

PENGLASIFIKASIAN GENRE MUSIK INDONESIA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK

Carl Ray Wairata¹⁾, Ericks Rachmat Swedia²⁾ dan Margi Cahyanti³⁾

^{1,2}Teknik Informatika, Fakultas Ilmu komputer dan Teknologi Informasi, Universitas Gunadarma

³Sistem Informasi, Fakultas Ilmu komputer dan Teknologi Informasi, Universitas Gunadarma

^{1,2,3}Jl. Margonda Raya No.100, Pondok Cina, Depok

E-mail : cray@student.gunadarma.ac.id¹⁾, erick_rs@staff.gunadarma.ac.id²⁾, margi@staff.gunadarma.ac.id³⁾

ABSTRAK

Pada zaman sekarang sudah terdapat banyak sekali teknologi *Artificial Intelligence*. *Artificial Intelligence* sendiri memiliki beberapa sub bab, salah satunya adalah *Machine Learning* dan *Deep Learning* merupakan salah sub bab dari *Machine Learning* itu sendiri. *Convolutional Neural network* (CNN) adalah salah satu jenis *neural network* yang biasa digunakan pada data gambar. Pada penelitian ini akan digunakan untuk mengklasifikasikan *genre* musik dengan cara mengonversi data pada lagu menjadi sebuah gambar yang kita sebut spektogram. Pada penelitian akan mengimplementasikan CNN dalam mengategorikan 3 *genre* musik di Indonesia yakni; dangdut, *Jazz* dan *Pop*. Pada penelitian ini terdapat 100 lagu untuk masing-masing *genre* sebagai data setnya. Tujuan dari penelitian ini sendiri adalah mengoptimalkan tingkat akurasi dalam pengategorian *genre* musik menggunakan model CNN. Berdasarkan penelitian yang sudah dilakukan didapatkan bahwa penggunaan 35 *epoch* memiliki tingkat akurasi yang optimal yakni; tingkat akurasi tes sebesar 81,33% dan tingkat akurasi validasi sebesar 100%. Implementasi ini diharapkan dapat menentukan kategori dalam musik.

Kata Kunci: *Convolutional Neural network, Deep Learning, Lagu, Musik*

1. PENDAHULUAN

Musik dikenal sejak kehadiran manusia modern *Homo sapiens* yakni sekitar 180.000 hingga 100.000 tahun yang lalu. Tidak ada yang tahu kapan manusia mulai mengenal seni dan musik (Mahganna, 2020). Musik merupakan bunyi yang diterima oleh individu dan berbeda-beda berdasarkan sejarah, lokasi, budaya dan selera individu. Definisi sejati tentang musik juga bermacam-macam di antaranya bahwa musik adalah bunyi/kesan terhadap sesuatu yang ditangkap oleh indera pendengar, musik adalah suatu karya seni dengan segenap unsur pokok dan pendukungnya, dan musik adalah segala bunyi yang dihasilkan secara sengaja oleh seseorang atau oleh kelompok individu yang disajikan sebagai musik. Dari beberapa definisi tersebut, maka musik merupakan segala bunyi yang dihasilkan manusia secara sengaja yang disajikan sebagai musik (Ardipal, 2020).

Deep Learning merupakan cabang ilmu *Machine Learning* berbasis Jaringan Saraf Tiruan (JST) atau bisa dikatakan sebagai perkembangan dari JST. (Jaringan Saraf Tiruan) terdiri dari berbagai *layer* dan beberapa neuron pada masing-masing *layer*. Kedua hal tersebut tidak dapat ditentukan menggunakan aturan yang pasti dan berlaku berbeda-beda pada data yang digunakan.

Dalam *Deep Learning*, sebuah komputer belajar mengklasifikasi secara langsung dari gambar atau suara. Metode *Deep Learning* menggunakan CPU dan RAM dalam proses komputasi, dan juga memanfaatkan GPU

sehingga proses komputasi data yang besar dapat berlangsung lebih cepat.

