Dilated-Convolutional Recurrent Neural Network untuk Klasifikasi Genre Musik

http://dx.doi.org/10.28932/jutisi.v10i3.9347

Riwayat Artikel
Received: 06 Juli 2024 | *Final Revision*: 06 Desember 2024 | *Accepted*: 06 Desember 2024

Mochammad Rizqul Fatichin✉ #1, Alfado Rafly Hermawan#2, Raynaldi Anggiat Samuel Siahaan#3, Rarasmaya Indraswari#4

Sistem Informasi, Institut Teknologi Sepuluh Nopember
Sukolilo, Surabaya, 60111, Indonesia
1 6026231032@student.its.ac.id
2 6026231041@student.its.ac.id
3 5026201037@student.its.ac.id
✉Corresponding author: 6026231032@student.its.ac.id

Abstrak — **Dalam era digital, pemanfaatan teknologi untuk mengelompokkan genre musik secara otomatis menjadi sangat penting, terutama untuk aplikasi seperti rekomendasi musik, analisis tren musik, dan pengelolaan perpustakaan musik digital. Penelitian ini mengevaluasi penggunaan *Dilated-Convolutional Recurrent Neural Network* (D-CRNN) dalam mengklasifikasi genre musik. Metode ini menggabungkan keunggulan *Dilated*-CNN dalam menangkap konteks temporal yang lebih panjang dengan kemampuan pengenalan urutan temporal dari CRNN. Data yang digunakan adalah dataset GTZAN yang terdiri dari 1.000 rekaman audio berdurasi 30 detik, dikategorikan ke dalam 10 genre musik. Proses data preprocessing melibatkan konversi rekaman audio menjadi gambar *Mel-Frequency Cepstral Coefficients* (MFCC). Model diuji menggunakan data tanpa augmentasi dan dengan augmentasi, menghasilkan total 15.991 gambar untuk pelatihan. Hasil penelitian menunjukkan bahwa penggunaan D-CRNN dapat meningkatkan akurasi klasifikasi genre musik dibandingkan dengan metode CRNN konvensional.**

Kata kunci— **CRNN; *Dilated Convolutional*; Klasifikasi genre musik; MFCC; *Music Information Retrieval***

Dilated-Convolutional Recurrent Neural Network for Music Genre Classification

Abstract — ***In the digital era, utilizing technology to automatically classify music genres has become very important, especially for applications such as music recommendation, music trend analysis, and digital music library management. This research evaluates the use of Dilated-Convolutional Recurrent Neural Network (D-CRNN) in classifying music genres. This method combines the advantages of Dilated-CNN in capturing longer temporal context with the temporal sequence recognition capability of CRNN. The data used is the GTZAN dataset consisting of 1,000 30-second audio recordings, categorized into 10 music genres. Data preprocessing involved converting the audio recordings into Mel-Frequency Cepstral Coefficients (MFCC) images. The model was tested using data without augmentation and with augmentation, resulting in a total of 15,991 images for training. The results show that the use of D-CRNN can improve the accuracy of music genre classification compared to the conventional CRNN method.***

Keywords— **CRNN; *Dilated Convolutional*; MFCC; *Music Genre Classification*; *Music Information Retrieval***

I. PENDAHULUAN

Dalam era digital, pemanfaatan teknologi untuk mengelompokkan genre musik secara otomatis menjadi sangat penting, terutama dalam aplikasi seperti rekomendasi musik [1], analisis tren musik, dan pengelolaan perpustakaan musik digital. Penelitian di bidang *Music Information Retrieval* (MIR) telah banyak dilakukan, termasuk studi yang mengevaluasi penggunaan berbagai fitur audio untuk klasifikasi genre musik. Salah satu fitur yang diketahui paling efektif dalam analisis audio dan pengenalan genre musik adalah *Mel-Frequency Cepstral Coefficients* (MFCC) [2]. MFCC mampu menangkap karakteristik spektral dari sinyal audio yang berkorelasi dengan persepsi pendengaran manusia, sehingga menjadikannya sangat cocok untuk klasifikasi genre musik. Salah satu tantangan utama dalam MIR adalah melakukan ekstraksi fitur audio secara efektif guna memperoleh informasi genre musik yang akurat.

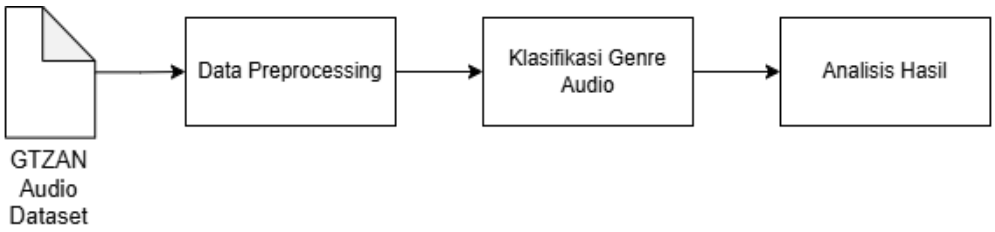
Dalam tugas klasifikasi musik, kerangka kerja *deep learning* seperti *Convolutional Neural Network* (CNN), yang umumnya digunakan untuk pemrosesan visual, telah banyak diterapkan untuk mengekstrak informasi dari representasi audio dengan memodelkan cara otak manusia mempersepsikan musik berdasarkan citra fitur audio [3]. Beberapa penelitian mengintegrasikan *Recurrent Neural Network* (RNN) dan CNN untuk mengidentifikasi pola data audio yang berkembang seiring waktu. Kombinasi RNN dan CNN, yang dikenal sebagai *Convolutional Recurrent Neural Network* (CRNN), telah terbukti memberikan akurasi klasifikasi genre musik yang lebih baik dibandingkan dengan penggunaan CNN saja [1] [4].

