.

3.1.1.1 Bluegrass

Bluegrass music emerged in the 1930s and is highly characterized by string instruments such as the mandolin, guitar, fiddle, banjo, and stand-up bass. The strings are commonly picked, and the vocals are a high-pitched and solo sound. Bill Monroe was the pioneer of bluegrass and the style was named after his band, The Blue Grass Boys. Other artists, mostly in the mountains and hills of the Southeastern U.S. popularized the style.

3.1.1.2 Country Pop

In the 1950s, with the growth of rock and roll, artists and producers were pressured to conform their sounds to be suitable for wider audiences. A new kind of country music, country pop, was created in Nashville, Tennessee and became known as the "Nashville Sound" [5]. This Nashville Sound took old elements of country but replaced the harsher stringed instruments with smoother sounds like background vocals and the electric guitar and piano. One thing to note about pop country is that it varies more than the other subgenres because it changes over time. As pop music changes over the decades, country pop reflects these changes. While country pop adopted a rock 'n' roll sound in the 1950s, it has adopted more of an electronic and hip-hop feel in the 2000s [26].

3.1.1.3 Country Rock

In the 1970s, Austin was a crossroads for rock and country [5], and this new country rock started to gain national recognition. Bands like the Eagles took a spin on rock music by producing it with country themes. The pedal steel guitar played an especially significant role in creating this new musical feel. Over time, the country rock feel has come to include elements of heavier rock too with emphasis on electric guitar and drums.

3.1.1.4 Traditional

Traditional country music is the root of many of the other types of country music. It started in rural areas in the Southern U.S. in the eighteenth and nineteenth centuries, focusing on singing about the life of working-class Americans. Until the 1920s, it wasn't popularized outside of the South. Musically, it is characterized by a combination of string instruments such as the banjo, fiddle, and one or more guitars, and being culturally southern, the vocals have a distinct southern accent. In the second half of the 20th century, artists like George Strait and Alan Jackson helped traditional country music make a comeback as part of the neotraditional country music movement [27]. The movement was an attempt to get back to its roots after pop country had begun to dominate the country music industry. Other well-known styles of country like honky tonk and

Texas country are included as part of this subgenre [28]. As neotraditional country songs are more accessible on music streaming platforms, a heavy amount of neotraditional songs were used to train the machine learning models in this study.

3.1.2 Spotify API

The data to populate the model was harvested using the Spotify API [29]. Using Python, I connected to the Spotify API and obtained song data through the API. Basic metadata for a song such as song title, artist, and unique id was extracted for several thousand songs before narrowing the dataset down to 1000 songs using the method described below. A link to a 30-second audio sample from the given song is available as well. This mp3 audio file is the key element in training my machine-learning model to hear and classify country music.

3.1.3 Voting System for Determining Labels

Interestingly enough, Spotify does not make genres or subgenres of songs readily available through the API. Therefore, to determine genre labels for each of the songs in the database, I had to devise a voting system using user-generated playlists on the Spotify app. First, I selected a number of playlists that were clearly marked as part of a certain country subgenre in the playlist title or description. For example, a playlist named "Country Pop Hits" would be considered for identifying songs labeled as country pop. A playlist with the description "bluegrass songs that you will love" would be considered for identifying songs labeled as bluegrass. For each subgenre, I selected ten to twenty different playlists and used the Spotify API to make a cumulative Pandas data frame of all of the songs for the given subgenre. Then, I utilized a voting type of system to determine the songs that would go into the database. For a song to make it into the database, it had to meet two different criteria.

1. The song must belong to two or more of the selected user-curated playlists I have selected.
2. The song must be unique to the subgenre. Songs that were found to belong to multiple subgenres were excluded from the database.

The top 250 most frequently occurring songs for each subgenre were included in the final database. Four subgenres with 250 songs each yield a final count of 1000 songs. Each song was represented by two fields: one class label (subgenre) and one 30-second mp3 audio clip. Having an evenly distributed dataset in terms of classes is important in machine learning because it prevents the models from becoming biased toward the majority class by relying on the frequencies of the samples. Instead, the model can rely on the characteristics of the dataset that are actually important for predicting the class.

3.1.4 MFCC Feature Extraction

To compress the mp3 data and make it more interpretable to the computer, I processed each 30-second audio clip into 40 MFCCs. Mel-Frequency Cepstral Coefficients or MFCCs represent the rate changes in the spectrum bands in an audio signal. Figure 1 and Figure 2 visualize these MFCC values for two different songs. These values are then summarized into the machine learning model by taking the mean. Negative values represent spectral energy concentrated at high frequencies while positive values represent spectral energy concentrated at low frequencies. MFCCs have been proven to work well for speech recognition [30] and genre recognition purposes [3].