Convolutional Neural network (CNN) adalah salah satu jenis *neural network* yang biasa digunakan pada data gambar atau citra. Pada penelitian ini akan digunakan untuk mengklasifikasikan *genre* musik dengan cara mengonversi data pada lagu menjadi sebuah suara yang kita sebut spektogram (Lina, 2019). CNN dapat belajar langsung dari citra sehingga mengurangi beban dari pemrograman. CNN termasuk dalam jenis *Deep Neural network* karena dalamnya tingkat jaringan dan banyak diimplementasikan dalam data citra. CNN memiliki dua metode; yakni klasifikasi menggunakan *feed forward* dan tahap pembelajaran menggunakan *back propagation*. Cara kerja CNN memiliki kesamaan pada MLP, namun dalam CNN setiap *neuron* ditampilkan dalam bentuk dua dimensi (Triwijoyo, 2019).

Pada penelitian lain yang berjudul "Klasifikasi Genre Lagu dengan Fitur Akustik Menggunakan Metode *K-Nearest Neighbor*" bertujuan untuk mengukur kedekatan dua buah lagu akan digunakan persamaan *normalized cross correlation* (NCC) yang menggantikan persamaan penghitungan jarak pada metode K-NN. Penelitian ini menggunakan 5 *genre* lagu yaitu *Electronic*, *Rock*, *Jazz*, *Pop* dan *Hip Hop*, di mana masing-masing *genre*-nya memiliki 100 lagu. Dari 100 lagu tersebut 70 lagu akan digunakan sebagai data latih dan 30 lagu sebagai data uji. Penelitian ini dilakukan pengujian terhadap 3 *Genre* lagu yaitu *Pop*, *Jazz* dan *Dance*.

Diharapkan dari hasil penelitian ini dapat mengklasifikasikan *genre* musik dengan menggunakan persamaan *normalized cross correlation* (NCC) yang menggantikan persamaan penghitungan jarak pada metode K-NN yang telah dilakukan oleh beberapa peneliti sebelumnya seperti Nugraha, Atmaja, Ramatryana dengan judul "*Simulasi Dan Analisis Klasifikasi Genre Musik Berbasis Fft Dan K-nearest Neighbor*" dan Rahmawati, Magdalena, Ramatryana dengan judul "*Perbandingan dan analisis k-nearest neighbor dan linear discriminant analysis untuk klasifikasi genre musik*".

2. RUANG LINGKUP

Ruang lingkup dari penelitian ini adalah membandingkan model CNN dengan metode dari 2 penelitian sebelumnya dan mengetahui cara untuk mengoptimalkan tingkat akurasi baik pada proses latihan maupun validasi.

Pada penelitian ini peneliti mengimplementasikan model *Convolutional Neural network* dalam pengklasifikasian 3 *genre* musik di Indonesia yakni; dangdut, *Jazz* dan *Pop*. *Data set* yang digunakan berjumlah 300 lagu di mana 1 *genre*-nya mempunyai masing-masing 100 data lagu. Pada proses latihan, terdapat 90 data digunakan untuk data latihan dan 10 data yang akan digunakan untuk data tes. Pada proses validasi 80 data digunakan untuk data latihan dan terdapat 20 data akan digunakan untuk memvalidasi data. Berdasarkan dari penelitian ini menghasilkan nilai akurasi tes dan nilai akurasi validasi dan *output* yang dihasilkan diharapkan dapat sesuai dengan kebutuhan.

3. BAHAN DAN METODE

Pada bagian ini menjelaskan bahan-bahan dan metode apa yang akan digunakan dalam penelitian ini, di antaranya :

3.1 Musik

Musik dikenal sejak kehadiran manusia modern Homo sapiens yakni sekitar 180.000 hingga 100.000 tahun yang lalu. Tidak ada yang tahu kapan manusia mulai mengenal seni dan musik (Mahganna, 2020). Musik merupakan bunyi yang diterima oleh individu dan berbeda-beda berdasarkan sejarah, lokasi, budaya dan selera individu. Definisi sejati tentang musik juga bermacam-macam di antaranya bahwa musik adalah bunyi/kesan terhadap sesuatu yang ditangkap oleh indera pendengar, musik adalah suatu karya seni dengan segenap unsur pokok dan pendukungnya, dan musik adalah segala bunyi yang dihasilkan secara sengaja oleh seseorang atau oleh kelompok individu yang disajikan sebagai musik. Dari beberapa definisi tersebut, maka musik merupakan segala bunyi yang dihasilkan manusia secara sengaja yang disajikan sebagai musik (Ardipal, 2020).