Penelitian lain menggunakan *Dilated-CNN* untuk menangkap informasi spektral dari representasi audio secara lebih efektif [5]. Penggunaan kernel konvolusi dengan dilatasi memungkinkan peningkatan ukuran *receptive field* selama proses konvolusi tanpa menambah jumlah parameter. Untuk mencapai ukuran *receptive field* yang sama, konvolusi dengan dilatasi memerlukan jauh lebih sedikit lapisan dibandingkan dengan konvolusi konvensional, sehingga dapat mengurangi risiko *overfitting* yang sering terjadi pada jaringan yang lebih dalam dengan banyak parameter [6]. Dalam analisis data audio, jaringan dengan konvolusi dilatasi digunakan untuk menangkap konteks temporal representasi audio yang lebih panjang dibandingkan dengan konvolusi konvensional. Namun, hingga saat ini, belum ada penelitian yang secara khusus menyelidiki efektivitas operasi dilatasi pada arsitektur CRNN dalam konteks klasifikasi genre musik.

Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi penggunaan Dilated-Convolutional Recurrent Neural Network (D-CRNN) dalam klasifikasi genre musik. Penelitian ini mengusulkan penggabungan Dilated-CNN dan CRNN untuk memadukan keunggulan Dilated-CNN dalam menangkap konteks temporal yang lebih panjang dengan kemampuan CRNN dalam mengenali urutan temporal. Pendekatan ini diharapkan dapat menghasilkan model dengan performa yang lebih unggul dalam klasifikasi genre musik.

II. METODE PENELITIAN

Metode penelitian ini terdiri atas beberapa tahapan yaitu *data preprocessing*, klasifikasi genre audio, dan analisis hasil sebagaimana pada Gambar 1.



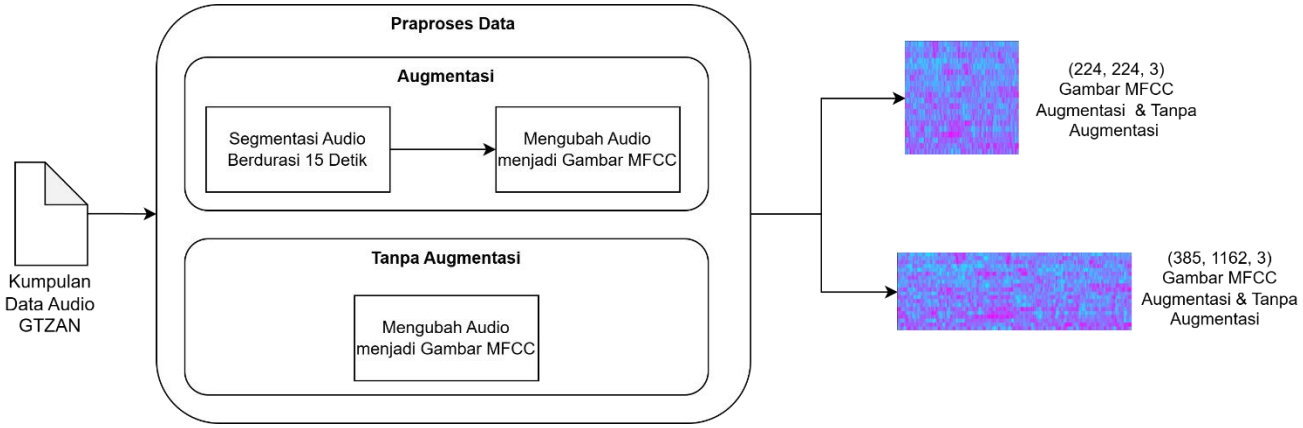
Gambar 1. Metode Penelitian

Dalam penelitian ini, data yang digunakan adalah *GTZAN - Genre Classification Dataset* yang diperoleh melalui situs Kaggle. Dataset GTZAN terdiri atas 1.000 rekaman audio, masing-masing berdurasi 30 detik. Rekaman-rekaman ini dikategorikan ke dalam 10 genre musik yang berbeda, yaitu *blues*, *classical*, *country*, *disco*, *hip-hop*, *jazz*, *metal*, *pop*, *reggae*, dan *rock*. Setiap genre diwakili oleh 100 trek audio. Semua file dalam dataset ini berformat WAV dengan frekuensi sampling 22.050 Hz dan menggunakan saluran mono [7].

A. Data Preprocessing

Tahap *data preprocessing* bertujuan untuk mempersiapkan data audio agar dapat diproses oleh model *deep learning* dalam tugas klasifikasi genre musik. Secara rinci, *data preprocessing* mengolah data audio sehingga menghasilkan keluaran berupa citra MFCC, seperti yang ditunjukkan pada Gambar 2. MFCC merupakan representasi visual dari sinyal suara yang telah diolah. Representasi ini menggambarkan spektrum daya jangka pendek dari suara, yang diperoleh melalui transformasi

kosinus linier dari spektrum daya log pada skala frekuensi mel nonlinier [8] [9]. Penelitian ini menggunakan dua jenis data untuk pengujian, yaitu data dengan proses augmentasi dan data tanpa proses augmentasi. Proses augmentasi dilakukan berdasarkan penelitian sebelumnya yang menerapkan mekanisme *sliding-window* [10], yang terbukti mampu meningkatkan akurasi model pada tugas deteksi emosi dalam pidato.



