JATIKOM: Jurnal Aplikasi dan Teori Ilmu Komputer | ISSN: 2615-577X (Online)
https://ejournal.upi.edu/index.php/JATIKOM

Klasifikasi Genre Musik Dengan *Mel Frequency Cepstral Coefficient* Dan Spektogram Menggunakan *Convolutional Neural Network*

Sifa Marcella Fardhani[1], Yaya Wihardi[2], Erna Piantari[3]

Prodi Ilmu Komputer Departemen Ilmu Komputer Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam
Universitas Pendidikan Indonesia
Bandung, Indonesia
[1]marcellassifa@student.upi.edu, [2,]yaya@upi.edu, [3,]erna@upi.edu,

Abstrak—**Musik sudah menjadi suatu kebutuhan bagi sebagian besar orang karena manfaatnya yang dapat menimbulkan relaksasi dan dapat menjadi hiburan bagi sebagian orang. Kebutuhan akan informasi yang terkandung dalam musik yang didengarkan seringkali dibutuhkan, salah satunya adalah informasi dari jenis genre musik yang sedang didengarkan. Untuk mengetahui jenis genre musik tersebut maka pada penelitian ini dilakukan klasifikasi genre musik yang diharapkan akan memenuhi kebutuhan tersebut dengan melalui proses pengenalan pola dari masing-masing genre. Metode *Convolutional Neural Network* (CNN) digunakan untuk melatih pola dari jumlah 500 data lagu GTZAN berbahasa Inggris dan 500 data lagu berbahasa Indonesia yang mencakup 5 genre. Teknik ekstraksi fitur digunakan pada praproses data untuk mendapatkan hasil ekstraksi *Mel-Frequency Cepstral Coefficient* (MFCC) dan spektogram. Kemudian arsitektur CNN akan dibandingkan dengan dua jenis data masukan yang berbeda, yaitu data berbentuk vektor sebagai representasi dari hasil ekstraksi MFCC dan data berbentuk citra spektogram. Setelah itu, data akan melalui proses validasi untuk mengetahui nilai evaluasi dari kinerja model yang dihasilkan masing-masing arsitektur dengan data masukan berbeda. Hasil validasi terbaik ditunjukkan oleh eksperimen dengan data masukan spektogram menggunakan dataset GTZAN yang memiliki nilai akurasi sebesar 76%.**

Kata kunci: *Music Genre Classification*, *Convolutional Neural Network*, MFCC, Spektogram, *Deep Learning*.

I. PENDAHULUAN

Dewasa ini, industri media sebagai salah satu penghasil musik di seluruh dunia mengalami perkembangan yang sangat pesat. Salah satu bentuk musik yang sering digunakan pada teknologi masa kini adalah jenis *music on demand*, dimana para penggunanya dapat bebas memilih musik juga memutarnya dimana saja dan kapan saja serta dapat disimpan secara *offline*. Di Indonesia ada beberapa aplikasi *music on demand* yang populer, diantaranya Spotify, Joox, dan iTunes.

Tingginya peminat musik di Indonesia dikarenakan musik bisa digunakan sebagai terapi relaksasi untuk memperbaiki, memelihara, dan mengembangkan mental, fisik, juga menstabilkan kesehatan emosi [1]. Dengan demikian, muncul beberapa masalah bagi para pengguna aplikasi. Adanya fitur pencari lagu pada aplikasi *music on demand* berdasarkan genre musik membuat sebagian orang tidak tahu apa genre dari musik yang akan mereka cari. Seringkali ketidaktahuan pendengar dari detail genre musik yang sedang didengar merupakan sebuah masalah saat mereka mendapatkan rekomendasi musik berdasarkan genrenya.

Solusi yang dapat dilakukan untuk masalah tersebut adalah klasifikasi musik berdasarkan genre. Permasalahan ini berkaitan dengan detail sebuah musik, dan mendorong munculnya sebuah ilmu interdisipliner yang bernama *Music Information Retrieval* (MIR). MIR adalah sebuah sistem pemanggilan kembali informasi dari suatu file musik agar dapat memberikan informasi musik yang lebih kompleks [2] dan mulai populer untuk menjawab permasalahan-permasalahan terkait musik dengan melakukan kombinasi dengan metode-metode pendukung pengambilan informasi. Dalam perkembangannya, MIR telah digunakan untuk melakukan pemanggilan informasi sebuah musik dengan detailnya seperti jenis mood, estimasi tempo, jenis instrumen, judul, dan sampai saat ini MIR terus berkembang untuk klasifikasi genre musik.

Untuk melakukan klasifikasi genre, data harus dilakukan ekstraksi fitur terlebih dahulu. Ekstraksi fitur bisa dilakukan dengan dua teknik ekstraksi yang paling sering digunakan pada klasifikasi genre, yaitu dengan teknik *Mel Frequency Cepstral Coefficient* (MFCC) dan pembentukkan spektogram. Kedua teknik tersebut dinilai efisien untuk menghasilkan ekstraksi fitur pada data klasifikasi genre, dan disimpan dengan representasi data yang berbeda. MFCC dapat direpresentasikan dalam

bentuk vektor dan spektogram direpresentasikan dalam bentuk file jpg. Dalam MIR ekstraksi fitur akan sangat mempengaruhi hasil karena semakin banyak dan semakin unik fitur dari setiap data maka akan semakin baik informasi yang didapat dari data audio. Pada penelitian ini, kedua teknik ekstraksi fitur tersebut akan digunakan untuk dilakukan perbandingan data hasil ekstraksinya agar mendapat arsitektur jaringan yang optimal pada klasifikasi genre.