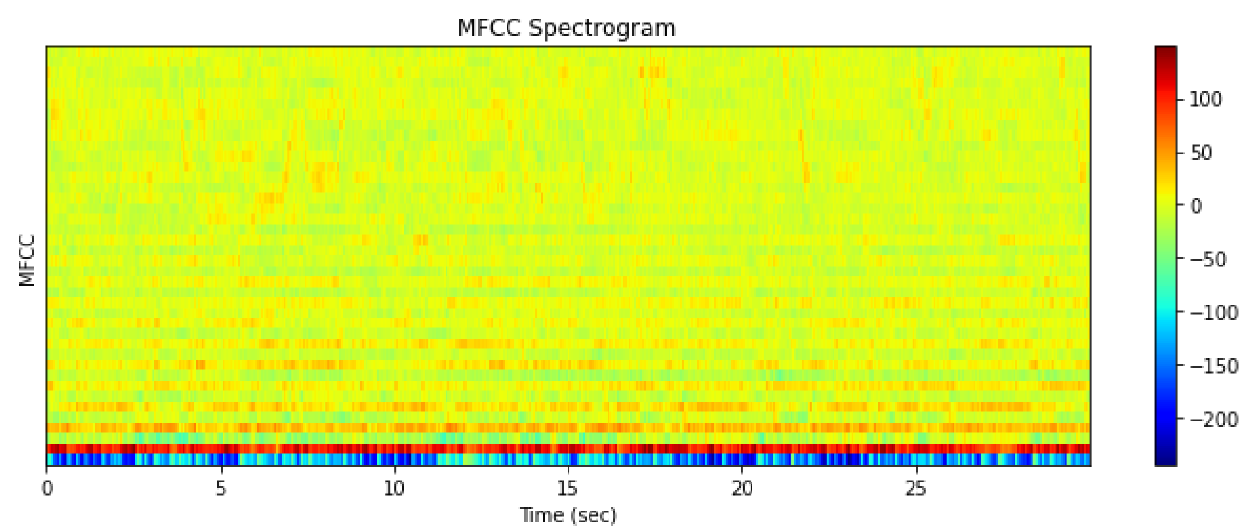

Figure 1: MFCC spectrogram of a 30-second clip of Amarillo By Morning by George Strait (Traditional Country).

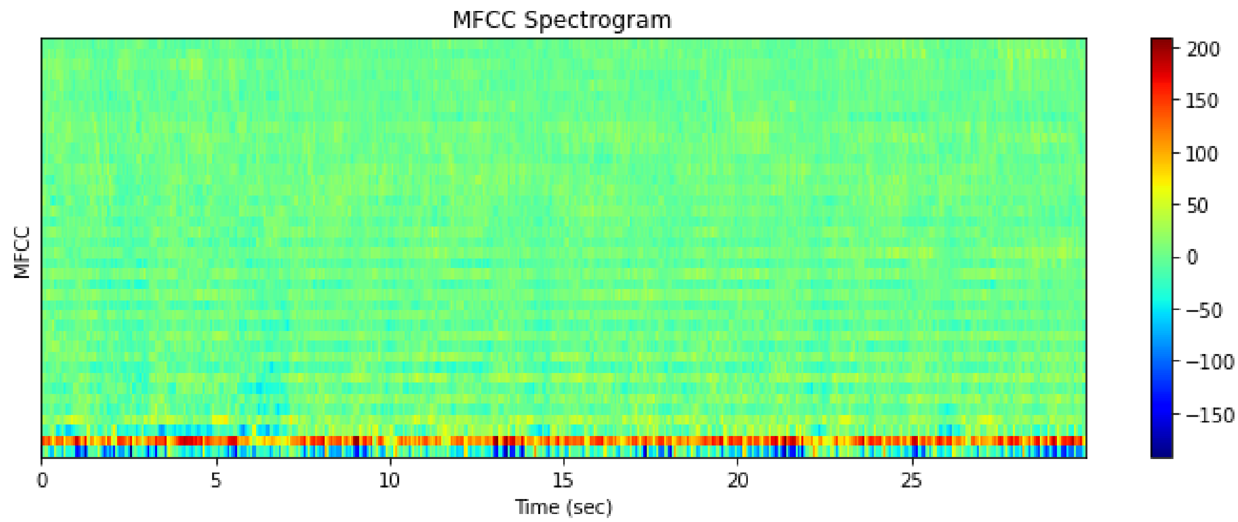

Figure 2: MFCC spectrogram of a 30-second clip of You Proof by Morgan Wallen (Country Pop).