Gambar 2. Detail Proses *Data Preprocessing*

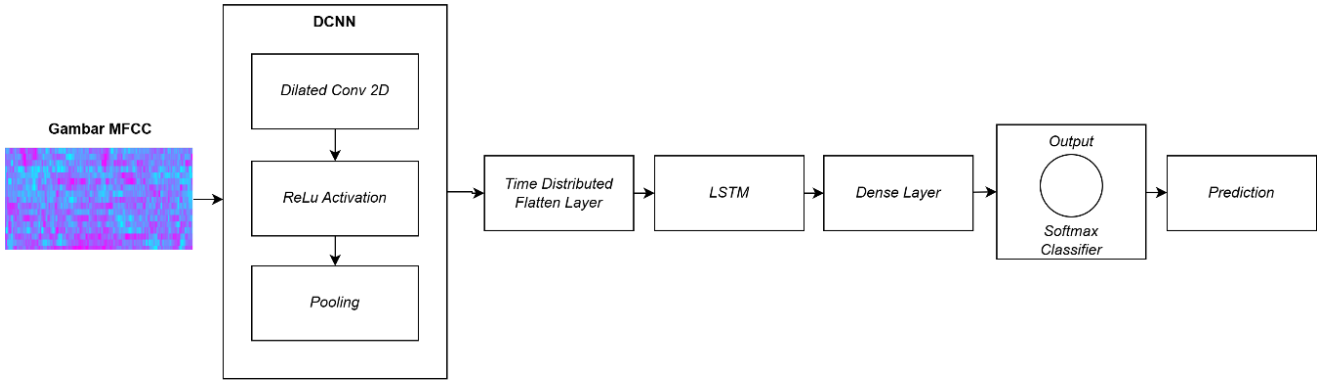
Berdasarkan Gambar 2, pada bagian tanpa augmentasi, data audio langsung diubah menjadi gambar MFCC. Sedangkan pada keluaran yang menggunakan augmentasi, data audio dengan durasi 30 detik dipotong menjadi beberapa segmen berdurasi 15 detik. Pemotongan dilakukan dengan menggeser interval 1 detik setiap kali, sehingga dari satu data audio berdurasi 30 detik dihasilkan 16 data audio berdurasi 15 detik. Selanjutnya, data audio tersebut dikonversi menjadi gambar MFCC. Total gambar MFCC yang diperoleh tanpa proses augmentasi adalah 1000 gambar, sedangkan yang menggunakan proses augmentasi berjumlah 15.991 gambar.

Gambar MFCC yang dihasilkan, baik dengan maupun tanpa augmentasi, kemudian diskalakan sesuai dengan ukuran input dari masing-masing model *deep learning* yang akan diuji. Dalam penelitian ini, terdapat empat keluaran data, yaitu bagian tanpa augmentasi dan augmentasi dengan ukuran input (224, 224, 3) dan (385, 1162, 3). Pada tahapan selanjutnya, gambar MFCC dengan format RGB atau dengan rentang warna dari 0 hingga 255 untuk masing-masing keluaran dilakukan proses normalisasi sehingga menghasilkan rentang warna untuk masing-masing piksel berada di rentang 0 hingga 1. Gambar MFCC yang telah melalui proses prapemrosesan data kemudian dibagi menjadi data latih dan data uji dengan perbandingan 80:20, sehingga data tanpa augmentasi akan berjumlah 800 dan 200 gambar serta yang menggunakan augmentasi berjumlah 12.793 dan 3.198 gambar.

B. Klasifikasi Genre Musik

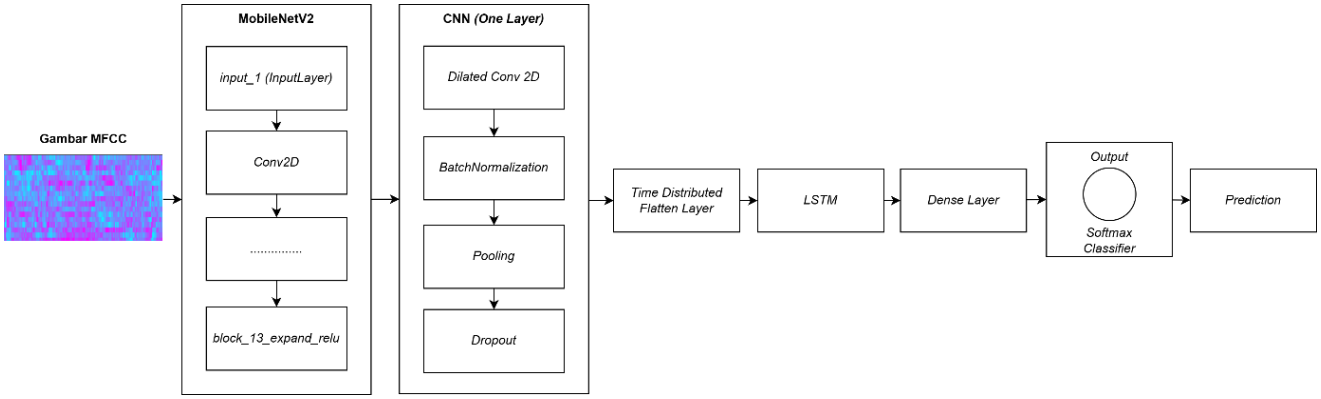
Pada tahap ini, keluaran data dari tahap prapemrosesan akan digunakan untuk pengujian *dilated-convolutional* pada tugas klasifikasi genre musik. Model yang digunakan untuk pengujian adalah model *baseline* dari penelitian sebelumnya, yaitu *recurrent convolutional neural network* (CRNN) yang menggunakan input data MFCC pada data audio untuk identifikasi kasus COVID-19 [11]. Model ini dibandingkan dengan model *Dilated*-CRNN yang digunakan dalam penelitian sebelumnya untuk tugas deteksi peristiwa suara [12], yang saat ini diterapkan untuk pengujian klasifikasi genre musik. Arsitektur model *baseline* dan *dilated-convolutional* RNN dapat dilihat pada Gambar 3.