Metode Jaringan Syaraf Tiruan sebelumnya digunakan sebagai metode klasifikasi, namun metode tersebut masih memiliki kekurangan seperti proses *learning* masih memiliki komputasi yang belum terlalu cepat, dan belum optimal jika digunakan untuk klasifikasi data berbentuk citra. Hal ini membuat munculnya perkembangan dari *Neural Network* untuk menggunakan model *Deep Learning*. *Deep Learning* merupakan sebuah metode pembelajaran terhadap mesin yang berbasis *learning* dan pengolahan data dengan terus belajar terhadap dataset yang baru ketika menemui hal-hal yang mirip dengan data sebelumnya. Pengembangan *Deep Learning* yang cukup populer digunakan yaitu *Convolutional Neural Network* (CNN), dimana metode ini dinilai cukup cepat dalam melakukan proses *training* dalam hal mengenali sebuah pola pada dataset terutama dataset berbentuk citra karena jaringannya yang lebih mendalam dan lebih kompleks. Oleh karena itu, tujuan dilakukannya penelitian ini untuk melakukan klasifikasi genre musik dengan teknik klasifikasi CNN dan menggunakan dua jenis ektraksi fitur yaitu MFCC dan spektogram yang kemudian akan dilakukan perbandingan kombinasi antara CNN dan ekstraksi fitur MFCC dengan CNN dan ekstraksi fitur spektogram untuk mengetahui ektraksi fitur mana yang memberikan hasil lebih baik saat dikombinasikan dengan teknik klasifikasi CNN.

Genre dipilih untuk proses klasifikasi karena genre merupakan salah satu pelabelan paling mudah yang digunakan untuk mengkategorikan jenis musik berdasarkan instrumentasi, ritmik yang dimiliki, dan konten *pitch* musik [3]. Ekstraksi fitur MFCC akan mengambil frekuensi spektrum dari gelombang data audio yang kemudian ditransformasi dan diambil frekuensi dari spektrum yang dihasilkan, lalu metode *Short-Time Fourier Transform* (STFT) akan membentuk fitur spektogram dengan membagi sinyal waktu yang lebih panjang menjadi segmen yang lebih pendek dengan panjang yang sama dan kemudian menghitung transformasi *Fourier* secara terpisah pada setiap segmen yang lebih pendek. Perbandingan data hasil dari ekstraksi fitur dilakukan untuk mencari kinerja model yang paling baik dalam melakukan klasifikasi genre selain menggunakan Spektogram dimana fitur Spektogram sudah banyak digunakan dan dinilai baik untuk melakukan pengklasifikasian data berupa audio. Selanjutnya, data hasil ekstraksi fitur akan masuk ke dalam proses *training* menggunakan CNN. Metode CNN dipilih karena dinilai memiliki kinerja optimal dan cepat dalam melakukan klasifikasi data, khususnya data berupa citra dan sudah terbukti pada penelitian terkait dengan menghasilkan akurasi yang lebih baik dibandingkan penelitian sebelumnya yang belum memakai CNN.

II. PENELITIAN TERKAIT

Saat ini, sudah cukup banyak penelitian yang membahas tentang klasifikasi genre musik menggunakan metode pengembangan *deep learning* CNN. Setelah rilis berbagai kemudahan untuk para peneliti, mulai bermunculan perkembangan teknologi-teknologi yang berhubungan dengan MIR serta metodenya. Beberapa penelitian terkait tersebut diantaranya penelitian yang melakukan klasifikasi genre musik dengan *Local Feature Maps* [4]. Metode yang digunakan adalah CNN dengan teknik ekstraksi fitur *Short Time Fourier Transform* (STFT) ditambah dengan *Grey Level Co-occurrence Matrix* (GLCM) yang diaplikasikan pada spektogram hasil ekstraksi fitur STFT. GLCM membuat spektogram sebagai data masukan berubah warna menjadi *greyscale*. Data yang digunakan adalah GTZAN dataset berjumlah 1000 file lagu yang mencakup 10 jenis genre. Selanjutnya penelitian yang dilakukan tahun 2014 yang melakukan klasifikasi genre terhadap musik dengan melakukan teknik pembelajaran terhadap fitur musik menggunakan *Deep Neural Network* [5]. Proses *learning* dilakukan menggunakan *Stochastic Gradient Descent* (SGD) dan dengan fungsi aktivasi ReLU. Data yang digunakan adalah gabungan dari GTZAN dataset berjumlah 1000 file lagu dan ISMIR 2004 dataset yang berjumlah 1458 file lagu yang mencakup 6 genre. Ekstraksi fitur dilakukan dengan Teknik *Fast Fourier Transform* (FFT).

Selain teknologi *deep learning* yang disebutkan sebelumnya, ada juga beberapa penelitian yang menjadi referensi dari penelitian paper ini namun tidak menggunakan metode CNN. Seperti penelitian tahun 2012, yang menggunakan perbandingan metode Jaringan Syaraf Tiruan *Learning Vector Quantization* (LVQ) dan *Euclidean Distance* untuk melakukan deteksi judul lagu pada musik instrumental piano [6]. Teknik ekstraksi fitur pada dataset yang digunakan adalah *Mel-Frequency Cepstral Coefficient* (MFCC) dimana data akan dilakukan pembentukan sinyal mel-spektrum dan direpresentasikan dalam bentuk vektor. Kemudian pada tahun yang sama juga terdapat penelitian yang menggunakan Jaringan Syaraf Tiruan LVQ untuk melakukan klasifikasi genre musik berdasarkan file audio yang berbentuk *waveform* (.wav) [7]. Teknik ekstraksi fitur yang digunakan pada dataset saat praproses adalah *Sort Time Energy* (STE) dan *Zero Crossing Rate* (ZCR).

Kemudian pada tahun 2009 ada yang menggunakan metode Jaringan Syaraf Tiruan dengan algoritma *Self-Organizing Maps* (SOMS) untuk melakukan klasifikasi genre music [8]. Penelitian ini memanfaatkan algoritma SOMS untuk mengklasifikasikan genre musik dengan file suara berbentuk *waveform* (.wav). Dataset yang digunakan melalui praproses untuk pengambilan ekstraksi fitur

berdasarkan konten frekuensi dan tekstur timbral terhadap spektrum yang terbentuk dari file suara.