3.2 Deep Learning

Pemelajaran dalam (bahasa *Inggris: Deep Learning*) atau sering dikenal dengan istilah pemelajaran struktural mendalam (bahasa *Inggris: deep structured learning*) atau pemelajaran hierarki (bahasa *Inggris: hierarchical learning*) adalah salah satu cabang dari ilmu pemelajaran mesin (bahasa *Inggris: Machine Learning*) yang terdiri algoritme pemodelan abstraksi tingkat tinggi pada data menggunakan sekumpulan fungsi transformasi non-linear yang ditata berlapis-lapis dan mendalam (Haixiang, 2017).

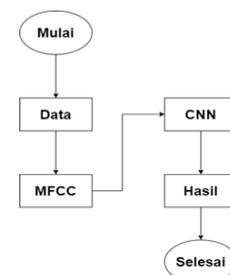
Deep Learning merupakan cabang ilmu *Machine Learning* berbasis Jaringan Saraf Tiruan (JST) atau bisa dikatakan sebagai perkembangan dari JST. Dalam *Deep Learning*, sebuah komputer belajar mengklasifikasi secara langsung dari gambar atau suara. *Convolutional Neural network* (CNN/ConvNet) (Sena, 2017) adalah salah satu algoritma *Deep Learning* yang merupakan pengembangan dari *Multilayer Perceptron* (MPL) (Azizah, 2018) yang dirancang untuk mengolah data dalam bentuk dua dimensi, misalnya gambar atau suara. CNN dapat belajar langsung dari citra sehingga mengurangi beban dari pemrograman. Metode *Deep Learning* menggunakan CPU dan RAM dalam proses komputasi, dan juga memanfaatkan GPU sehingga proses komputasi data yang besar dapat berlangsung lebih cepat.

3.3 Supervised Learning

Teknik *supervised learning* merupakan teknik yang bisa diterapkan pada pembelajaran mesin yang bisa menerima informasi yang sudah ada pada data dengan memberikan label tertentu. Diharapkan teknik ini bisa memberikan target terhadap *output* yang dilakukan dengan membandingkan pengalaman belajar di masa lalu.

3.4 Gambaran Umum

Klasifikasi 3 *genre* lagu indonesia menggunakan python 3.8.5 dan dijalankan menggunakan Anaconda *Prompt* (*miniconda3*). Pada bagian gambar 1 ini menjelaskan proses dari pengklasifikasian *genre* lagu menggunakan *Deep Learning* dengan model CNN. Penelitian ini sendiri terdiri dari 4 tahap yaitu: pengumpulan data, MFCC, CNN dan hasil.



Gambar 1. Alur Sistem

3.5 Data Lagu

Pada Tahapan ini telah diunduh 300 lagu indonesia, di mana 300 lagu-lagu tersebut terbagi menjadi 3 kelompok *genre* yaitu : Dangdut, *Pop* dan *Jazz*. Sebelum dilakukan proses ekstraksi, data format lagu-lagu tersebut harus dikonversi terlebih dahulu menjadi format *.wav*. Hal ini dilakukan agar python bisa membaca data lagu-lagu tersebut dan melakukan ekstraksi data.

Dari 300 lagu yang sudah diunduh akan dibagi menjadi 2 direktori yakni; direktori latih dan direktori tes dengan perbandingan 90% untuk direktori latih dan 10% untuk direktori tes pada tabel 1.