p-ISSN : 2443-2210
e-ISSN : 2443-2229

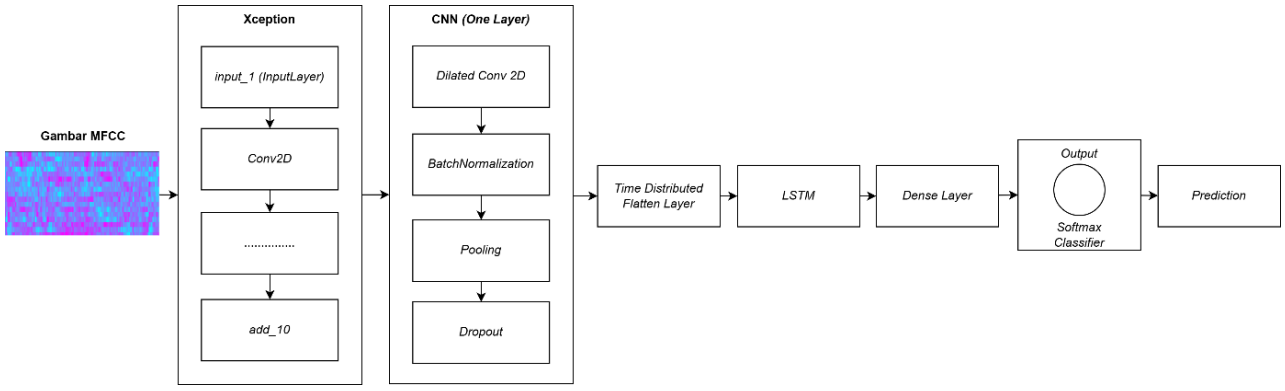


Gambar 3. Arsitektur Dilated-Convolutional RNN

Selain itu, pengujian *dilated-convolutional* juga dilakukan menggunakan pendekatan *transfer learning* dengan model *Xception* dan *MobileNetV2*. *Xception* merupakan pengembangan dari model *Inception* yang mampu mengklasifikasikan gambar ke dalam 1000 kategori objek [13]. *MobileNetV2* adalah model yang dioptimalkan untuk perangkat seluler dan efektif untuk ekstraksi fitur deteksi objek dan segmentasi [14]. Untuk mengurangi kompleksitas model dalam pengujian dilated-convolutional menggunakan transfer learning, dilakukan pemotongan layer. Pada model *Xception*, pemotongan dilakukan pada layer '*add_10'*, sedangkan pada model *MobileNetV2*, pemotongan dilakukan pada layer '*block_13_expand_relu'*. Arsitektur model *MobileNetV2* dan *Xception* dapat dilihat pada Gambar 4 dan Gambar 5.



Gambar 4. Arsitektur *MobileNetV2* + DCRNN



Gambar 5. *Arsitektur Xception* + DCRNN

TABEL 1
VARIABLE TETAP PENELITIAN

Train:Test	80:20
Batch Size	32
Learning Rate	0.001
Epoch	30
Optimizer	Adam

Dalam pengujian ini, beberapa variabel telah ditentukan sebelumnya untuk mempermudah proses pengujian berapa model pada tugas klasifikasi genre musik. Tabel 1 menunjukan beberapa variabel tetap yang digunakan dalam penelitian ini. Untuk mengevaluasi performa *Dilated*-CRNN, model dibandingkan dengan model CRNN konvensional dalam memprediksi data testing. Perbandingan dilakukan berdasarkan nilai akurasi sebagaimana pada persamaan (1), dimana TP A adalah jumlah kelas A yang diprediksi dengan benar, TP B adalah jumlah kelas B yang diprediksi benar, dan seterusnya.

$$Accuracy = \frac{\text{TP A} + \text{TP B} + \text{TP C} + \text{TP D} + \text{TP E}}{\text{Jumlah total data}} \quad (1)$$

III. HASIL & PEMBAHASAN

A. *Uji Coba Model Dasar Dilated-CRNN*

Pada tahap ini, dilakukan uji coba terhadap beberapa konfigurasi model CRNN dan *Dilated*-CRNN. Hasil pengujian disajikan pada Tabel 2 dan Tabel 3 sebagai berikut. Pada Tabel 2, dapat dilihat setelah dilakukan percobaan sebanyak 5 kali dengan mengubah lapisan filter pada *Conv2D*, *LSTM*, dan *Dense*. Tujuan dari penguijan ini adalah mencari konfigurasi terbaik untuk beberapa parameter. Diperoleh Konfigurasi terbaik untuk pengujian adalah model 1 dengan nilai akurasi validasi sebesar 0.275 atau 27,50%.

Pada Tabel 3, dapat dilihat setelah dilakukan percobaan sebanyak 15 kali dengan mengubah lapisan filter, *dilation rate*, dan jumlah lapisan konvolusi. Konfigurasi yang dinilai memiliki akurasi validasi yang tertinggi adalah model 9, yaitu model dengan 3 lapisan konvolusi dengan filter (16-32-32) dan *dilation rate* (2-2-2) dengan akurasi validasi sebesar 35%.

Gambar 6 menunjukkan grafik *training* dan *validation* accuracy serta *training* and *validation loss* dari model dasar D-CRNN terbaik. Pada grafik terlihat bahwa model *overfit*, ditunjukkan dengan *training accuracy* yang hampir mencapai 1 namun *validation accuracy* memiliki nilai yang jauh lebih rendah yaitu 35%. Hal ini dapat disebabkan karena model yang terlalu kompleks atau kurangnya data latih. Model cenderung mempelajari pola-pola yang spesifik pada data latih, termasuk *noise*, sehingga performanya sangat baik di data latih namun tidak *generalizable* ke data validasi.

TABEL 2
HASIL EVALUASI MODEL DASAR CRNN

Model	Lapisan			Akurasi Validasi
	Conv2D (Filter)	*LSTM (Unit)*	*Dense (Unit)*	
1	**16**	**128**	**128**	**0.2750**
2	32			0.2400
3	16	64	128	0.2750
4		32		0.2700
5	16	128	64	0.2600