III. METODE

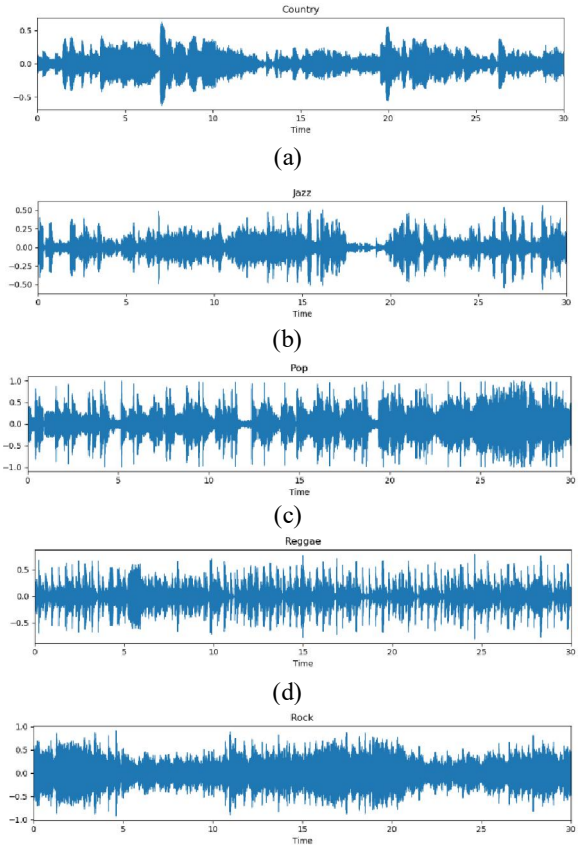
A. Pengumpulan Data

Pada penelitian ini data yang dibutuhkan adalah file musik dengan format *waveform*. Data yang digunakan didapat dari GTZAN dataset dengan mengunduhnya melalui situs http://marsyas.info/downloads/datasets.html. GTZAN dataset dibuat dan digunakan oleh G. Tzanetakis dan P. Cook tahun 2002 pada saat melakukan penelitian klasifikasi genre musik untuk dipublikasikan kepada IEEE [3]. GTZAN dataset terdiri dari 10 jenis genre yaitu Blues, Classical, Country, Disco, Hip Hop, Jazz, Metal, Pop, Reggae dan Rock dengan masing-masing genre memiliki 100 file lagu berdurasi 30 detik. Saat diunduh, dataset GTZAN dikemas dalam bentuk file .tar berukuran 1,5 GB yang didalamnya berisi 10 folder genre dengan total jumlah data terdiri dari 1000 potongan musik dalam format .au dan dengan *channel* Mono 16-bit.

TABEL I.
ISI DARI GTZAN DATASET

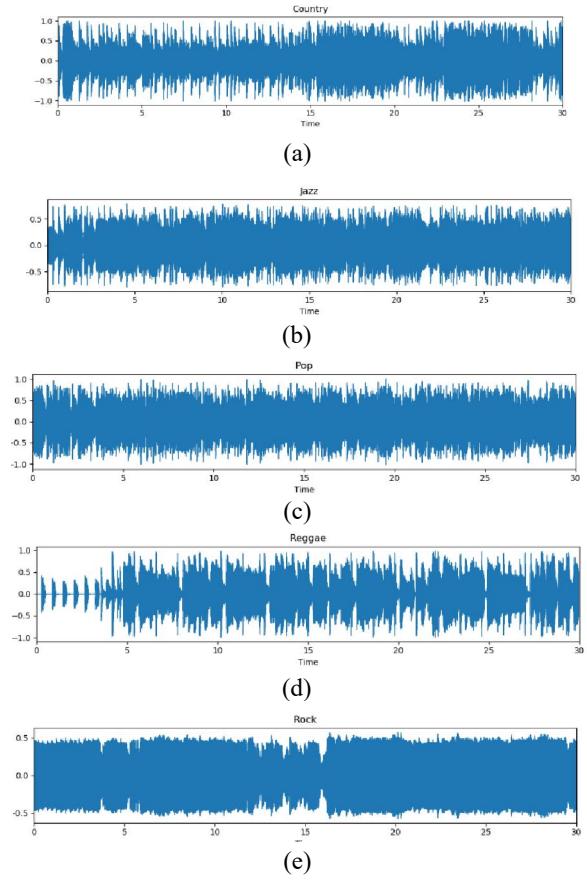
Folder ke-	Genre	Jumlah File	Format File
1	Blues	100	.au
2	Classical	100	.au
3	Country	100	.au
4	Disco	100	.au
5	Hip Hop	100	.au
6	Jazz	100	.au
7	Metal	100	.au
8	Pop	100	.au
9	Reggae	100	.au
10	Rock	100	.au

Namun, karena genre yang dipakai pada penelitian ini hanya 5, maka total data dari dataset GTZAN yang dipakai hanya 500 file musik. Lima genre tersebut adalah Country, Jazz, Pop, Reggae dan Rock. Bentuk file pada dataset GTZAN masih berupa .au, sehingga perlu dilakukan konversi untuk mengubah formatnya menjadi .wav dengan menggunakan konverter *online*. Konverter yang bisa digunakan untuk mengubah format .au menjadi .wav tersedia di www.doscpal.com. Setelah dilakukan konversi, sinyal dari setiap data lagu dapat dilihat dengan cara melakukan *waveplot* sebagai representasi setiap data lagu yang berbentuk sinyal mentah dengan memetakan sinyal berdasarkan domain waktu dan amplitudo, dan dibuat dengan menggunakan kode program Python dengan bantuan *library* Librosa seperti Gambar 1.



Gambar 1. Bentuk *Waveplot* Dari Sampel Data Lagu Pada Dataset GTZAN Dalam 5 Genre, (a) Country, (b) Jazz, (c) Pop, (d) Reggae dan (e) Rock

Kemudian data kedua yang digunakan adalah dataset lagu Indonesia yang dikumpulkan dengan cara *handmade*, yakni membuat sendiri dataset dari lagu-lagu yang telah tersimpan dalam *harddisk* atau dari lagu yang baru saja diunduh untuk kemudian dilakukan proses konversi dan *chunking* secara manual. Genre yang dipakai pun berjumlah sama dengan dataset GTZAN, yakni 5 genre yang terdiri dari Country, Jazz, Pop, Reggae, Rock. Karena lagu yang dimiliki belum dikelompokkan berdasarkan genre, maka pengelompokkan dan *labelling* dilakukan manual dengan melihat kemiripan sinyal *waveplot* audio dari setiap genre seperti Gambar 2, lalu melakukan pencocokan dengan melihat sumber-sumber melalui situs pencarian www.google.com, dan dicocokkan dengan beberapa situs lagu-lagu populer Indonesia yaitu https://dansmedia.net/musik/chart-lagu-indonesia-terbaru/, www.musikpopuler.com, www.nandahero.com, dan id.wikipedia.org sebagai acuan untuk mengelompokkan musik Indonesia berdasarkan genrenya. Sebelumnya data akan melalui proses konversi untuk mengubah ekstensi file MP3 menjadi .wav.



