

# Pengembangan Aplikasi Klasifikasi Sepuluh Genre Musik

Achmad Naufal Ilyasa

Informatika, Fakultas Ilmu Komputer dan Desain, Universitas Kalbis  
Jalan Pulomas Selatan Kav. 22, Jakarta 13210  
Email: naufal.ilyasa7@gmail.com

**Abstract:** This study aims to classify music by genre, which is used to group music by genre to make it easy to search. The data used audio data and then converted into melspectogram images of 10,000 data divided into ten genres. They are blues, classical, country, disco, hiphop, jazz, metal, pop, reggae dan rock. The data validation using the split test with the training and testing data ratio is 9:1. The method used in this study is a Convolution Neural Network (CNN) using the Tensorflow Keras library. This study conducted experiments with three different architectural models to compare and find the best model. Based on the results, the model that produced the best test accuracy value was the CNN\_2 model, with an accuracy value of 81.7% in the test data and the accuracy of the training data of 95.5% by undergoing an epoch of 120 and with a configuration using adam optimizer, loss categorical-cross entropy and learning rate of 0.00005.  
**Keywords:** Convolutional Neural Network, Music, Genre, Classification

**Abstrak:** Penelitian ini bertujuan untuk mengklasifikasikan musik berdasarkan genre yang digunakan untuk mengelompokkan musik berdasarkan genre agar mudah dalam melakukan pencarian. Data penelitian yang digunakan adalah data audio yang diubah menjadi citra melspectogram sebanyak 10.000 data yang terbagi dalam sepuluh genre. Genre tersebut yaitu blues, classical, country, disco, hiphop, jazz, metal, pop, reggae dan rock. Validasi data menggunakan split test yaitu data dibagi menjadi latih dan data uji dengan rasio 9:1. Metode yang digunakan pada penelitian ini adalah Jaringan Syaraf Konvolusi menggunakan pustaka Tensorflow Keras. Penelitian ini melakukan percobaan dengan tiga model arsitektur yang berbeda guna untuk membandingkan dan mencari model terbaik. Berdasarkan hasil penelitian, model yang menghasilkan nilai akurasi uji terbaik adalah model CNN\_2 dengan nilai akurasi sebesar 81,7% pada data uji dan akurasi data latih sebesar 95,5% dengan menjalani epoch sebesar 120 serta dengan konfigurasi menggunakan adam optimizer, loss categorical-crossentropy dan learning rate sebesar 0,00005.

**Kata kunci:** Jaringan Saraf Konvolusi, Musik, Genre, Klasifikasi

## I. PENDAHULUAN

### A. Latar Belakang

Salah satu media hiburan populer di era *digital* adalah musik. Musik adalah hasil kreatifitas manusia dalam mengekspresikan ide dan emosi dengan membentuk kumpulan suarayang terdiri dari melodi, harmoni dan ritme [1]. Pada era *digital* ini jumlah musik digital pada *database* layanan *streaming* musik sudah sangat besar dan ditambah musik-

musik keluaran terbaru yang akan terus bertambah setiap harinya. Berdasarkan artikel berjudul “*Spotify Revenue and Usage Statistics (2022)*” oleh Iqbal menunjukkan jumlah musik pada layanan *streaming* musik Spotify berjumlah 70 juta dan lebih dari 60.000 musik rilis setiap harinya [2]. Sehingga, dibutuhkan pengorganisasian musik-musik *digital* berdasarkan artis, album atau berdasarkan *genre*.

Klasifikasi genre musik merupakan

salah satu cara untuk mengorganisasi *database* musik *digital*. *Database* musik *digital* yang sudah diatur berdasarkan genre musik akan memudahkan pendengar musik di layanan *streaming* musik tersebut untuk menemukan musik yang sesuai dengan preferensinya. Maka, penelitian ini dibuat untuk melakukan klasifikasi genre yang kemudian ketika sudah melakukan klasifikasi maka pengguna dapat melakukan pengelompokan musik berdasarkan genre, pengelompokan ini berguna untuk mempermudah pengguna menemukan musik yang disukai atau ingin mendengarkan musik genre tertentu saja.