3.2 Machine Learning Algorithms

Three different machine learning algorithms were trained with the data from the Spotify dataset. These consisted of the neural network, random forest, and K-nearest neighbors models. The train-test split was 80:20, which means that 80% of the dataset was used to train the model and the other 20% was used to evaluate it. Since the dataset consisted of 4 subgenres with 250 songs each, there were about 200 songs from each subgenre used to train the models and about 50 songs from each subgenre used to evaluate the models.

3.3 System

All of the code for this project was written in the Python language and executed on the Google Colab online platform. First, I used the Spotipy library [31] to connect to the Spotify API and create my database. Next, I used the Pandas [32] and Librosa [33] libraries to manage the data and convert audio into MFCCs. Lastly, the Tensorflow library [34] was used to train the neural network, and the Scikit-learn library [35] was used to train the random forest and KNN models.

4. RESULTS

For each model, I constructed a confusion matrix that represents how accurately the model performed on classifying the test data. The vertical axis of the confusion matrix represents the true labels according to the classifications in the dataset. The horizontal axis represents the predicted labels that the machine learning model has output based on its analysis of the MFCC features. Usually, confusion matrices contain count values and can be used to evaluate how many songs were either classified correctly or incorrectly, and if so, which subgenres they were misclassified as. However, I have normalized the confusion matrices below to give an idea of the percentages of songs that were classified/misclassified according to the four subgenres. Darker areas represent higher percentages while lighter areas indicate lower percentages, with the diagonal running between the top-left and bottom-right corners representing correct classifications.

Table 1: Performance comparison of the machine learning models.

Accuracy		
Neural Network	Random Forest	KNN
66.5%	70%	65.5%

4.1 Neural Network

The neural network achieved an overall accuracy of 66.5% (Figure 3). The model consists of 3 hidden layers with rectified linear unit (ReLU) activation functions. The layers also have varying dropout rates which prevent overfitting of the model. The final output layer has a Softmax activation function which is useful for multiclass prediction. The softmax outputs the probability distribution for the given audio data belonging to each subgenre of country music. We then determine the class with the highest probability to be the predicted subgenre for the song.

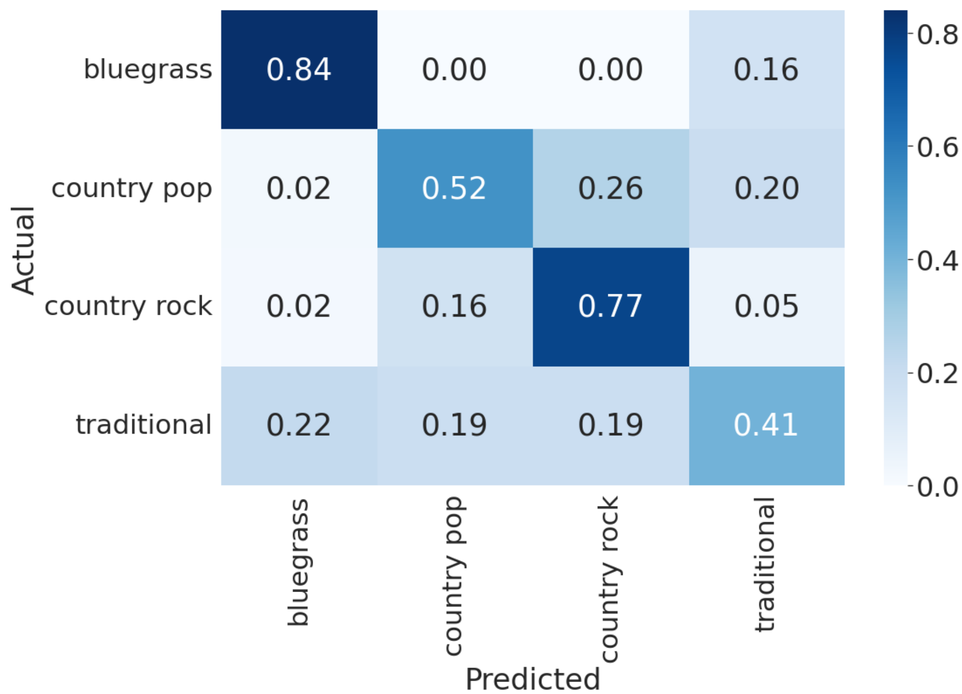

Figure 3: Confusion matrix for the neural network model. The neural network performed well on classifying the bluegrass and country rock subgenres. However, it performed moderately when classifying country pop and traditional country.

4.2 Random Forest

The overall accuracy for the random forest was 70% (Figure 4). This was the highest accuracy that was achieved of the three different models. 256 estimators (decision trees) were used in the model with a maximum tree depth of 20. A higher number of decision trees tends to make the model more accurate [36] as each tree is trained on a different subset of data and factoring in multiple decision trees produces a more stable prediction system. This comes at the expense of performance time, but since the dataset is relatively small, it only took a couple of minutes to train even with tuning the number of estimators and maximum depth hyperparameters.