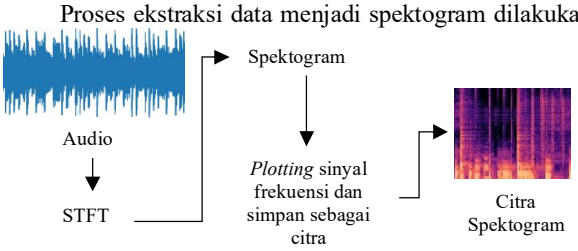
Gambar 2. Bentuk *Waveplot* Dari Sampel Data Lagu Pada Dataset Indonesia Dalam 5 Genre, (a) Country, (b) Jazz, (c) Pop, (d) Reggae dan (e) Rock

B. Ekstraksi Fitur Spektogram

Pada proses eksperimen, kedua dataset akan menjadi data input untuk dilakukan proses pengklasifikasian dengan CNN agar menghasilkan output berupa genre hasil klasifikasi. Sebelumnya, data melalui praproses terlebih dahulu untuk dilakukan ekstraksi fitur pada kedua dataset. Teknik pertama yang dilakukan untuk mendapat ekstraksi fitur adalah *Short Time Fourier Transform* (STFT). STFT merupakan sebuah metode untuk melakukan ekstraksi fitur dari data berupa sinyal audio yang diuraikan menjadi gelombang sinusoidal, dengan menggunakan analisis Fourier [10]. Transformasi Fourier sendiri pada dasarnya mengubah sinyal deret berupa gelombang sinusoidal dengan domain waktu dan amplitudo ke dalam domain frekuensinya. Pada saat yang bersamaan STFT akan memberikan informasi waktu dan frekuensi secara temporal. Dalam prakteknya, prosedur untuk menghitung STFT adalah dengan membagi sinyal waktu yang panjang menjadi beberapa segmen yang lebih pendek atau *chunk*, namun dengan panjang yang masih sama di setiap *chunk* nya.

Gambar 3. Proses Ekstraksi Fitur Spektogram

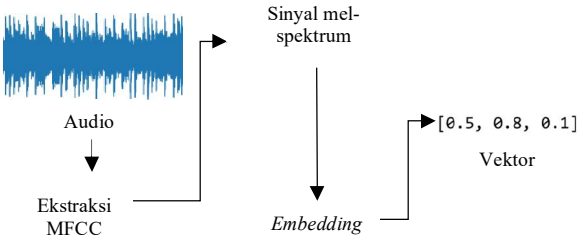
Proses ekstraksi data menjadi spektogram dilakukan



seperti Gambar 3 dengan bantuan kode program Python serta *library* Librosa dan Matplotlib. Hasil *plotting* sinyal disimpan sebagai citra spektogram dalam bentuk file.jpg.

C. Ekstraksi Fitur Mel Frequency Cepstral Coefficient

Teknik selanjutnya yang dilakukan untuk mendapat ekstraksi fitur adalah *Mel Frequency Cepstral Coefficient* (MFCC). MFCC adalah sebuah fitur untuk melakukan ekstraksi data audio yang berdurasi pendek dan ringkas yang proses ekstraksinya dilakukan berdasarkan spektogram dari hasil ekstraksi fitur STFT [10]. MFCC merupakan representasi dari spektrum daya jangka pendek dari suatu suara, berdasarkan pada transformasi kosinus linier spektrum daya log pada skala frekuensi non linear dimana representasi ini lebih sederhana dan kompresibel dari hasil ekstraksi fitur STFT. Proses ekstraksi fitur MFCC dilakukan dengan bantuan kode program Python serta *library* Librosa dan Pydub.



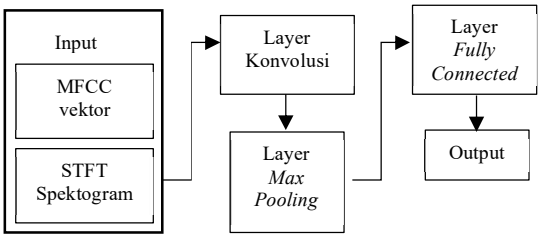
Gambar 4. Proses Ekstraksi Fitur MFCC

Proses ekstraksi MFCC dilakukan untuk mendapatkan sinyal spektrum mel-frekuensi yang akan direpresentasikan ke dalam bentuk vektor dari file lagu dengan menggunakan kode program Python seperti pada Gambar 3, dengan melakukan *embedding*. *Embedding* adalah sebuah proses pemetaan dari objek diskrit yang dalam penelitian ini adalah data audio, menjadi vektor dengan bilangan *real*. *Embedding* dilakukan supaya data dapat direpresentasikan dengan vektor melalui proses pemetaan dari data audio yang bersifat diskrit ke dalam bentuk vektor bilangan real.

D. Convolutional Neural Network

Metode klasifikasi yang digunakan dalam penelitian ini adalah *Convolutional Neural Network* (CNN) dengan beberapa arsitektur hasil dari *tune up* parameter. Kedua jenis data baik vektor atau spektogram akan dilatih dengan beberapa teknik konvolusi, dengan melewati layer konvolusi, kemudian layer *pooling* untuk selanjutnya masuk ke layer *fully connected* agar dapat menghasilkan

output sesuai lima kelas genre yang sudah dilatih sebelumnya. Garis besar metode CNN yang digunakan dijelaskan pada Gambar 5.



Gambar 5. Alur Kerja Metode CNN

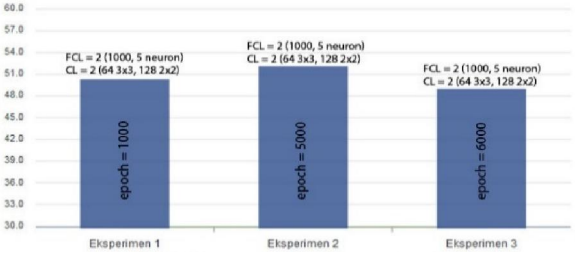
Eksperimen kali ini menggunakan tiga layer utama, yakni layer konvolusi, layer *pooling*, dan layer *fully connected*. Lalu dilakukan *tuning parameter* untuk mencari sendiri tahap demi tahap arsitektur yang menghasilkan nilai paling optimal, dengan menambah jumlah layer konvolusi, mengubah jumlah filternya, menambah jumlah layer *fully connected* dan mengubah jumlah neuronnya di akhir jaringan. Keseluruhan arsitektur dijalankan dengan menentukan jumlah *epoch*.