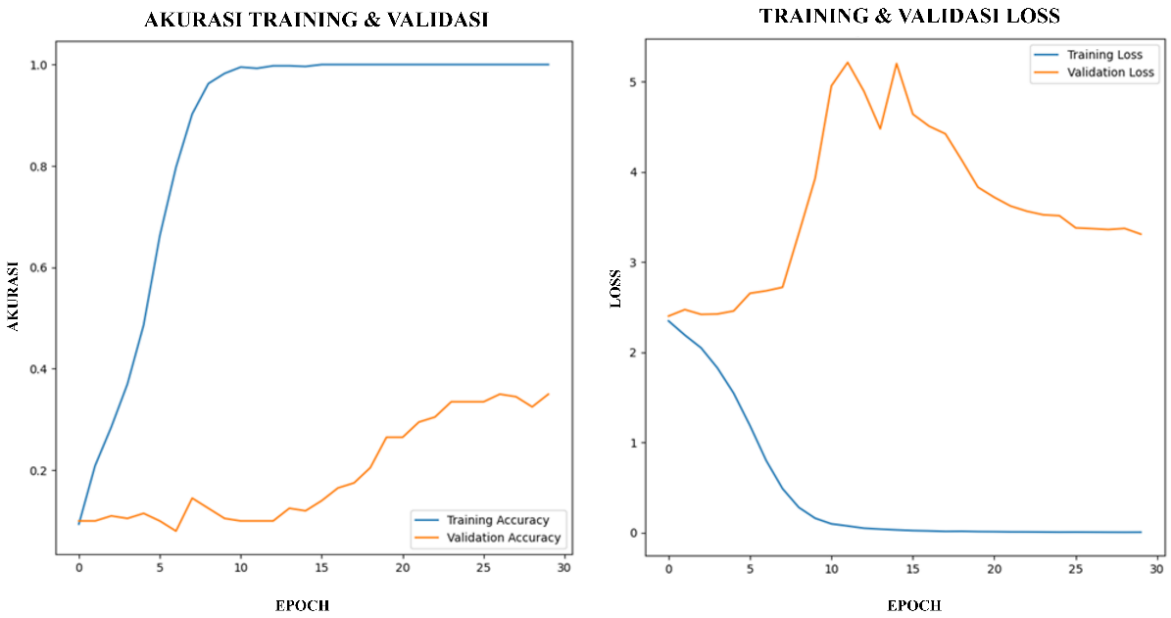
p-ISSN : 2443-2210
e-ISSN : 2443-2229

TABEL 3
HASIL EVALUASI MODEL DASAR D-CRNN

Model	Konvolusi 1		Konvolusi 2		Konvolusi 3		Konvolusi 4		Akurasi Validasi
	Filter	*Dilation Rate*	Filter	*Dilation Rate*	Filter	*Dilation Rate*	Filter	*Dilation Rate*	
1	16	1	-	-	-	-	-	-	0,2750
2		2							0,3350
3		4							0,3150
4		8							0,2750
5	16	1	32	1	-	-	-	-	0,2550
6		2		2					0,2600
7		2		4					0,2600
8	**16**	1	**32**	1	**32**	1	-	-	0,2900
9		**2**		**2**		**2**			**0,3500**
10		2		2		4			0,3250
11		2		4		8			0,3000
12	16	1	32	1	32	1	64	1	0,3050
13		2		2		2		2	0,3050
14		2		4		4		8	0,3100
15		2		4		8		16	0,2750

B. *Uji Coba Model MobileNetV2 + Dilated-CRNN*

Pada percobaan ini, dilakukan *transfer learning* dengan menambahkan model *MobileNetV2* sebelum dilakukan CRNN dan D-CRNN. Dari table 4, dapat dilihat hasil percobaan menggunakan model *MobileNetV2* dengan menambahkan CRNN memiliki akurasi senilai 0.48 dan *F1-Score* senilai 0.49. Sedangkan, setelah dilakukan percobaan sebebanyak 4 kali dengan mengubah *Dilation Rate* serta Filter pada D-CRNN, model terbaik masih memberikan akurasi senilai 0.445 dan *F1-Score* senilai 0.44. Dari percobaan ini dapat disimpulkan penambahan D-CRNN pada model *MobileNetV2* malah memperburuk kinerja model dibandingkan dengan hanya menambahkan CRNN.

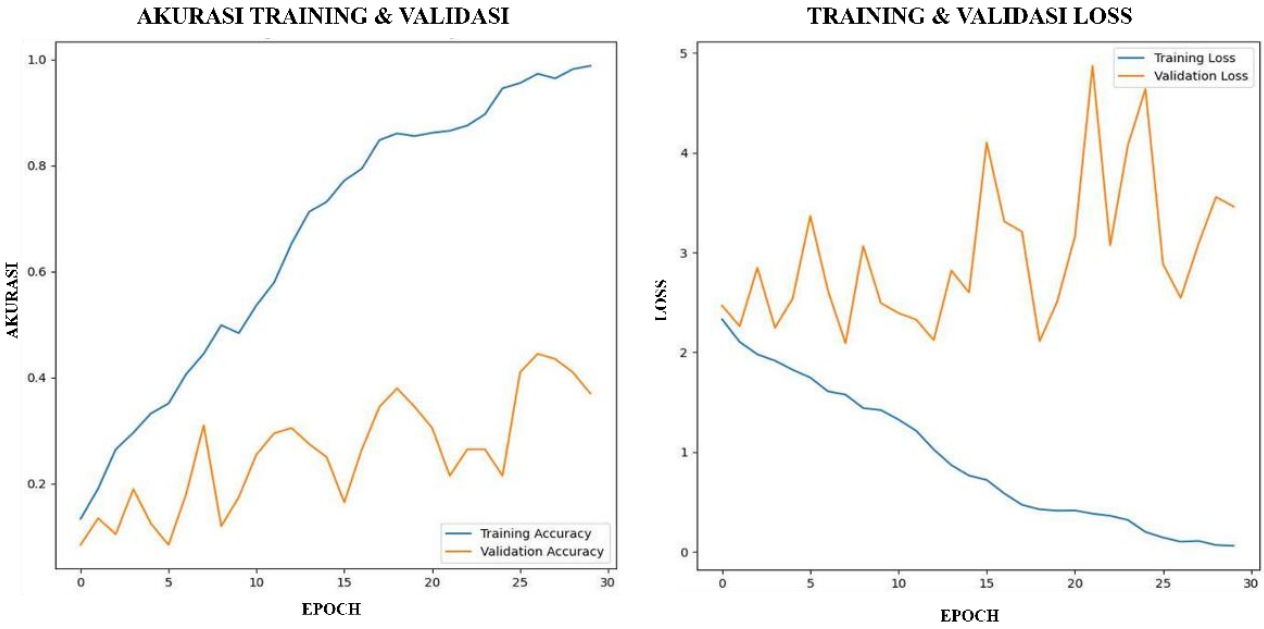


Gambar 6 Grafik *training* dan *validation loss* model dasar D-CRNN terbaik

TABEL 4
HASIL EVALUASI MODEL MOBILENETV2 + D-CRNN

MobileNetV2 + CRNN			
		Akurasi	*F1-Score*
Full-layer MobileNetV2		0,4800	0,4900
MobileNetV2 + D-CRNN			
Dilation Rate	Filter	Akurasi	*F1-Score*
2	16	0,4250	0,4000
	32	**0,4450**	**0,4400**
4	16	0,4600	0,4200
	32	0,4500	0,4300