Tabel 1. Data Lagu

Genre	Jumlah Lagu	Latih	Tes
Dangdut	100	90	10
Jazz	100	90	10
Pop	100	90	10

3.6 Mel-Frequency Cepstrum Coefficients

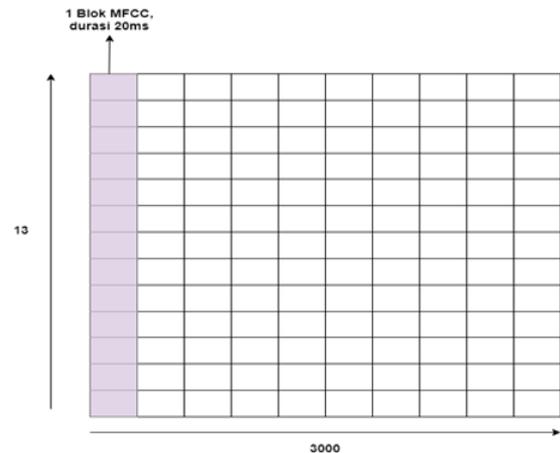
MFCC adalah representasi yang didefinisikan sebagai nilai *real-cepstrum* dari *short-time* signal yang berasal dari spektrogram sinyal suara (Kasih, 2018). Perbedaan dari *cepstrum* adalah bahwa pada MFCC skala frekuensi *nonlinier* digunakan, yang mendekati perilaku sistem pendengaran manusia (Gumelar, 2020).

Spektrogram sendiri telah berhasil mendapatkan berbagai macam frekuensi yang ada dalam sinyal suara, namun mengikuti struktur *koklea* pada pendengaran manusia yang menangkap frekuensi pada beberapa titik yang berbeda ditemukan formulasi yang mendekati perilaku sistem pendengaran manusia yang disebut dengan Skala mel (*mel-scale*) (Nurarinda, 2020).

Skala mel pada dasarnya adalah melakukan filter pada koefisien frekuensi yang didapatkan dari spektrogram. Filter yang dilakukan adalah *linier* pada frekuensi 1000Hz dan *logaritmik* pada frekuensi di atasnya (Tomchuk, 2018) formula (1) digunakan untuk menghitung mel-Frekuensi (Swedia, 2018).

$$M(f) = 2595 + \text{LOG} \left(1 + \left(\frac{f}{700} \right) \right) \quad (1)$$

Gelombang suara dari lagu tersebut akan dibuat 13 MFCC agar dapat membentuk matriks. Dimana pada sumbu Y menyatakan MFCC-nya dan sumbu X menyatakan durasi lagu yang ditentukan. Dari setiap blok yang terbentuk memiliki nilai-nilai koefisien. Dari gambar III.4 akan terbentuk 13 x 3000 *pixel*. Nilai 3000 sendiri ada setiap 1 Blok MFCC memiliki waktu yakni 20ms, sehingga 1 detiknya terdapat 50 *pixel*. Dikarenakan lagu berdurasi 60 detik sehingga terdapat 3000 *pixels* yang dapat dilihat pada 2 gambar (Astuti, 2019).



Gambar 2. Model MFCC

Dalam menyimpan data menggunakan modul *Json*, data akan disimpan dalam sebuah file. Data dari 300 lagu akan disimpan dan dibagi menjadi 3 bagian data yaitu: Mapping, MFCC dan Label.