Terdapat penelitian terdahulu untuk melakukan pengklasifikasian pada genre musik, yaitu Pengklasifikasian Genre Musik Indonesia Menggunakan *Convolutional Neural Network* (CNN) oleh Wairata dkk. [3]. Penelitian Wairata bertujuan untuk mengoptimalkan tingkat akurasi menggunakan metode CNN, dengan cara mengkonversi data musik menjadi gambar atau yang dapat disebut sebagai *spectrogram*. Penelitian ini mengkategorikan tiga genre musik yakni; Dangdut, *Jazz* dan *Pop*. Kemudian didapatkan hasil tingkat akurasi tes sebesar 81,33%. Penelitian lainnya adalah penelitian yang dilakukan oleh Sarofi dkk. [4]. Yang menggunakan metode *random forest*, dengan *Mel-Frequency Cepstral Coefficient* (MFCC) sebagai fitur ekstraksi karena mampu mengadaptasi pendengaran manusia. Menggunakan lima genre musik yang berbeda yaitu *disco*, *hiphop*, *jazz*, *pop* dan *reagge*. Penelitian ini memperoleh tingkat akurasi tes 88,83%. Selain itu, penelitian lain yang dilakukan oleh Giri yang berjudul *Klasifikasi Musik Berdasarkan Genre dengan Metode K-Nearest Neighbor* [5]. Menggunakan metode *k- nearest neighbor* dan menggunakan sembilan genre musik yakni; *punk*, *classical*, *rap*, *metal*, *EDM*, *hip-hop*, *R&B*, *pop*, dan *rock*. Juga,

menggunakan sebelas fitur musik yakni; *mode*, *acoustic-ness*, *danceability*, *tempo*, *valence*, *loudness*, *instrumentalness*, *energy*, *liveliness* dan *speechiness*. Penelitian ini mendapatkan hasil tingkat akurasi klasifikasi sebesar 44,8%. Penelitian yang terkait dengan metode CNN adalah penelitian tentang perbandingan akurasi klasifikasi genre dari berbagai metode yakni; *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Feed-forward Neural Network* (FNN) dan CNN oleh Huang [6]. Menggunakan sepuluh genre musik dan mengubah audio menjadi *mel-spectrogram*. Penelitian ini menghasilkan tingkat akurasi tes tertinggi didapatkan pada metode CNN sebesar 82%. Dari penelitian ini didapatkan bahwa metode CNN memiliki tingkat akurasi tes tertinggi dibanding ketiga metode yang dijadikan perbandingan. Dari penelitian terdahulu terlihat masih memiliki kekurangan yaitu hasil tingkat akurasi tes masih kurang memuaskan dan genre yang diklasifikasikan terbilang sedikit. Maka dari itu, pada penelitian ini akan mengembangkan dari penelitian- penelitian terdahulu yaitu dengan menggunakan metode CNN. Tidak hanya itu, pada penelitian ini akan mengklasifikasikan sepuluh genre musik yang berbeda.

Berdasarkan latar belakang dan penelitian terdahulu, maka dibuatlah penelitian Pengembangan Aplikasi Klasifikasi Sepuluh Genre Musik. *Dataset* yang digunakan adalah *dataset GTZAN*, *dataset* ini menyediakan seratus audio *clip* berdurasi 30 detik per satu genre dari sepuluh genre musik [7]. Lalu, mengubah data *audio clip* tersebut menjadi citra *mel spectrogram* sehingga, data tersebut dapat diklasifikasikan menggunakan *library CNN*. Kemudian, dibuatkan *graphical user interface* (GUI) untuk memudahkan pengguna untuk klasifikasi genre musik.

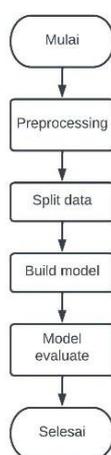
Berdasarkan dari latar belakang

penelitian ini, maka dibuatlah perumusan masalah yaitu, bagaimana hasil tingkat akurasi dari penerapan menggunakan metode CNN pada klasifikasi genre musik?

Berdasarkan rumusan masalah tersebut maka tujuan dari penelitian ini adalah mengetahui hasil tingkat akurasi dari penerapan menggunakan metode CNN pada klasifikasi genre musik.