4.3 KNN

The K-Nearest Neighbors model achieved a 63.5% accuracy which was the lowest of the models (Figure 5). The number of neighbors hyperparameter was tuned to 36, which means that

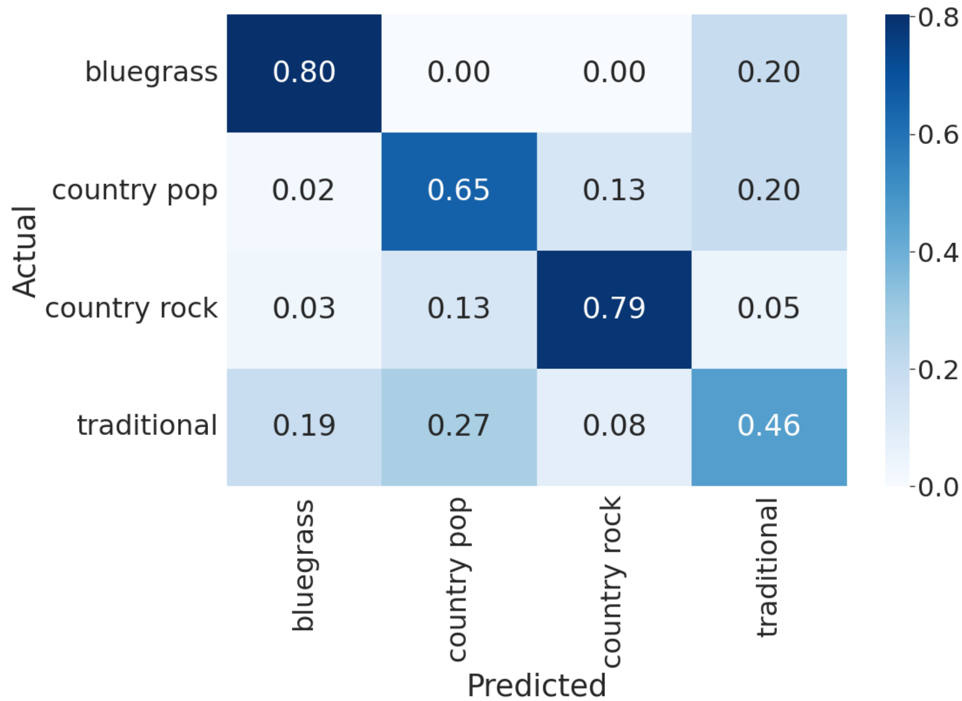

Figure 4: Confusion matrix for the random forest model. The random forest had the best overall performance of all of the models. It had good accuracy for the bluegrass, country pop, and country rock subgenres, but performed moderately for traditional country.

an unknown value is determined by the majority subgenre of its 36 known neighbors which have the most similar MFCC values.

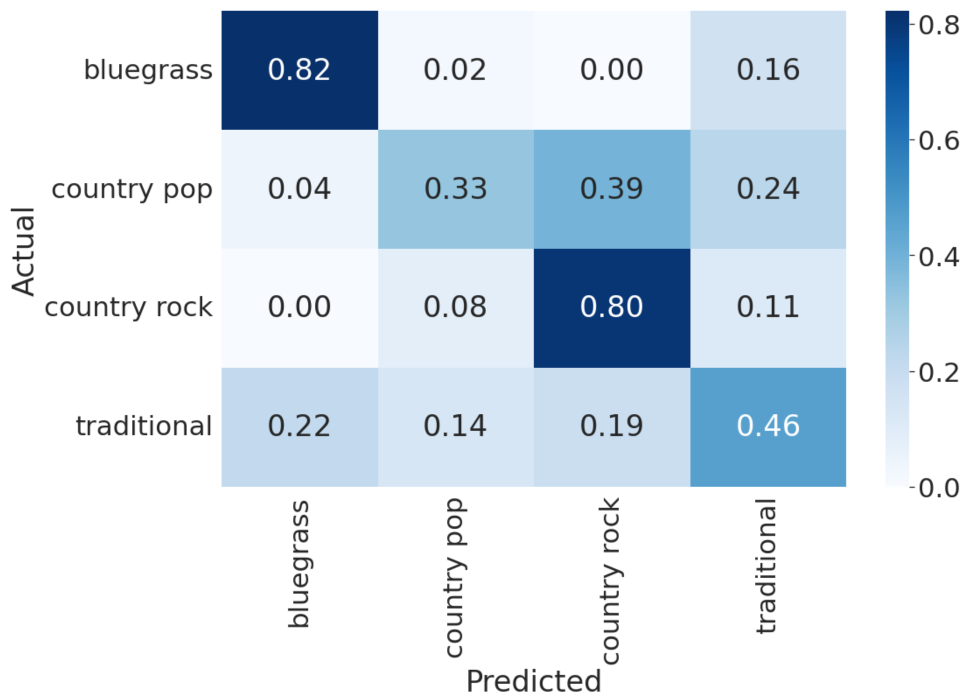

Figure 5: Confusion matrix for the KNN model. The KNN performed the lowest of the three models. It had good performance for classifying bluegrass and country rock while performing moderately for traditional country. It was especially weak at classifying country pop, which it commonly confused with country rock and traditional country.