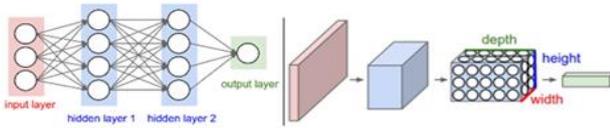
3.7 Convolutional Neural network

Convolutional Neural network adalah salah satu metode *Machine Learning* dari pengembangan *Multi Layer Perceptron* (MLP) yang didesain untuk mengolah data dua dimensi. Sebuah *perceptron multilayer* (MLP) adalah kelas dari *feed forward* jaringan saraf tiruan (JST). Istilah MLP digunakan secara ambigu, kadang-kadang secara longgar untuk setiap JST *feed forward*, kadang-kadang hanya mengacu pada jaringan yang terdiri dari beberapa lapisan *perceptrons* (dengan aktivasi ambang). *Perceptron multilayer* terkadang dalam bahasa sehari-hari disebut sebagai jaringan saraf "vanilla", terutama jika mereka memiliki satu lapisan tersembunyi. CNN terbukti berhasil mengungguli metode *Machine Learning* lainnya seperti SVM pada kasus klasifikasi objek pada citra. MLP terdiri dari setidaknya tiga lapisan node: lapisan masukan, lapisan tersembunyi, dan lapisan keluaran. Kecuali *node input*, setiap node merupakan neuron yang menggunakan fungsi aktivasi *nonlinear*. MLP menggunakan teknik pembelajaran yang diawasi yang disebut *propagative mundur* untuk pelatihan. Beberapa lapisan dan aktivasi *non-liniernya* membedakan MLP dari *perceptron linier* itu dapat membedakan data yang tidak dapat dipisahkan secara *linier* (Triwijoyo, 2019).

CNN termasuk dalam jenis *Deep Neural network* karena dalamnya tingkat jaringan dan banyak diimplementasikan dalam data citra. CNN memiliki dua metode; yakni klasifikasi menggunakan *feed forward* dan tahap pembelajaran menggunakan *back propagation*. Cara kerja CNN memiliki kesamaan pada MLP, namun dalam CNN setiap neuron ditampilkan dalam bentuk dua dimensi (Triwijoyo, 2019).

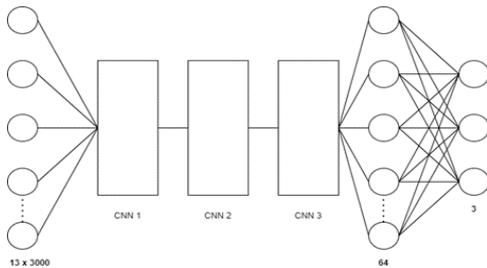
Konsep CNN ataupun kerja CNN memiliki kesamaan pada MLP, di dalam CNN setiap neuron mempresentasikan dalam bentuk dua dimensi, tidak

seperti MLP yang setiap neuron hanya berukuran satu dimensi (Darmanto, 2019) seperti gambar 3.



Gambar 3. Alur CNN

Pada CNN, data yang dipropagasikan pada jaringan adalah data dua dimensi, sehingga operasi linier dan parameter bobot pada CNN berbeda. Pada CNN operasi linier menggunakan konvolusi, sedangkan bobot tidak lagi satu dimensi saja, namun berbentuk empat dimensi yang merupakan kumpulan *kernel konvolusi*. Penelitian ini dibuat 3 model CNN, dimana pada convolutional layer 1, 2 dan 3 menggunakan aktivasi ReLu. Setelah proses CNN selesai akan dilanjutkan ke *hidden layer* (Abdulbar, 2019). Pada penelitian ini digunakan 64 simpul pada *hidden layer*. Proses dilanjutkan pada *output layer* yang berjumlah 3 simpul karena kita mengklasifikasikan 3 *genre* lagu (Putra, 2020), seperti pada Gambar 4.



Gambar 4. Model CNN

3.8 Arsitektur Jaringan CNN

JST (Jaringan Saraf Tiruan) terdiri dari berbagai *layer* dan beberapa neuron pada masing-masing *layer*. Kedua hal tersebut tidak dapat ditentukan menggunakan aturan yang pasti dan berlaku berbeda-beda pada data yang digunakan.