## II. METODE PENELITIAN

### A. Pembuatan Model



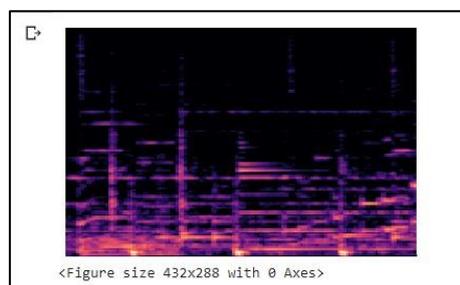
Gambar 3 Flowchart pembuatan model

Berdasarkan Gambar 3 merupakan desain *flowchart* dalam pembuatan model klasifikasi genre musik. Sebelum masuk tahapan *preprocessing* data yaitu melakukan persiapan *dataset* dengan membagi *dataset* GTZAN yang sebelumnya berjumlah 1000 berdurasi 30 detik menjadi sepuluh bagian sehingga, total data menjadi 10.000 *audio clip* berdurasi tiga detik.

#### 1. Preprocessing

Tahap pertama dalam proses *preprocessing* adalah menghasilkan *dataset* berjumlah 10.000. Cara yang dilakukan untuk menghasilkan *dataset* sebanyak 10.000 dari sebelumnya yang berjumlah hanya seribu data *audio* yaitu, dengan membagi setiap *audio* berdurasi

30 detik menjadi tiga detik sehingga, setiap *audio* berdurasi 30 detik menghasilkan sepuluh *audio*. Maka, *dataset* yang sebelumnya berjumlah seribu *audio* berdurasi 30 detik menjadi 10.000 *audio* berdurasi tiga detik. Pembagian ini dilakukan agar menghasilkan lebih banyak data sehingga, pelatihan model akan menghasilkan akurasi yang lebih tinggi. Setelah melakukan pembagian data, dilakukan perubahan data dari *audio* menjadi citra *mel spectrogram*.



Gambar 4 Hasil dari perubahan data audio menjadi *mel spectrogram*

Gambar 4 merupakan *spectrogram* dimana frekuensi (Hz) yang terkandung pada *spectrogram* diubah menjadi *mel scale* (Mel) sehingga, perubahan tersebut hasilnya disebut sebagai *mel spectrogram*. *Spectrogram* adalah visualisasi rentang spektrum frekuensi yang dikandung dari *audio*. Ketika warna frekuensi yang dihasilkan lebih cerah pada citra maka, suara yang dihasilkan lebih pekat sedangkan, ketika warna yang dihasilkan lebih gelap maka, suara yang dihasilkan lebih redup. Kemudian, *Mel scale* adalah satuan unit suara dimana suara dengan jarak yang sama pada skala *mel* dianggap memiliki jarak yang sama dengan pendengar suara.

#### 2. Split Data

Tahap selanjutnya adalah pembagian data *train* dan *test* dengan cara memindahkan seratus citra pada setiap

genre secara *random* ke direktori *test*. Setelah memindahkan citra ke direktori *test* maka, total keseluruhan data citra pada direktori *test* terdapat seribu citra. Dari hasil tersebut pembagian data *train* dan data *test* dengan perbandingan 9:1 yaitu data *train* sebanyak 9.000 citra dan data *test* sebanyak 1.000 citra.

### 3. Build Model

Dalam penelitian ini akan digunakan tiga model arsitektur yang berbeda

	Normalizat ion		
Max pooling	2 x 2	2 x 2	
Convolution 32	3 x 3	1 x 1	3
Max pooling	2 x 2	2 x 2	
Convolution 64	3 x 3	1 x 1	3
Max pooling	2 x 2	2 x 2	
Convolution 128	3 x 3	1 x 1	-1

untuk mencari model arsitektur terbaik. Tiga model yang digunakan adalah sebagai berikut.

Tabel 1 Model Arsitektur CNN\_1  
Layer Kernel Stride Batch

	Normaliza tion		
Input (citra melspectogram 432 x 288)			
Convolution 8	3 x 3	1 x 1	3
Convolution 8	3 x 3	1 x 1	3
Max pooling	2 x 2	2 x 2	
Convolution 16	3 x 3	1 x 1	3
Convolution 16	3 x 3	1 x 1	3
Max pooling	2 x 2	2 x 2	
Convolution 32	3 x 3	1 x 1	3
Convolution 32	3 x 3	1 x 1	3
Max pooling	2 x 2	2 x 2	
Convolution 64	3 x 3	1 x 1	-1
Convolution 64	3 x 3	1 x 1	-1
Max pooling	2 x 2	2 x 2	
Flatten			
Dropout 0,5			
Dense (10, softmax)			

Tabel 1 adalah model arsitektur

CNN\_1 yang disusun dengan pola dua *convolution layer* dan satu *pooling pooling*. Model arsitektur ini menggunakan *adam optimizer* dengan *learning rate* sebesar 0,00005, menggunakan loss function *sparse categorical cross-entropy* dan dengan 70 *epoch*.