5. DISCUSSION

5.1 Analysis of Results

Although the random forest had the best performance of the three models, all three models performed similarly in terms of their overall accuracy (all within 10%). If we examine the confusion matrices between the models, we see that there are multiple common trends between the classification tendencies of the models. For one, bluegrass has the highest percentage of correct classifications (all above 80%). Bluegrass is heavily characterized by distinct plucking sounds of the banjo and fiddle which may make it easily identifiable to the machine learning models. A close second for the highest accuracy in all three models was the country rock subgenre. 77-80% of country rock songs were classified correctly. Like bluegrass, country rock is characterized by very distinct sounds such as the electric guitar and drums. This may contribute to the high accuracy of classification for country rock.

Interestingly enough, bluegrass and country rock were also seldom confused with each other in the model. In no model was bluegrass ever misclassified as country rock, and there were few cases where country rock was misclassified as bluegrass. If we imagine a kind of spectrum of country music subgenres, bluegrass and country rock sounds would be on opposite ends.

As for the classifications of the country pop and traditional country subgenres, the accuracies were significantly worse. In fact, the only case where a subgenre had a greater amount of misclassifications than correct classifications was in the case of country pop for the KNN. Only 33% of country pop songs were classified correctly while 39% of country pop songs were classified as country rock. In the other models, it also appears that country pop was commonly confused with country rock, possibly because there are similar electronic elements between the two subgenres. In the context of the history of country pop and country rock, both were popularized around the same time in the 60s and 70s [5] and were influenced by mainstream rock 'n' roll. This might be a key factor in the high misclassifications between the two subgenres. On the other hand, traditional country tended to have a more even distribution of incorrect classifications over the other subgenres. This could be attributed to the fact that traditional country music was prevalent across the Southern United States before most of the other subgenres, and the other subgenres have evolved alongside it [5]. Therefore, there are a lot of basic elements of traditional country music that can be found throughout the other subgenres of country.

5.2 Comparison to Previous Research

When compared to previous research in the area of music genre recognition, this study fares extremely well. Of course, compared to music genre recognition models that have been trained on the GTZAN dataset which reach accuracies upwards of 80-90%, the country subgenre model does not seem to compare. However, the difference is that the GTZAN dataset is a much broader dataset where the individual classes (rock, classical, country, etc) have significantly different sounds and are easier to differentiate. When we examine subgenres within country music, we are looking at a much narrower range of styles, which is a more difficult recognition task for both humans and computers.

We can compare, however, the performance of the country subgenre recognition models to other studies which focus on subgenre recognition. One notable study from 2011 [25] obtained a 45.7% accuracy on subgenre classification of heavy metal music. Another study from 2016 obtained a 31% accuracy on various subgenres from a multi-genre dataset. From these two studies, we can see that 70% for subgenre recognition is significantly better than past research.

This study's relatively high accuracy may be attributed to a number of reasons. First, feature extractions of MFCCs and using the random forest classifier have proved to be extremely effective in past music genre recognition as well as within this study. Until more effective methods are developed, MFCCs and random forest models are good choices for music genre recognition problems. Second, the methods of creating the database from Spotify may be an effective way of building a good dataset that can train machine learning models well. Not only are the feature extractions and ML models important to obtaining accurate classifiers, but the quality of the underlying data is extremely important. The databases in the other research explored earlier in the paragraph were different but still meticulous in their song selection methods. There are improvements that can be made, however, specifically by eliminating possible accuracy inflation due to the "artist effect". I elaborate more on this in the future work section. Third, the country music subgenres prove to be diverse in sound although they all fall under the umbrella genre of country music.

6. FUTURE WORK

6.1 Improvement of Methods

With machine learning, there are many processes and steps along the way that can be adjusted, so there are many points in this research that can be extended or improved in the future.

6.1.1 Database Creation

The method for obtaining data used a voting kind of system to determine the subgenre labels of songs. Although this was fairly accurate, crowdsourced subgenre data from Spotify might not be the most reliable source for labeling songs. Additionally, since only the most frequently seen songs were put into the database, there is a level of popularity required for all the songs that were used to train the models. This might not be a great wholistic representation of a subgenre because there are many sources for music than just the top, professionally produced songs on Spotify. To make the model more versatile, it would be good to add data from a variety of sources including but not limited to concert recordings, home-produced music, and songs from other databases such as the Library of Congress. The difficulty with getting this data, however, is that it cannot simply be accessed using an API, and it may be hard to obtain in a format where we know the exact subgenre labels.