Sebuah jaringan tanpa *hidden layer* dapat memetakan persamaan linier, sedangkan jaringan dengan satu atau dua *hidden layer* dapat memetakan sebagian besar persamaan pada data yang sederhana. Namun pada data yang lebih kompleks, MLP memiliki kekurangan atau keterbatasan. Pada permasalahan jumlah *hidden layer* di bawah tiga *layer*, terdapat pendekatan untuk menentukan jumlah neuron pada masing-masing *layer* untuk mendekati hasil optimal. Penggunaan *layer* di atas dua pada umumnya tidak direkomendasikan dikarenakan akan menyebabkan *over fitting* serta kekuatan *back propagation* berkurang secara signifikan. Berkembangnya *Deep Learning*, ditemukan bahwa untuk mengatasi kekurangan MLP dalam menangani kasus data kompleks, diperlukan fungsi untuk mentransformasikan data *input* menjadi bentuk yang lebih mudah dimengerti

oleh MLP. Hal tersebut memicu berkembangnya *Deep Learning* di mana dalam satu model diberikan beberapa *layer* untuk melakukan transformasi data sebelum data diolah menggunakan metode klasifikasi. Hal tersebut memicu berkembangnya model *neural network* dengan jumlah *layer* di atas tiga. Namun dikarenakan fungsi *layer* awal sebagai metode ekstraksi fitur, maka jumlah *layer* dalam sebuah DNN tidak memiliki aturan universal dan berlaku berbeda-beda tergantung *dataset* yang digunakan. Sebuah CNN terdiri dari beberapa *layer*. Berdasarkan arsitektur LeNet5, terdapat empat macam *layer* utama pada sebuah CNN namun yang diterapkan pada TA ini hanya tiga macam lapisan antara lain (Nurhikmat, T, 2018) :

1. Convolution Layer
2. Subsampling Layer
3. Fully Connected Layer

4. PEMBAHASAN

Tahap MFCC adalah tahap di mana dilakukan ekstraksi data terhadap lagu-lagu yang telah diunduh dan diubah menjadi bentuk numerik atau dalam bentuk matriks lebih tepatnya. Proses ekstraksi data ini menggunakan *python* dan beberapa modul yang harus diinstal terlebih dahulu yaitu : *os*, *Librosa*, *Math* dan *Json*. Proses ekstraksi data pada gambar 5.

```
Indonesia\DANGDUT\80 Juta - Nella Kharisma.wav, segment : 0
Indonesia\DANGDUT\80 Juta - Nella Kharisma.wav, segment : 1
Indonesia\DANGDUT\80 Juta - Nella Kharisma.wav, segment : 2
Indonesia\DANGDUT\80 Juta - Nella Kharisma.wav, segment : 3
Indonesia\DANGDUT\80 Juta - Nella Kharisma.wav, segment : 4
Indonesia\DANGDUT\80 Juta - Nella Kharisma.wav, segment : 5
Indonesia\DANGDUT\80 Juta - Nella Kharisma.wav, segment : 6
Indonesia\DANGDUT\80 Juta - Nella Kharisma.wav, segment : 7
Indonesia\DANGDUT\80 Juta - Nella Kharisma.wav, segment : 8
Indonesia\DANGDUT\80 Juta - Nella Kharisma.wav, segment : 9
Indonesia\DANGDUT\Cukup Rogo Isun - Nella Kharisma.wav, segment : 0
Indonesia\DANGDUT\Cukup Rogo Isun - Nella Kharisma.wav, segment : 1
Indonesia\DANGDUT\Cukup Rogo Isun - Nella Kharisma.wav, segment : 2
Indonesia\DANGDUT\Cukup Rogo Isun - Nella Kharisma.wav, segment : 3
Indonesia\DANGDUT\Cukup Rogo Isun - Nella Kharisma.wav, segment : 4
Indonesia\DANGDUT\Cukup Rogo Isun - Nella Kharisma.wav, segment : 5
Indonesia\DANGDUT\Cukup Rogo Isun - Nella Kharisma.wav, segment : 6
Indonesia\DANGDUT\Cukup Rogo Isun - Nella Kharisma.wav, segment : 7
Indonesia\DANGDUT\Cukup Rogo Isun - Nella Kharisma.wav, segment : 8
Indonesia\DANGDUT\Cukup Rogo Isun - Nella Kharisma.wav, segment : 9
```