Tabel 2 Model Arsitektur CNN\_2  
Layer Kernel Stride Batch

	Norma lizatio n		
Input (citra melspectogram 432x 288)			
Convolution 8	3 x 3	1 x 1	3
Max pooling	2 x 2	2 x 2	
Convolution 16	3 x 3	1 x 1	3

#### Layer Kernel Stride Batch

Tabel 2 adalah model arsitektur CNN\_2 disusun dengan pola satu *convolution layer* dan satu *pooling layer* namun, *convolution layer* hanya sampai *convolution 128*. Model arsitektur ini menggunakan *adam optimizer* dengan *learning rate* sebesar 0,00005, menggunakan loss function *categorical cross-entropy* dan dengan 120 *epoch*.

Tabel 3 Model Arsitektur CNN\_3

	Normal ization		
Input (citra mel spectogram 432 x 288)			
Convolution 8	3 x 3	1 x 1	3
Max pooling	2 x 2	2 x 2	
Convolution 16	3 x 3	1 x 1	3
Max pooling	2 x 2	2 x 2	
Convolution 32	3 x 3	1 x 1	3
Max pooling	2 x 2	2 x 2	
Convolution 64	3 x 3	1 x 1	-1
Max pooling	2 x 2	2 x 2	

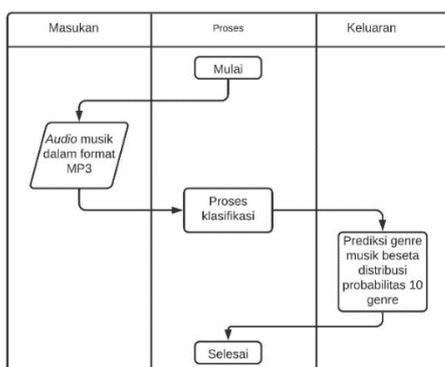
<i>Flatten</i>
<i>Dropout 0,5</i>
<i>Dense (10, softmax)</i>

Tabel 3 adalah model arsitektur CNN\_3 yang disusun dengan pola satu *convolution layer* dan satu *pooling layer* namun, *convolution layer* hanya sampai *convolution 64*. Model arsitektur ini menggunakan *adam optimizer* dengan *learning rate* sebesar 0,00005, menggunakan *loss function categorical cross-entropy* dan dengan 120 *epoch*.

### 3. Model Evaluate

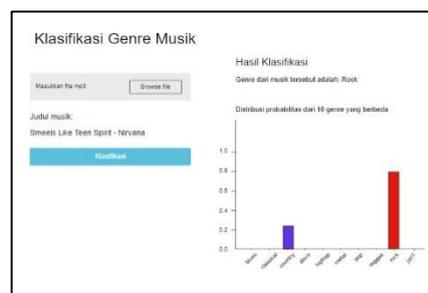
Terakhir, evaluasi model menggunakan *confusion matrix* dan dilakukan pencarian nilai akurasi, *recall*, presisi serta *F-score*.

### B. Pembuatan GUI



Gambar 5 Flowchart proses klasifikasi pada GUI

Berdasarkan Gambar 5 desain alur program dimulai dari pengguna memasukkan audio yang ingin diklasifikasikan dalam *format MP3*. Kemudian, dilakukan proses klasifikasi menggunakan model terbaik yang sudah dibuat pada tahap *increment* satu. Selanjutnya program akan mengeluarkan hasil prediksi genre musik serta distribusi probabilitas dari sepuluh genre.



Gambar 6 Mockup GUI klasifikasi genre musik

Gambar 6 adalah *mockup* GUI klasifikasi genre musik, terdapat fungsi untuk memasukkan *file audio* dengan *format MP3*. Kemudian, ketika berhasil memasukkan file audio akan muncul judul dari file audio tersebut. Selanjutnya, menekan tombol klasifikasi untuk mengklasifikasikan musik tersebut. Lalu, hasilnya adalah prediksi genre dari musik tersebut dan distribusi probabilitas dari sepuluh genre yang berbeda.