1) Layer *input*
   Layer *input* pada penelitian ini terdiri dari dua jenis data, yaitu data vektor dan data spektogram. Data vektor berupa audio yang sebelumnya melalui proses *embedding* untuk memetakannya ke dalam vektor menggunakan metode MFCC. Vektor ini secara tidak langsung telah terkonversi layaknya data citra berukuran 3 dimensi berukuran 20x11 dengan proses *embedding* tersebut. Sedangkan data spektogram yang digunakan berupa citra berukuran 128x128 dengan chanel 3 yang berarti RGB.
2) Layer Konvolusi (*Convolutional Layer*)
   Layer konvolusi melakukan operasi yang bertujuan untuk mengekstraksi fitur untuk mempelajari representasi fitur dari input citra. Dalam operasinya, konvolusi melakukan pengaplikasian *kernel* dan *filter* yang pada penelitian ini didapat dari *tuning parameter*, dengan menggunakan fungsi aktivasi ReLU.
3) Layer *Pooling*
   Layer *pooling* adalah jenis metode pada layer *subsampling* yang digunakan pada riset ini. Tujuan dilakukannya *pooling* adalah untuk melakukan *downsampling* pada representasi data citra, dengan melakukan pengurangan dimensi untuk melakukan penyesuaian menuju layer berikutnya dengan menyediakan bentuk representasi dengan nilai maksimum dari *input*.
4) Layer *Fully Connected* dan Layer *output*
   Layer ini menghubungkan setiap neuron pada layer sebelumnya menuju layer-layer berikutnya. Pada penelitian ini, layer *fully connected* mendapatkan input dari keluaran pada layer konvolusi dan layer *pooling* berupa hasil ekstraksi dari input citra. Agar mendapatkan *output*, layer ini menggunakan fungsi aktivasi *softmax* untuk menghasilkan klasifikasi berdasarkan kelas yang telah diinisialisasi, dan pada layer terakhir jumlah neuron disamakan dengan jumlah kelas.

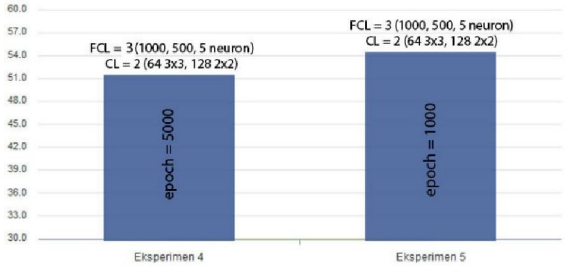
IV. HASIL PENELITIAN DAN PEMBAHASAN

A. *Tuning Parameter*

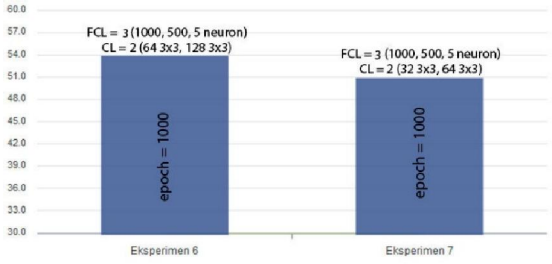
Proses *tuning* parameter ini dilakukan untuk menemukan arsitektur CNN yang menghasilkan akurasi paling optimal. Pencarian arsitektur dilakukan dari mulai membangun jaringan dengan kompleksitas tinggi hingga kompleksitas rendah juga sebaliknya dan mencoba setiap arsitekturnya. Eksperimen MFCC melakukan sebanyak 11 kali *tuning*, dan eksperimen Spektogram melakukan 4 kali *tuning*. Gambar 6 sampai 10 menunjukkan hasil *tuning parameter* pada eksperimen MFCC.



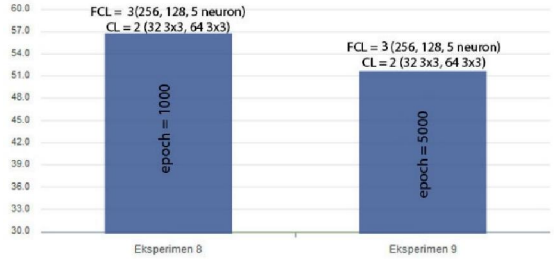
Gambar 6. Grafik Pengaruh Perubahan Epoch Pada Eksperimen MFCC ke 1, 2 dan 3



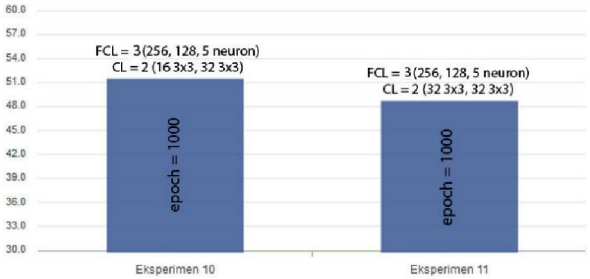
Gambar 7. Grafik Pengaruh Perubahan Epoch dan Fully Connected Layer Pada Eksperimen MFCC ke 4 dan 5



Gambar 8. Grafik Pengaruh Perubahan Kernel dan Jumlah Filter Convolution Layer Pada Eksperimen MFCC ke 6 dan 7

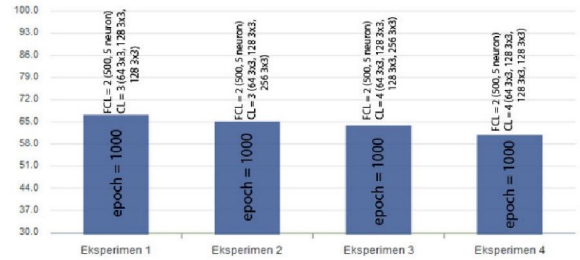


Gambar 9. Grafik Pengaruh Perubahan Neuron Fully Connected Layer Pada Eksperimen MFCC ke 8 dan 9



Gambar 10. Grafik Pengaruh Perubahan Filter Convolution Layer Pada Eksperimen MFCC ke 10 dan 11

Sedangkan Gambar 11 menunjukkan hasil *tuning parameter* pada eksperimen Spektogram.