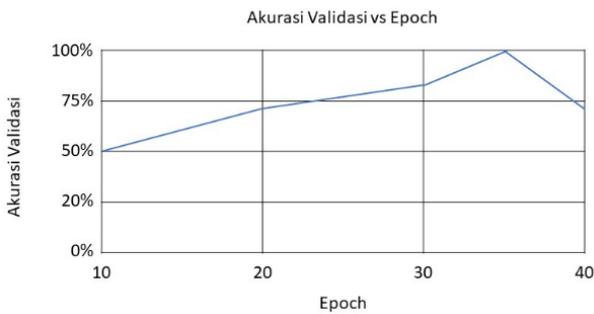
Gambar 5. Ekstraksi Data Lagu

Setelah proses ekstraksi data selesai dilakukan maka akan menghasilkan Mapping MFCC dan Label pada gambar 6 di mana 1 Blok MFCC memiliki waktu yakni 20ms.

epoch. Nilai akurasi validasi dari data *array* di tabel 2 pertama sampai dengan kesepuluh antara lain, 10 *Epoch* memiliki nilai akurasi validasi 50%, 20 *Epoch* memiliki nilai akurasi validasi 70%, 30 *Epoch* memiliki nilai akurasi validasi 80%, 40 *Epoch* memiliki nilai akurasi validasi 70%, dan 35 *epoch* memiliki nilai akurasi validasi 100% (Tabel 6.). Nilai akurasi tes pada *epoch* 35 sendiri adalah 81.33%. *Output* dari Tabel penelitian bisa dilihat contohnya terdapat pada gambar 9, sedangkan *output* nilai akurasi tes dapat di lihat pada gambar 10 dan grafik akurasi validasi berdasarkan jumlah *epoch* bisa dilihat pada gambar 10.

```
Epoch 26/35
64/64 [*****] - 2s 34ms/step - loss: 0.3229 - accuracy: 0.9274 - val_loss: 0.6113 - val_accuracy: 0.7984
Epoch 27/35
64/64 [*****] - 2s 35ms/step - loss: 0.3408 - accuracy: 0.9181 - val_loss: 0.6326 - val_accuracy: 0.7985
Epoch 28/35
64/64 [*****] - 2s 36ms/step - loss: 0.3165 - accuracy: 0.9384 - val_loss: 0.6195 - val_accuracy: 0.8059
Epoch 29/35
64/64 [*****] - 2s 34ms/step - loss: 0.3194 - accuracy: 0.9284 - val_loss: 0.6099 - val_accuracy: 0.8044
Epoch 30/35
64/64 [*****] - 2s 35ms/step - loss: 0.3108 - accuracy: 0.9223 - val_loss: 0.6051 - val_accuracy: 0.8015
Epoch 31/35
64/64 [*****] - 2s 35ms/step - loss: 0.2996 - accuracy: 0.9402 - val_loss: 0.611 - val_accuracy: 0.8059
Epoch 32/35
64/64 [*****] - 2s 35ms/step - loss: 0.2929 - accuracy: 0.9412 - val_loss: 0.6188 - val_accuracy: 0.8059
Epoch 33/35
64/64 [*****] - 2s 34ms/step - loss: 0.2984 - accuracy: 0.9407 - val_loss: 0.6232 - val_accuracy: 0.8164
Epoch 34/35
64/64 [*****] - 2s 34ms/step - loss: 0.2725 - accuracy: 0.9551 - val_loss: 0.6221 - val_accuracy: 0.7956
Epoch 35/35
64/64 [*****] - 2s 34ms/step - loss: 0.2749 - accuracy: 0.9531 - val_loss: 0.6574 - val_accuracy: 0.7896
10-Oct [*****]
10/10 [*****] - 0s 6ms/step - loss: 0.6701 - accuracy: 0.8133
accuracy on test is 81.33%
```

Gambar 9. Nilai Akurasi Tes



Gambar 10. Grafik Akurasi Validasi

5. KESIMPULAN

Model *Convolutional Neural network* merupakan salah satu metode *Machine Learning* yang banyak diimplementasikan dalam data citra. Untuk penggunaan CNN dalam data *audio*, pertama data *audio* akan diubah menjadi sebuah gambar dengan koordinat x sebagai Amplitudo dari *audio* itu sendiri, sedangkan koordinat y sebagai durasi waktu dari *audio* tersebut. Setelah diubah menjadi gambar maka akan dilakukan proses *Deep Learning* menggunakan model CNN.