## III. HASIL DAN PEMBAHASAN

### A. Hasil Pembuatan Model

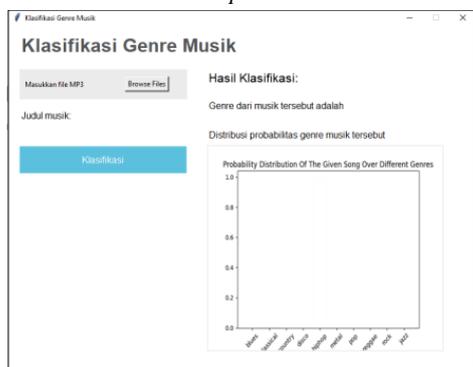
Tabel 4 Hasil perbandingan performa model

No	Nama Model	Precisio	Rec	F1- score	Accu racy	Loss
1	CNN1	0,79	0,78	0,78	78,2%	0,646
2	CNN_2	0,82	0,82	0,82	81,7%	0,576
3	CNN_3	0,8	0,79	0,79	79,4%	0,677

Berdasarkan Tabel 4 model yang menghasilkan nilai *accuracy* terbesar dan nilai *loss* terkecil adalah model CNN\_2 dengan nilai *accuracy* sebesar 81,7%, nilai *loss test* sebesar 0,576, nilai *precision* rata-rata sebesar 0,82, *recall* rata-rata sebesar 0,82 dan *F1-score* sebesar 0,82. Maka, pada aplikasi GUI akan digunakan model CNN\_2 untuk melakukan prediksi atau klasifikasi genre musik.

## B. Hasil Pembuatan GUI

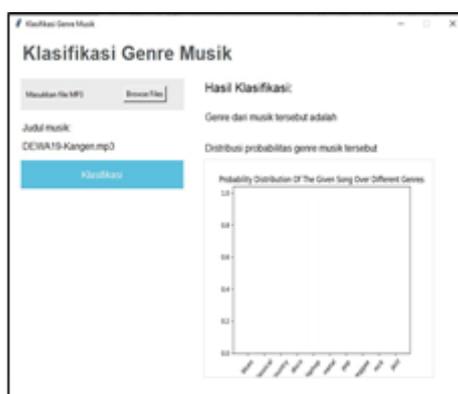
Gambar 7 Tampilan awal GUI



Gambar 7 adalah hasil tampilan awal GUI ketika dijalankan. Tampilan GUI sudah terdapat label judul, tombol pencarian *file* musik, label judul musik, tombol klasifikasi dan label hasil prediksi. Dari hasil tersebut terlihat tampilan sudah sesuai dengan rancangan *mockup*.

Gambar 8 Hasil ketika melakukan browse file

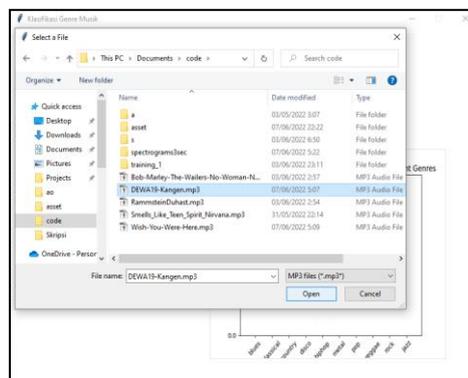
Gambar 8 adalah hasil ketika menekan tombol browse file pada GUI. Tampilan *menu dialog* pencarian *file* sudah diatur *default* dalam format MP3 sehingga, akan langsung menampilkan hanya file dalam format MP3.



Gambar 9 Hasil ketika pengguna sudah memasukkan file musik

Gambar 9 adalah hasil ketika pengguna sudah memasukkan *file* musik. Terlihat GUI akan

menampilkan judul musik dari *file* musik yang pengguna masukkan. Hal ini dilakukan agar pengguna dapat mengetahui *file* yang akan diklasifikasikan. Selanjutnya, pengguna dapat langsung menekan tombol klasifikasi untuk dapat melakukan prediksi agar mengetahui genre dari musik yang sebelumnya sudah dimasukkan.