Dari Gambar 7 juga dapat dilihat bahwa model masih mengalami *overfiting*. Hal ini ditandai dengan nilai *training accuracy* yang hampir mencapai 1 dan nilai *validation Accuracy* yang masih mencapai 0.44. Hal ini dapat disebabkan karena model yang terlalu kompleks atau kurangnya data latih. Model cenderung mempelajari pola-pola yang spesifik pada data latih, termasuk *noise*, sehingga performanya sangat baik di data latih namun tidak *generalizable* ke data validasi.



Gambar 7. Grafik t*raining* dan *validation loss* model *MobileNetV2* + D-CRNN terbaik

C. Uji Coba Model Xception + Dilated-CRNN

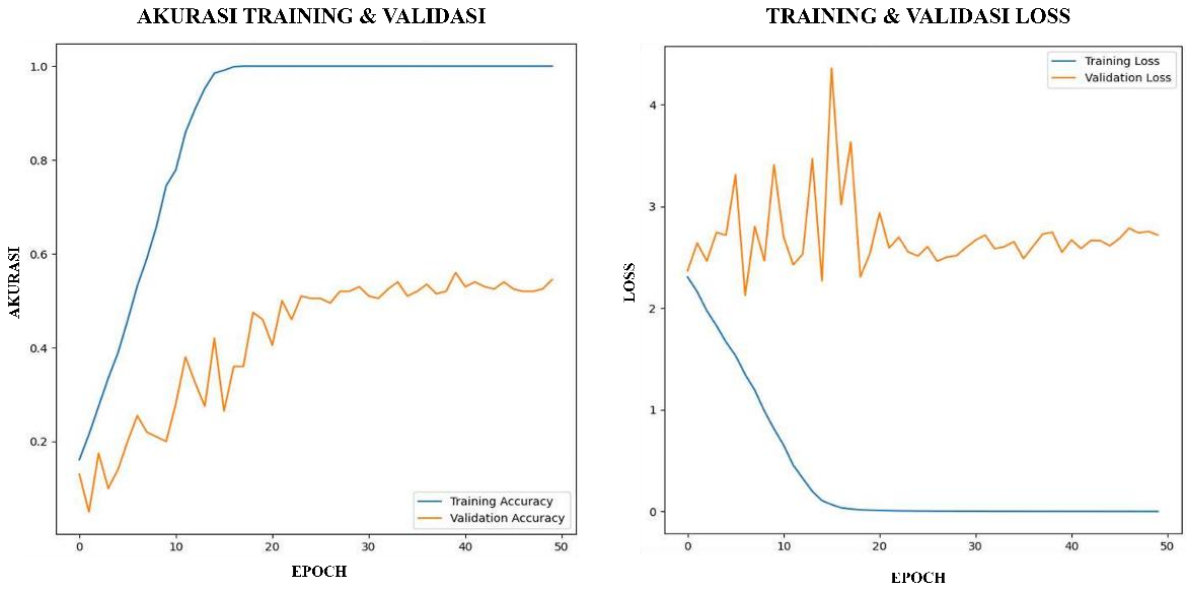
Pada percobaan ini, dilakukan *transfer learning* dengan menambahkan model *Xception* sebelum dilakukan CRNN dan D-CRNN. Dari tabel 5, dapat dilihat hasil percobaan menggunakan model *Xception* dengan menambahkan CRNN memiliki akurasi senilai 0.425 dan F1-Score senilai 0.41. Sedangkan, setelah dilakukan percobaan sebanyak 4 kali dengan mengubah *Dilation Rate* serta Filter pada D-CRNN, model terbaik memberikan akurasi senilai 0.54 dan *F1-Score* senilai 0.56. Dari percobaan ini dapat disimpulkan penambahan D-CRNN pada model *Xception* memberikan hasil yang lebih memuaskan dengan adanya peningkatan nilai akurasi dan *F1-Score*.

p-ISSN : 2443-2210
e-ISSN : 2443-2229

TABEL 5
HASIL EVALUASI MODEL XCEPTION + D-CRNN

Xception + CRNN			
		Akurasi	*F1-Score*
Full-layer Xception		0,4250	0,4100
Xception + D-CRNN			
Dilation Rate	Filter	Akurasi	*F1-Score*
2	16	0,4900	0,5200
	32	**0,5400**	**0,5600**
4	16	0,4900	0,5200
	32	0,5200	0,4800