Gambar 11. Grafik Pengaruh Perubahan Convolution Layer Pada Eksperimen Spektogram

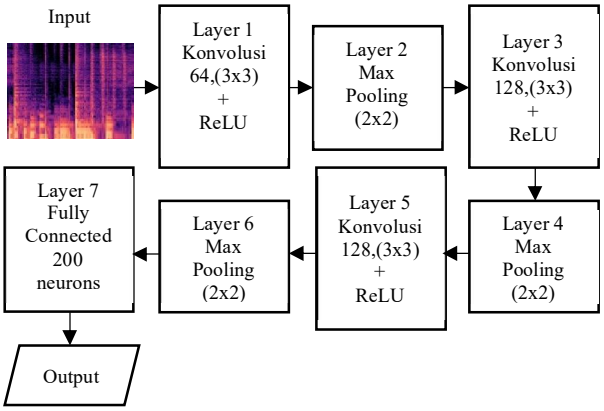
Dari grafik-grafik tersebut dapat disimpulkan, semakin besar nilai parameter diubah tidak menjamin hasil akurasi akan lebih baik. Hal ini disebabkan kemungkinan bisa terjadi *overfitting* karena jaringan yang dibangun terlalu kompleks untuk data masukan. Kesesuaian parameter dapat dilakukan dengan melakukan pencarian parameter yang tepat agar menghasilkan arsitektur yang optimal.

B. Eksperimen Spektogram

Eksperimen ini menggunakan data input berupa citra spektogram berukuran 128x128 hasil dari ekstraksi data audio menggunakan metode STFT dan direpresentasikan dengan spektogram. Data yang digunakan adalah lagu dengan Bahasa Inggris yang diperoleh dari GTZAN dataset berjumlah 500 data dengan 5 kelas. Arsitektur CNN yang digunakan didapat dari proses *tuning parameter*. Hasil *tuning parameter* terbaik didapat seperti pada Gambar 12.

Gambar 12. Arsitektur CNN Untuk Eksperimen Spektogram

Selanjutnya dilakukan validasi model yang dihasilkan dari arsitektur CNN pada Gambar 12. Validasi ini dilakukan untuk mengukur kinerja model dengan menghitung nilai dari akurasi, presisi, *recall* dan f_1-*measure*. Proses validasi dilakukan dengan menggunakan metode *cross-validation* dimana data terlebih dahulu dibagi menjadi beberapa bagian dan proses validasi dilakukan sesuai *fold* nya. Pada proses validasi ini menggunakan 5



fold yang artinya pengujian dilakukan sebanyak 5 kali dengan 5 bagian data yang berbeda. Pembagian data berjumlah 400 untuk data *training* dan 100 untuk data *testing* yang mencakup 5 kelas. Dari 100 data testing, hasil validasi yang didapat dari data tersebut maka dihitung rata-rata nilainya sebagai hasil evaluasi pada model tersebut.

TABEL II.
HASIL VALIDASI TERHADAP ARSITEKTUR CNN EKSPERIMEN SPEKTOGRAM GTZAN

***Fold* ke-**	**Akurasi**	**Presisi**	***Recall***	***F_1-Measure***
1	71.00%	73.96%	71.00%	70.95%
2	67.00%	67.83%	67.00%	66.74%
3	76.00%	76.81%	76.00%	75.96%
4	72.00%	72.79%	72.00%	71.88%
5	68.00%	70.91%	68.00%	68.69%
Rata-Rata	70.80%	72.46%	70.80%	70.84%

Hasil validasi terhadap arsitektur CNN pada eksperimen spektogram menunjukkan rata-rata nilai dari kelima *fold*. Rata-rata yang didapat pada validasi ini yaitu akurasi 70.80%, presisi 72.46%, *recall* 70.80% dan f_1-*measure* 70.84%. Kemudian, hasil validasi tertinggi didapat oleh *fold* ke 3 dengan nilai akurasi 76.00%, presisi 76.81%, *recall* 76.00% dan f_1-*measure* 75.96% dengan rincian data pada Tabel 3 dan 4.

TABEL III.
HASIL PENGUJIAN DATA PADA EKSPERIMEN SPEKTOGRAM GTZAN

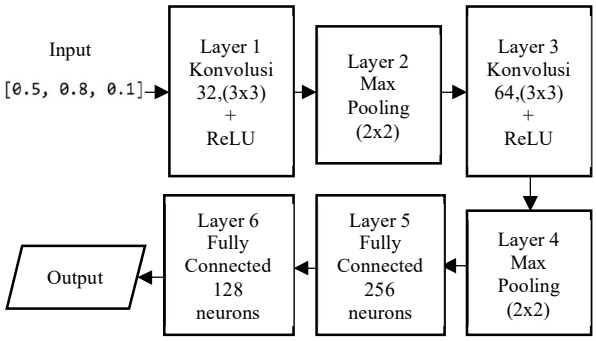
***Fold* ke-**	**Jumlah Data Testing**	**Benar**	**Salah**
1	100	71	29
2	100	67	33
3	100	76	24
4	100	72	27
5	100	68	32

TABEL IV.
HASIL PENGUJIAN TIAP KELAS DATA PADA EKSPERIMEN SPEKTOGRAM GTZAN

Kelas	Akurasi	Presisi	Recall	*F1-Measure*
Country	75.00%	75.00%	75.00%	75.00%
Jazz	90.00%	85.71%	90.00%	87.80%
Pop	65.00%	86.67%	65.00%	74.28%
Reggae	80.00%	66.67%	80.00%	72.73%
Rock	70.00%	70.00%	70.00%	70.00%

C. *Eksperimen MFCC*

Eksperimen ini menggunakan data input berupa vektor berukuran 20x11 hasil dari ekstraksi menggunakan metode MFCC. Data yang digunakan adalah lagu dengan Bahasa Inggris yang diperoleh dari GTZAN dataset berjumlah 500 data yang mencakup 5 kelas. Arsitektur CNN yang digunakan didapat dari proses *tuning parameter*. Hasil *tuning parameter* terbaik didapat seperti pada Gambar 13.