6.1.2 Features

There are a number of parameters within the data that can be changed to improve the performance of the models. A few are listed below.

6.1.2.1 Number of MFCCs

40 MFCCs were used in feature extraction from the data, but we could probably filter down or increase the number of MFCCs to include more relevant features to the classification problem, therefore increasing the accuracy of the model.

6.1.2.2 Audio Clip Length

All of the audio clips from Spotify were 30 seconds long. At least from Spotify, it is not possible to retrieve longer clips. However, it would be interesting to see how longer or shorter clips affect the performance of the models. Realistically, a computer application using the model trained in this research would listen to an entire song before labeling it as a given subgenre on Spotify or a given music database. However, if the model was trained to only require listening to a few seconds of audio to determine the subgenre, this could save significant computational time and be used in cases where the entire song data is not available.

6.1.3 Artist Effect

One potential weakness in my model is the introduction of artist bias. The "artist effect" or "album effect" states that a classifier is going to achieve better results if the same artist is included in both the training and test data. This is exaggerated when multiple songs from the same artist are included in the test data. Essentially, the model is being trained to classify based on the artist's voice and not the other elements of the music. Luckily, the artist effect may not be as much of a problem for this study, as other research has shown that classifier performance was only affected by 0.06 as opposed to upwards of 0.2 for non-western music [37]. Nevertheless, it is good to remove this bias if we want to have the best model possible.

One way to reduce the artist effect could be to ensure that there are no shared artists between the training and test data. The reported performance accuracy will not be affected by the "artist effect" because the model will have never heard the artists that it is being tested on. However, that does not guarantee that the model might be biased toward artists who are more prevalent in the data. A more thorough way to remove the "artist effect" is to have a one-to-one ratio of artists to songs. No artist will be repeated, allowing for no certain artist to dominate and skew the data. The difficulty with this is that certain artists tend to naturally dominate subgenres because they are either pioneers of the subgenre or have gained large influence as an artist. The most popular examples of a song within a subgenre are generally going to be limited to the most influential artists. The method I used of extracting data from Spotify did not work well to eliminate the artist effect. Unfortunately, the basic premise of it was influenced by user preferences and the "most popular" songs. Still, getting 1000 songs using the automated voting process was tough, so I imagine that creating a database with 1000 songs and 1000 different artists would be a tedious but beneficial process. It would be necessary and beneficial to seek data from other sources as well as opposed to just Spotify. During my public presentation, I received feedback that the Library of Congress would be a good place to sample more music. Additional sources that could be useful are homemade or concert recordings from SoundCloud [38] and YouTube [39].

6.2 Areas for Expansion

Due to time constraints, I was not able to accomplish everything I wanted to in my research. In this section, I will detail future experiments and potential areas of expansion for my research.

6.2.1 Human Subject Research

One future experiment is the comparison of the performance of my machine learning model to the performance of human subjects. As we saw from the test data results, we would expect the model to achieve about 70% accuracy on real-world data. But how would this stack up against real people? In this future research, I would identify three kinds of human subjects using a preliminary survey. The first group would claim to be very familiar with country music and consider themselves "experts" at the subject. A second group would know of country music but not well-versed in it. A third group would have very little to no knowledge of country music at all. Once sufficient numbers of people were found for each group, I would give each of them a test of 10 different songs, each belonging to one of the four subgenres of country music. To be on par with how the ML model is tested, each song clip would be 30 seconds long, and after each clip is played, the participant would be asked to identify which of the four subgenres of country music the song belongs to. This participant data would give us a good gauge of how the ML model performs compared to humans. It would show how "expert" my model actually is, and if it truly excels at the problem it has been trained to solve.

6.2.2 *Expanding Subgenres*

The more in-depth one explores country music, the more complex the categorization of subgenres gets. For the purposes of this study, I categorized country music into four main subgenres. However, Carney [5] cites three additional subgenres including Singing Cowboy, Western Swing, and Honky Tonk. A quick Internet search will reveal several more. It would be interesting to retrain the models to see how they perform with the addition of more subgenres. One foreseeable difficulty would be getting sufficient data for these other subgenres since they are not as prevalent on Spotify playlists. Another difficulty is that most of these additional subgenres overlap or relate closely to the current subgenres in the model. Determining exactly which subgenres should be used as the boundaries in the model is not an easy task.