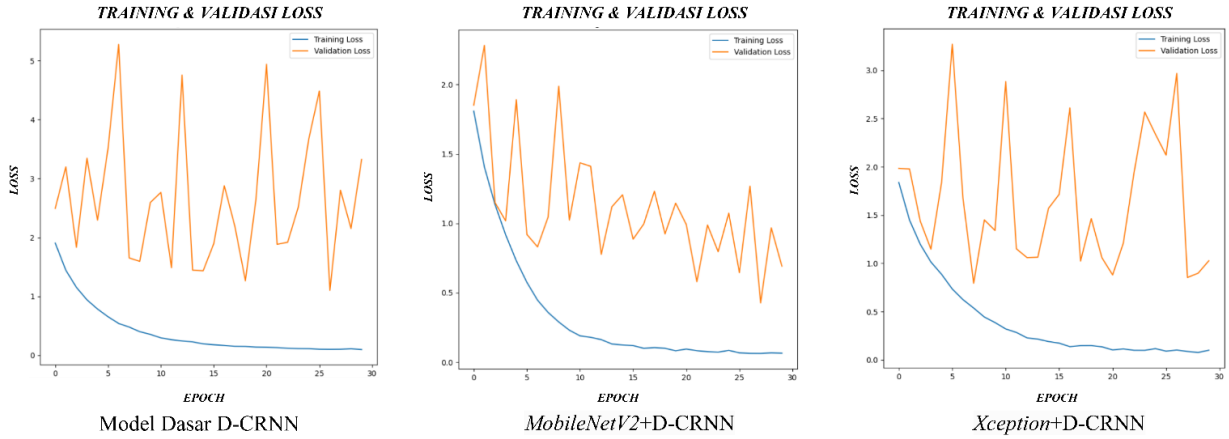
Namun, dari Gambar 8 juga diketahui bahwa model masih mengalami *overfiting*. Hal ini ditandai dengan nilai *training accuracy* yang hampir mencapai 1 dan nilai *validation accuracy* yang masih mencapai 0.54. Hal ini dapat disebabkan karena model yang terlalu kompleks atau kurangnya data latih. Model cenderung mempelajari pola-pola yang spesifik pada data latih, termasuk *noise*, sehingga performanya sangat baik di data latih namun tidak *generalizable* ke data validasi.



Gambar 8. Grafik *training* dan *validation loss* model *Xception* + D-CRNN terbaik

D. *Uji Coba Augmentasi Data*

Percobaan ini dilakukan menggunakan data hasil augmentasi untuk menguji performa dari model D-CRNN. Setiap arsitektur model terbaik yang diperoleh dari percobaan model dasar D-CRNN, *MobileNetV2* + D-CRNN, dan *Xception* + D-CRNN dilatih kembali menggunakan data yang telah diaugmentasi. Gambar 9 menunjukkan *training loss* dan *validation loss* dari hasil pelatihan ketiga model tersebut. Dari gambar tersebut dapat diketahui bahwa augmentasi data dapat menurunkan nilai *validation loss* sehingga model tidak *overfitting*. Hal ini disebabkan karena data augmentasi menambah variasi pada dataset dengan menciptakan versi baru dari data yang sudah ada. Hal ini membuat model lebih terpapar pada berbagai macam skenario yang memungkinkan, sehingga model dilatih untuk mengenali pola yang lebih umum dan *robust*. Akibatnya, model tidak hanya mempelajari data latih secara spesifik, tetapi juga belajar untuk menyesuaikan diri dengan data baru, yang pada akhirnya menurunkan nilai *training loss* dan *validation loss* secara bersamaan. Proses ini membantu mengurangi risiko *overfitting* dengan membuat model lebih *generalizable* terhadap data yang tidak terlihat sebelumnya.



Model Dasar D-CRNN *MobileNetV2*+D-CRNN *Xception*+D-CRNN

Gambar 9. Grafik training dan validation loss tiga model terbaik yang dilatih dengan data hasil augmentasi

Tabel 6 menunjukkan laporan hasil klasifikasi dari ketiga model yang diuji. Berdasarkan tabel tersebut dapat diketahui bahwa ketiga model D-CRNN memiliki performa yang baik untuk mengklasifikasi genre musik. Model dasar D-CRNN memiliki akurasi 77% dengan presisi 79% dan recall 79%. Model *MobileNetV2* + D-CRNN memiliki performa terbaik diantara kedua model lainnya dengan akurasi 89% dan presisi dan recall 89%. Sementara model *Xception* + D-CRNN memiliki akurasi 80% dengan presisi 84% dan recall 80%. Ketiga model ini sangat baik untuk mengklasifikasi seluruh genre pada data, kecuali *rock* pada model dasar D-CRNN dan *Xception* + D-CRNN.

TABEL 6
HASIL KLASIFIKASI MODEL TIGA D-CRNN

	Model Dasar D-CRNN			***MobileNetV2* + D-CRNN**			***Xception* + D-CRNN**		
	Precision	*Recall*	*F1-Score*	*Precision*	*Recall*	*F1-Score*	*Precision*	*Recall*	*F1-Score*
Blues	0,91	0,72	0,80	0,86	0,93	0,89	0,96	0,62	0,75
Classical	0,66	1,00	0,79	0,99	0,96	0,98	0,89	1,00	0,94
Country	0,86	0,62	0,72	0,86	0,92	0,89	0,95	0,48	0,64
Disco	0,75	0,86	0,80	0,80	0,93	0,86	0,79	0,83	0,81
Hiphop	0,83	0,84	0,83	0,94	0,80	0,86	0,71	0,94	0,81
Jaz	0,80	0,78	0,79	0,93	0,94	0,93	0,94	0,84	0,89
Metal	0,70	0,90	0,79	0,80	0,97	0,88	0,84	0,94	0,89
Pop	0,80	0,66	0,72	0,90	0,85	0,88	0,93	0,65	0,77
Reggae	0,96	0,76	0,85	0,96	0,84	0,90	0,83	0,86	0,84
Rock	0,62	0,56	0,59	0,88	0,76	0,81	0,52	0,83	0,64
Accuracy			0,77			0,89			0,80
Macro avg	0,79	0,77	0,77	0,89	0,89	0,89	0,84	0,80	0,80
Weighted avg	0,79	0,77	0,77	0,89	0,89	0,89	0,84	0,80	0,80

Tabel 7 menunjukkan performa beberapa algoritma dalam tugas klasifikasi genre musik sebagai perbandingan dengan penelitian terdahulu. Berdasarkan tabel tersebut dapat diketahui bahwa performa dari model yang diusulkan mengungguli beberapa model sebelumnya. Model dasar D-CRNN yang diusulkan memiliki 3% akurasi lebih tinggi dibandingan dengan model dasar CRNN konvensional. Sementara model modifikasi MobileNetV2 + D-CRNN menghasilkan akurasi yang lebih baik daripada model klasifikasi genre pada penelitian-penelitian sebelumnya. Penggunaan *dilated* pada CRNN ini meningkatan efektivitas model pada proses ekstraksi fitur dalam tugas klasifikasi genre musik menggunakan MFCC.