Gambar 13. Arsitektur CNN Untuk Eksperimen MFCC

Selanjutnya dilakukan validasi model yang dihasilkan dari arsitektur CNN pada Gambar 13. Validasi ini dilakukan untuk mengukur kinerja model dengan menghitung nilai dari akurasi, presisi, *recall* dan f_1-*measure*. Proses validasi dilakukan dengan menggunakan metode *cross-validation* dimana data terlebih dahulu dibagi menjadi beberapa bagian dan proses validasi dilakukan sesuai *fold* nya. Pada proses validasi ini menggunakan 5 *fold* yang artinya pengujian dilakukan sebanyak 5 kali dengan 5 bagian data yang berbeda. Pembagian data berjumlah 300 untuk data *training* dan 200 untuk data *testing* yang mencakup 5 kelas. Dari 200 data testing, hasil validasi yang didapat dari data tersebut maka dihitung rata-rata nilainya sebagai hasil evaluasi pada model tersebut.

TABEL V.
HASIL VALIDASI TERHADAP ARSITEKTUR CNN EKSPERIMEN MFCC GTZAN

Fold ke-	Akurasi	Presisi	*Recall*	*F1-Measure*
1	53.00%	52.69%	53.00%	52.04%
2	53.00%	52.84%	53.00%	52.57%
3	58.00%	57.97%	58.00%	57.64%
4	53.00%	52.93%	53.00%	52.67%
5	54.00%	53.98%	54.00%	53.48%
Rata-Rata	54.20%	54.08%	54.20%	53.68%

Hasil validasi terhadap arsitektur CNN pada eksperimen spektogram menunjukkan rata-rata nilai dari kelima *fold*. Rata-rata validasi dari eksperimen MFCC lebih rendah daripada eksperimen spektogram, dengan selisih rata-rata akurasi 16%. Hal ini dikarenakan data hasil ekstraksi fitur MFCC merupakan data kompresibel sehingga representasinya lebih sederhana dibandingkan dengan spektogram.

Rata-rata validasi pada eksperimen 10 yakni akurasi sebesar 54.20%, presisi sebesar 54.08%, *recall* sebesar 54.20% dan f_1-*measure* sebesar 53.68%. Dari kelima *fold*, hasil evaluasi yang terbaik didapat oleh *fold* ke 3 dengan hasil akurasi 58.00%, presisi 57.97%, *recall* 58.00% dan f_1-*measure* sebesar 57.64% dengan rincian data pada Tabel 6 dan 7.

TABEL VI.
HASIL PENGUJIAN DATA PADA EKSPERIMEN MFCC GTZAN

Fold ke-	Jumlah Data Testing	Benar	Salah
1	200	106	94
2	200	106	94
3	200	116	84
4	200	106	94
5	200	108	92

TABEL VII.
HASIL PENGUJIAN TIAP KELAS DATA PADA EKSPERIMEN MFCC GTZAN

Kelas	Akurasi	Presisi	Recall	*F1-Measure*
Country	55.00%	52.38%	55.00%	53.65%
Jazz	70.00%	57.14%	70.00%	62.91%
Pop	72.50%	69.04%	72.50%	70.72%
Reggae	47.50%	61.29%	47.50%	55.00%
Rock	45.00%	50.00%	45.00%	47.36%

D. *Uji Coba Pada Musik Indonesia*

Selanjutnya, data yang akan diuji adalah dataset lagu Indonesia yang dibuat secara *handmade* dengan dua jenis data masukan yakni vektor dan citra spektogram. Dataset lagu Indonesia ini akan diuji menggunakan arsitektur CNN yang sama seperti eksperimen pada data GTZAN, yang diambil dari arsitektur paling optimal di proses *tuning*

parameter pada Gambar 6 dan 7. Setelah itu data juga akan divalidasi dengan metode yang sama yaitu *cross-validation*. Uji coba ini dilakukan untuk mengetahui apakah dengan arsitektur yang sama, hasil yang didapat dari dataset lagu Indonesia memiliki persamaan dengan dataset lagu berbahasa Inggris atau memiliki perbedaan dari nilai evaluasinya. Data yang digunakan berjumlah 500 lagu Indonesia yang mencakup 5 kelas dengan durasi masing-masing 30 detik.

1) Eksperimen Spektogram
   Data masukan berupa citra spektogram berukuran 128x128. Kemudian validasi dilakukan sama seperti sebelum-sebelumnya yaitu menggunakan *cross-validation* dengan 5 *fold* yang berarti validasi dilakukan sebanyak 5 kali pengujian menggunakan 1000 epoch. Masing-masing *fold* pada uji coba ini terdiri dari 400 data training dan 100 data testing yang diambil secara acak setiap *fold* nya mencakup kelima kelas.

TABEL VIII.
HASIL VALIDASI TERHADAP ARSITEKTUR CNN EKSPERIMEN SPEKTOGRAM LAGU INDONESIA

***Fold* ke-**	**Akurasi**	**Presisi**	***Recall***	***F_1-Measure***
1	70.00%	72.41%	70.00%	70.45%
2	62.00%	64.13%	62.00%	60.49%
3	70.00%	74.58%	70.00%	68.48%
4	64.00%	66.32%	64.00%	63.90%
5	64.00%	68.36%	64.00%	64.11%
Rata-Rata	66.00%	69.16%	66.00%	65.48%

Rata-rata hasil validasi 5 *fold* pada uji coba lagu Indonesia dalam eksperimen spektogram adalah akurasi sebesar 66.00%, presisi sebesar 69.16%, *recall* sebesar 66.00% dan *f_1-measure* sebesar 65.48%. Dari kelima *fold*, hasil validasi terbaik dimiliki oleh *fold* ke 3 dengan akurasi 70.00%, presisi 74.58%, *recall* 70.00% dan *f_1-measure* 68.48% dengan rincian data pada Tabel 9 dan 10.

TABEL IX.
HASIL PENGUJIAN DATA PADA EKSPERIMEN SPEKTOGRAM LAGU INDONESIA

***Fold* ke-**	**Jumlah Data Testing**	**Benar**	**Salah**
1	100	70	30
2	100	62	38
3	100	70	30
4	100	64	36
5	100	64	36

TABEL X.
HASIL PENGUJIAN TIAP KELAS DATA PADA EKSPERIMEN SPEKTOGRAM LAGU INDONESIA

Kelas	**Akurasi**	**Presisi**	**Recall**	***F_1-Measure***
Country	30.00%	85.71%	30.00%	44.45%
Jazz	65.00%	68.42%	65.00%	66.67%
Pop	75.00%	75.00%	75.00%	75.00%
Reggae	85.00%	89.47%	85.00%	87.17%
Rock	95.00%	54.28%	95.00%	69.08%

2) Eksperimen MFCC
Data masukan berupa vektor berukuran 20x11. Validasi dilakukan dengan *cross-validation* menggunakan 5 *fold* yang berarti validasi ini melalui 5 kali proses pengujian dengan 1000 epoch. Masing-masing fold terdiri dari 300 data training dan 200 data testing yang mencakup kelima kelas.