This leads to the problem of multiple classifications. When adding more subgenres to the model, many songs could correctly fall into two or more of these subgenres. Of the models used in this study, the neural network is best suited for this task since it already calculates probabilities that a given clip falls into each subgenre. In the future, the model could address the complexity of subgenre classification by outputting the "top three subgenres" or something of that nature.

7. CONCLUSION

The subgenres of country music have distinguishing characteristics that experts are able to differentiate and classify just by listening to a song. Each subgenre has developed over the years in a unique way and has different bases of listeners who prefer it. While hearing and differentiating between subgenres of country music has been more of a human-based, qualitative task in the past, I have developed a model that takes a quantitative approach to subgenre recognition.

In this research, I trained a computer to take in 30-second audio clips and learn the difference between four subgenres of country music: bluegrass, country pop, country rock, and traditional. A country music dataset was extracted from Spotify and MFCC features were used to make the audio understandable to the computer. Of the neural network, random forest, and KNN models that were trained using the dataset, the random forest had the best performance. Bluegrass and country rock appeared to have the most unique sound within the genre of country music, while the sounds of country pop and traditional, while still relatively distinct, blended into other subgenres.

This paper demonstrated that there is a diversity of subgenres even within the genre of country music, and the differences in sound can be recognized by machine learning models. This is reflective of society today where although two listeners can both enjoy country music, they can have opposite opinions on whether they like a given song due to subgenre preference. Having a model that can accurately differentiate subgenres of country music not only saves time that experts might usually spend, ‘but it also helps artists who generate new content categorize their music and cater to a specific audience.

REFERENCES

[1] S.-Y. Chou, Y.-H. Yang, and Y.-C. Lin, "Evaluating music recommendation in a real-world setting: On data splitting and evaluation metrics," in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2015.

[2] A. Germain and J. Chakareski, "Spotify me: Facebook-assisted automatic playlist generation," in *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 025–028, 2013.

[3] Z. Qi, M. Rahouti, M. A. Jasim, and N. Siasi, "Music genre classification and feature comparison using ml," in *2022 7th International Conference on Machine Learning Technologies (ICMLT)*, ICMLT 2022, (New York, NY, USA), p. 42–50, Association for Computing Machinery, 2022.

[4] J. Dyck, "The aesthetics of country music," *Philosophy Compass*, vol. 16, no. 5, p. e12729, 2021.

[5] G. O. Carney, "Country music and the south: A cultural geography perspective," *Journal of Cultural Geography*, vol. 1, no. 1, pp. 16–33, 1980.

[6] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[7] B. L. Sturm, "An analysis of the gtzan music genre dataset," in *Proceedings of the Second International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies*, MIRUM '12, (New York, NY, USA), p. 7–12, Association for Computing Machinery, 2012.

[8] V. Prashanthi, S. Kanakala, V. Akila, and A. Harshavardhan, "Music genre categorization using machine learning algorithms," in *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, pp. 1–4, 2021.

[9] A. Ghildiyal and S. Sharma, "Music genre classification using data filtering algorithm: An artificial intelligence approach," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 922–926, 2021.

[10] R. Sarkar and S. K. Saha, “Music genre classification using emd and pitch based feature,” in *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*, pp. 1–6, 2015.

[11] M. Haggblade, Y. Hong, and K. Kao, “Music genre classification,” 2011.

[12] M. Defferrard, S. P. Mohanty, S. F. Carroll, and M. Salathé, “Learning to recognize musical genre from audio: Challenge overview,” in *Companion Proceedings of the The Web Conference 2018*, WWW ’18, (Republic and Canton of Geneva, CHE), p. 1921–1922, International World Wide Web Conferences Steering Committee, 2018.

[13] T. Li, M. Ogihara, and Q. Li, “A comparative study on content-based music genre classification,” in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, SIGIR ’03, (New York, NY, USA), p. 282–289, Association for Computing Machinery, 2003.

[14] P. Zhang, X. Zheng, W. Zhang, S. Li, S. Qian, W. He, S. Zhang, and Z. Wang, “A deep neural network for modeling music,” in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, ICMR ’15, (New York, NY, USA), p. 379–386, Association for Computing Machinery, 2015.

[15] M. R. Ansari, S. A. Tumpa, J. A. F. Raya, and M. N. Murshed, “Comparison between support vector machine and random forest for audio classification,” in *2021 International Conference on Electronics, Communications and Information Technology (ICECIT)*, pp. 1–4, 2021.

[16] T. Endo, S. ichi Ito, Y. Mitsukura, and M. Fukumi, “The music analysis method based on melody analysis,” in *2008 International Conference on Control, Automation and Systems*, pp. 2559–2562, 2008.

[17] Y. KIKUCHI, N. AOKI, and Y. DOBASHI, “A study on automatic music genre classification based on the summarization of music data,” in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 705–708, 2020.

[18] F. Rong, “Audio classification method based on machine learning,” in *2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pp. 81–84, 2016.