TABEL 7
PERBANDINGAN PERFORMA MODEL KLASIFIKASI GENRE

Algoritma	Referensi	Input	Dataset	Akurasi
KNN	Jakubec & Chmulik, 2019 [15]	MFCC	GTZAN	69%
CNN	Dong, 2019 [16]	*Spectrogram*	GTZAN	70%
CRNN	Adiyansjah et al., 2019 [4]	*Mel-Spectrogram*	GTZAN	74%
GLR-CRNN	Ashraf et al., 2020 [17]	*Mel-Spectrogram*	GTZAN	87%
D-CRNN	***Proposed***.	MFCC	GTZAN	77%
MobileNetV2 + D-CRNN	***Proposed***.	MFCC	GTZAN	89%
Xception + D-CRNN	***Proposed***.	MFCC	GTZAN	80%

IV. SIMPULAN

Berdasarkan beberapa uji coba yang dilakukan, dapat disimpulkan bahwa penggunaan algoritma *Dilated*-CRNN mampu menghasilkan model klasifikasi genre musik dengan performa yang lebih baik dibandingkan dengan penggunaan CRNN konvensional. Pengintegrasian model *transfer learning* seperti *MobileNetV2* dan *Xception* dengan algoritma *Dilated*-CRNN dapat meningkatkan performa model klasifikasi genre musik dibandingkan dengan model *MobileNetV2* dan *Xception* standar. Hal ini disebabkan karena penggunaan algoritma *Dilated* pada CRNN memperkaya representasi temporal dan spasial dari data audio dengan memperluas *receptive field* secara efisien tanpa kehilangan resolusi. Selain itu, diketahui juga bahwa augmentasi data audio menggunakan mekanisme *sliding window* mampu meningkatkan performa model klasifikasi secara signifikan. Model dasar *Dilated*-CRNN yang terdiri dari tiga lapisan konvolusi dengan filter (16-32-32) dan *dilation rate* (2-2-2) mampu mengklasifikasi genre musik dengan tingkat akurasi 77%. Sementara, model *MobileNetV2* + D-CRNN memiliki akurasi tertinggi yaitu 89%, dan *model Xception* + D-CRNN memiliki akurasi 80%.

DAFTAR PUSTAKA

[1] P. A. Nabila, R. A. Vinarti, E. Riksakomara, and R. Tyasnurita, "Soothing Music Recommendation System for Mothers with Postpartum Depression Using CRNN Method," *ICOSNIKOM 2022 - 2022 IEEE Int. Conf. Comput. Sci. Inf. Technol. Bound. Free Prep. Indones. Metaverse Soc.*, pp. 1–6, 2022.
[2] Y. Singh and A. Biswas, "Robustness of musical features on deep learning models for music genre classification," *Expert Syst. Appl.*, vol. 199, 2022.
[3] K. Zaman, M. Sah, C. Direkoglu, and M. Unoki, "A Survey of Audio Classification Using Deep Learning," *IEEE Access*, vol. 11, no. October, pp. 106620–106649, 2023.
[4] Adiyansjah, A. A. S. Gunawan, and D. Suhartono, "Music recommender system based on genre using convolutional recurrent neural networks," *Procedia Comput. Sci.*, vol. 157, pp. 99–109, 2019.
[5] M. Russo, L. Kraljević, M. Stella, and M. Sikora, "Cochleogram-based approach for detecting perceived emotions in music," *Inf. Process. Manag.*, vol. 57, no. 5, p. 102270, 2020.
[6] S. Kulkarni and R. Rabidas, "Detection of multiple abnormalities of breast cancer in mammograms using a deep dilated fully convolutional neural network," *Comput. Electr. Eng.*, vol. 120, p. 109662, 2024.
[7] A. A. Khamees, H. D. Hejazi, M. Alshurideh, and S. A. Salloum, "Classifying Audio Music Genres Using CNN and RNN," in *Advanced Machine Learning Technologies and Applications*, A.-E. Hassanien, K.-C. Chang, and T. Mincong, Eds., Cham: Springer International Publishing, 2021, pp. 315–323.
[8] B. Tracey *et al.*, "Towards interpretable speech biomarkers: exploring MFCCs," *Sci. Rep.*, vol. 13, no. 1, p. 22787, 2023.
[9] W. Seo, S. H. Cho, P. Teisseyre, and J. Lee, "A Short Survey and Comparison of CNN-Based Music Genre Classification Using Multiple Spectral Features," *IEEE Access*, vol. 12, pp. 245–257, 2024.
[10] M. A. Farias da Silva, R. L. De Carvalho, and T. da S. Almeida, "Evaluation of a Sliding Window mechanism as DataAugmentation over Emotion Detection on Speech," *Acad. J. Comput. Eng. Appl. Math.*, vol. 2, no. 1, pp. 11–18, 2021.
[11] A. Rayan *et al.*, "Utilizing CNN-LSTM techniques for the enhancement of medical systems," *Alexandria Eng. J.*, vol. 72, pp. 323–338, 2023.
[12] Y. Li, M. Liu, K. Drossos, and T. Virtanen, "Sound Event Detection Via Dilated Convolutional Recurrent Neural Networks," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May 2020, pp. 286–290.
[13] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, pp. 1800–1807.
[14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2018, pp. 4510–4520.
[15] M. Jakubec and M. Chmulik, "Automatic music genre recognition for in-car infotainment," *Transp. Res. Procedia*, vol. 40, pp. 1364–1371, 2019.
[16] M. Dong, "Convolutional Neural Network Achieves Human-level Accuracy in Music Genre Classification," pp. 1–6, 2019.
[17] M. Ashraf, G. Geng, X. Wang, F. Ahmad, and F. Abid, "A Globally Regularized Joint Neural Architecture for Music Classification," *IEEE Access*, vol. 8, pp. 220980–220989, 2020.