TABEL XI.
HASIL VALIDASI TERHADAP ARSITEKTUR CNN EKSPERIMEN MFCC LAGU INDONESIA

***Fold* ke-**	**Akurasi**	**Presisi**	***Recall***	***F_1-Measure***
1	48.00%	48.21%	48.00%	48.05%
2	52.00%	54.80%	52.00%	52.57%
3	52.50%	53.26%	52.50%	52.72%
4	50.50%	50.53%	50.50%	50.34%
5	49.50%	50.05%	49.50%	49.54%
Rata-Rata	50.50%	51.37%	50.50%	50.64%

Rata-rata nilai validasi yang didapat dari uji coba ini yaitu akurasi sebesar 50.50%, presisi sebesar 51.37%, *recall* sebesar 50.50% dan *f_1-measure* sebesar 50.64%. Hasil yang paling baik didapat oleh *fold* ke 3 yakni akurasi 52.50%, presisi 53.26%, *recall* 52.50% dan *f_1-measure* 50.34% dengan rincian data pada Tabel 12 dan 13.

TABEL XII.
HASIL PENGUJIAN DATA PADA EKSPERIMEN MFCC LAGU INDONESIA

***Fold* ke-**	**Jumlah Data Testing**	**Benar**	**Salah**
1	200	96	104
2	200	104	96
3	200	105	95
4	200	101	99
5	200	99	101

TABEL XIII.
HASIL PENGUJIAN TIAP KELAS DATA PADA EKSPERIMEN MFCC LAGU INDONESIA

Kelas	Akurasi	Presisi	Recall	*F1-Measure*
Country	47.50%	39.58%	47.50%	43.17%
Jazz	62.50%	60.97%	62.50%	61.72%
Pop	45.00%	52.94%	45.00%	48.64%
Reggae	55.00%	52.38%	55.00%	53.65%
Rock	52.50%	53.84%	52.50%	53.16%

Setelah melakukan uji coba pada dataset lagu Indonesia, hasil validasi menunjukkan eksperimen spektogram tetap menghasilkan nilai validasi yang lebih tinggi dibandingkan dengan eksperimen MFCC. Namun kedua eksperimen pada dataset lagu Indonesia memiliki nilai validasi lebih rendah dibandingkan eksperimen dengan dataset GTZAN. Hal ini karena lagu berbahasa Indonesia relatif memiliki irama dan melodi yang hampir mirip di beberapa genre dan hanya dibedakan dengan keberadaan beberapa instrument tertentu sebagai ciri khas dari genre tersebut.

V. KESIMPULAN

Dari penelitian ini dapat diambil kesimpulan bahwa klasifikasi musik berdasarkan genre berhasil dilakukan dengan metode *Convolutional Neural Network* (CNN) melalui proses *tuning parameter* untuk pencarian arsitektur yang paling optimal dan menggunakan 1000 epochs. Kemudian klasifikasi genre menggunakan Spektogram menghasilkan akurasi yang lebih besar dibandingkan dengan klasifikasi genre menggunakan MFCC, karena MFCC merupakan bentuk representasi kompresibel yang didapat dari hasil ekstraksi STFT sehingga menyebabkan data hasil ekstraksi MFCC menjadi lebih sederhana.

Adapun saran untuk penelitian untuk penelitian kedepan dalam klasifikasi genre musik:

1) Menaikkan performa terhadap dataset Indonesia dengan cara menambah lebih banyak lagi data lagu Indonesia dan mencari parameter yang lebih tepat lagi pada arsitektur CNN saat proses *training* agar hasil yang didapat lebih baik dan optimal
2) Teknik ekstraksi fitur dikombinasikan dengan fitur-fitur lain pada musik yang lebih detail seperti *pitch*, timbral, dinamika, tempo dan harmoni agar masing-masing genre memiliki perbedaan pola yang lebih signifikan sehingga menghasilkan akurasi yang lebih baik lagi.

REFERENSI

[1] Djohan. (2006). Terapi musik: Teori dan aplikasi. Yogyakarta: Galangpress

[2] Ridoean, J. A., Sarno, R., & Sunaryono, D. (2017). Rancang Bangun Aplikasi MusicMoo Dengan Metode MIR (Music Information Retrieval) Pada Modul Mood, Genre Recognition, dan Tempo Estimation. *Jurnal Teknik ITS*, *6*(1), 202-206.

[3] Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, *10*(5), 293-302.

[4] Nakashika, T., Garcia, C., & Takiguchi, T. (2012). Local-feature-map integration using convolutional neural networks for music genre classification. In *Thirteenth Annual Conference of the International Speech Communication Association*.

[5] Sigtia, S., & Dixon, S. (2014, May). Improved music feature learning with deep neural networks. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 6959-6963). IEEE.

[6] Peggy, A.T., Hidayat, B. & Atmaja, R.D., " Deteksi Lagu Pada Piano Berdasarkan Ekstraksi Ciri MFCC Dengan Metode *Learning Vector Quantization* dan *Euclidean Distance*". *Teknik Telekomunikasi, Universitas Telkom, 2012*.

[7] Dillak, R. Y., Pangestuty, D. M., & Bintiri, M. G. (2015, July). Klasifikasi Jenis Musik Berdasarkan File Audio Menggunakan Jaringan Syaraf Tiruan Learning Vector Quantization. In *Seminar Nasional Informatika (SEMNASIF)* (Vol. 1, No. 3).

[8] Agustine, T.R., Tritoasmoro, I.I. & Haryatno, J., " Analisis Pengenalan Klasifikasi Musik Berdasarkan Genre Dengan Menggunakan Jaringan Syaraf Tiruan *Self-Organizing Maps* (SOMS)". *Teknik Telekomunikasi, Universitas Telkom, 2009*.

[9] Li, T. L., Chan, A. B., & Chun, A. H. (2010). Automatic musical pattern feature extraction using convolutional neural network. *Genre*, *10*, 1x1.