Gambar 10 Hasil tampilan saat mengklasifikasikan genre musik

Gambar 10 adalah hasil tampilan saat mengklasifikasikan genre musik. Terlihat GUI dapat menampilkan label prediksi genre musik dalam bentuk teks. Kemudian, GUI juga dapat menampilkan hasil prediksi distribusi probabilitas dari sepuluh genre musik dalam bentuk citra. Dari hasil ini maka, tombol klasifikasi dapat berfungsi dengan semestinya kemudian, hasil prediksi dapat menampilkan sesuai dengan harapan.

#### IV. SIMPULAN

Berikut adalah kesimpulan dari hasil penelitian klasifikasi genre musik yang telah dilakukan.

1. Berdasarkan tujuan dari penelitian ini yaitu, menerapkan klasifikasi genre musik menggunakan CNN yang digunakan untuk mengelompokkan musik berdasarkan genre agar terorganisir sudah dapat dilakukan.
2. Hasil pengujian model menggunakan *white box* sudah berjalan sesuai harapan dimana fungsi-fungsi sudah terimplementasi dengan baik dan mendapat model terbaik.
3. Berdasarkan hasil penelitian, model terbaik yang didapat setelah melakukan pelatihan pada tiga arsitektur model adalah model CNN\_2 yang mendapat tingkat akurasi *test* sebesar 81.7% setelah menjalankan pelatihan sebanyak 120 *epoch* dengan konfigurasi menggunakan *adam optimizer*, *loss categorical-crossentropy*, *learning rate* sebesar 0.00005 dan menggunakan data latih sebanyak 9.000 citra serta data uji sebanyak 1.000 citra.
4. Hasil pengujian GUI menggunakan *black box* sudah berjalan sesuai harapan dimana fitur-fitur yang sebelumnya dirancang sudah terimplementasi dengan baik dan berfungsi dengan semestinya.
5. Berdasarkan hasil percobaan klasifikasi genre terhadap beberapa musik, terdapat hasil prediksi genre yang sering tertukar yaitu genre musik *hip hop*, *metal* dan *rock*, karena genre musik tersebut memiliki pola frekuensi *spectrum* yang mirip.

Berdasarkan hasil penelitian yang telah dilakukan, berikut adalah saran untuk penelitian sejenis selanjutnya.

1. Menggunakan arsitektur model

yang berbeda, karena tingkat akurasi training lebih *tinggi* dibandingkan tingkat akurasi *test* dimana model pada penelitian ini masih terjadi *overfitting*.

2. Menambah *class* atau genre yang akan diprediksi, karena masih banyak genre musik yang ada.
3. Menggunakan fitur-fitur yang terdapat pada musik seperti harmoni, *beat*, *tempo* dan sebagainya, agar mendapat tingkat akurasi *test* yang lebih baik.

#### DAFTAR RUJUKAN

- [1] Adiyansjah, A. A. S. Gunawan, and D. Suhartono, "Music Recommender System Based on Genre Using Convolutional Recurrent Neural Networks," in *Procedia Computer Science*, 2019, vol. 157, pp. 99–109. doi: 10.1016/j.procs.2019.08.146.
- [2] Mansoor Iqbal, "Spotify Revenue and Usage Statistics (2022)," *businessofapps.com*, Jan. 19, 2022. Accessed: Feb. 23, 2022. [Online]. Available: <https://www.businessofapps.com/data/spotify-statistics/>
- [3] C. R. Wairata, E. R. Swedia, and M. Cahyanti, "Pengklasifikasian Genre Musik Indonesia Menggunakan Convolutional Neural Network," *Sebatik*, vol. 25, no. 1, Jun. 2021, doi: 10.46984/sebatik.v25i1.1286.
- [4] M. A. A. Sarofi and A. Mukarromah, "Identifikasi Genre Musik dengan Menggunakan Metode Random Forest," *Sains dan Seni ITS*, vol. 09, no. 01, 2020.
- [5] A. V. M. Giri, "Klasifikasi Musik Berdasarkan Genre dengan Metode K-Nearest Neighbor," *Jurnal Ilmu Komputer*, vol. 11, no. 2, 2018, [Online]. Available: <https://developer.spotify.com/web-api/>
- [6] D. A. Huang, A. A. Serafini, and E. J. Pugh, "Music Genre Classification," 2018. [Online]. Available: <https://github.com/derekahuang/Music-Classification>
- [7] 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, Jul.2002, doi: 10.1109/TSA.2002.800560.