[19] J. Xu, X. Li, Y. Hao, and G. Yang, “Source separation improves music emotion recognition,” in *Proceedings of International Conference on Multimedia Retrieval*, ICMR ’14, (New York, NY, USA), p. 423–426, Association for Computing Machinery, 2014.

[20] M. Serwach and B. Stasiak, "Ga-based parameterization and feature selection for automatic music genre recognition," in *2016 17th International Conference Computational Problems of Electrical Engineering (CPEE)*, pp. 1–5, 2016.

[21] T. Li and G. Tzanetakis, "Factors in automatic musical genre classification of audio signals," in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, pp. 143–146, 2003.

[22] S. Amiriparian, N. Cummins, M. Gerczuk, S. Pugachevskiy, S. Ottl, and B. Schuller, """are you playing a shooter again?!" deep representation learning for audio-based video game genre recognition," *IEEE Transactions on Games*, vol. 12, no. 2, pp. 145–154, 2020.

[23] J. Hall, W. O'Quinn, and R. J. Haddad, "An efficient visual-based method for classifying instrumental audio using deep learning," in *2019 SoutheastCon*, pp. 1–4, 2019.

[24] B. Murauer and G. Specht, "Detecting music genre using extreme gradient boosting," in *Companion Proceedings of the The Web Conference 2018*, WWW '18, (Republic and Canton of Geneva, CHE), p. 1923–1927, International World Wide Web Conferences Steering Committee, 2018.

[25] V. Tsatsishvili, "Automatic subgenre classification of heavy metal music," *Master's Thesis*, 11 2011.

[26] J. Orosz, """straight outta nashville": Allusions to hip hop in contemporary country music," *Popular Music and Society*, vol. 44, no. 1, pp. 49–59, 2021.

[27] R. Shusterman, "Moving truth: Affect and authenticity in country musicals," *The Journal of Aesthetics and Art Criticism*, vol. 57, no. 2, pp. 221–233, 1999.

[28] D. Lackey, "Cowboy gathering, heart of texas events to present best of traditional country music," *Ranch and Rural Living*, vol. 102, p. 8, 08 2021.

[29] Spotify, "Web api | spotify for developers," 2023.

[30] K. Gupta and D. Gupta, "An analysis on lpc, rasta and mfcc techniques in automatic speech recognition system," in *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, pp. 493–497, 2016.

[31] P. Lamere, "Spotipy - a light weight python library for the spotify web api," 2021.

[32] W. McKinney, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference* (S. van der Walt and J. Millman, eds.), pp. 51 – 56, 2010.

[33] B. McFee, M. McVicar, D. Faronbi, I. Roman, M. Gover, S. Balke, S. Seyfarth, A. Malek, C. Raffel, V. Lostanlen, B. van Niekirk, D. Lee, F. Cwitkowitz, F. Zalkow, O. Nieto, D. Ellis, J. Mason, K. Lee, B. Steers, E. Halvachs, C. Thomé, F. Robert-Stöter, R. Bittner, Z. Wei, A. Weiss, E. Battenberg, K. Choi, R. Yamamoto, C. Carr, A. Metsai, S. Sullivan, P. Friesch, A. Krishnakumar, S. Hidaka, S. Kowalik, F. Keller, D. Mazur, A. Chabot-Leclerc, C. Hawthorne, C. Ramaprasad, M. Keum, J. Gomez, W. Monroe, V. A. Morozov, K. Eliasi, nullmightybofo, P. Biberstein, N. D. Sergin, R. Hennequin, R. Naktinis, beantowel, T. Kim, J. P. Åsen, J. Lim, A. Malins, D. Hereñú, S. van der Struijk, L. Nickel, J. Wu, Z. Wang, T. Gates, M. Vollrath, A. Sarroff, Xiao-Ming, A. Porter, S. Kranzler, Voodoohop, M. D. Gangi, H. Jinoz, C. Guerrero, A. Mazhar, toddrme2178, Z. Baratz, A. Kostin, X. Zhuang, C. T. Lo, P. Campr, E. Semeniuc, M. Biswal, S. Moura, P. Brossier, H. Lee, and W. Pimenta, "librosa/librosa: 0.10.0.post2," Mar. 2023.

[34] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[36] P. Probst and A.-L. Boulesteix, "To tune or not to tune the number of trees in random forest," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6673–6690, 2017.

[37] A. Kruspe, H. Lukashevich, and J. Abeßer, "Artist filtering for non-western music classification," in *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound*, AM '11, (New York, NY, USA), p. 82–87, Association for Computing Machinery, 2011.

[38] SoundCloud, "Stream and listen to music online for free with soundcloud," 2023.

[39] YouTube, "Youtube," 2023.