Setelah melakukan percobaan pada data *array* pertama sampai dengan yang kesepuluh menggunakan 35 *epoch* merupakan yang paling efektif karena mendapatkan akurasi validasi 100% dan akurasi tes sebesar 81.33%. Jika dibandingkan dengan penelitian sebelumnya yang menggunakan metode *K-Nearest Neighbor* yang mendapatkan akurasi sebesar 63%, maka metode CNN lebih mendapatkan tingkat akurasi yang lebih tinggi. Berbeda dengan penelitian berbasis FFT dan Support Vector Machine, model CNN mendapatkan tingkat akurasi yang sedikit lebih rendah.

6. SARAN

Saran untuk penelitian ini adalah ke depannya bisa memperbanyak *dataset* agar akurasi yang didapatkan lebih baik lagi, dikarenakan pada penelitian ini lagu-lagu yang pada *dataset* tidak memiliki batas yang jelas antara satu *genre* dengan *genre* lainnya. Selain itu diharapkan penelitian ini dapat dikolaborasi dengan penelitian-penelitian sebelumnya untuk mendapatkan hasil yang lebih baik lagi seperti menguji dengan metode dari Algoritma *deep learning* lainnya.

7. DAFTAR PUSTAKA

- Abdulbar, H., Adikara, P. P., & Adinugroho, S. 2019. Klasifikasi Genre Lagu dengan Fitur Akustik Menggunakan Metode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN, 2548, 964X.
- Ardipal, A. 2020. Pemanfaatan Perangkat Teknologi dalam Pembelajaran Musik Berbasis Tematik sebagai Peningkatan Keterampilan Abad 21 Bagi Guru Sekolah Dasar. *Musikolastika: Jurnal Pertunjukan Dan Pendidikan Musik*, 2(2), 77-84.
- Astuti, D. 2019. Aplikasi Identifikasi Suara Hewan Menggunakan Metode Mel-Frequency Cepstral Coefficients (MFCC). *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 1(2), 26-34
- Azizah, L. M., Umayah, S. F., & Fajar, F. 2018. Deteksi Kecacatan Permukaan Buah Manggis Menggunakan Metode *Deep Learning* dengan Konvolusi *Multilayer*. *Semesta Teknika*, 21(2), 230-236.
- Darmanto, H. (2019). Pengenalan Spesies Ikan Berdasarkan Kontur Otolith Menggunakan *Convolutional Neural network*. *Joined Journal (Journal of Informatics Education)*, 2(1), 41-59.
- Gumelar, A. B., Yuniarno, E. M., Anggraeni, W., Sugiarto, I., Kristanto, A. A., & Purnomo, M. H. 2020. Kombinasi Fitur Multispektrum Hilbert dan Cochleagram untuk Identifikasi Emosi Wicara. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 9(2), 180-189.
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. 2017. *Learning from class-imbalanced data: Review of methods and*

- applications. *Expert Systems with Applications*, 73, 220-239.
- Kasih, P. 2018. Voice Recognition untuk Sistem Keamanan PC Menggunakan Metode MFCC dan DTW. *Generation Journal, Department Of Informatics Engineering*, 2(1), 57-68.
- Lina, Qolbiyatul. 2019. Apa itu *Convolutional Neural network*?. <https://medium.com/@16611110/apa-itu-convolutional-neural-network-836f70b193a4> (diakses pada tanggal 6 Januari 2021).
- Mahganna, S. 2020. OLIOREANG: Entitas Ritmis dan Melodi Mandar (Vol. 1). Gerbang Visual.
- Nurarinda, T. A. P., Sahertian, J., & Mahdiyah, U. 2020. Rancangan Sistem Identifikasi Jenis Burung Kicau Berdasarkan Suara Burung dengan Mel Frequency Cepstrum Coefficiens (MFCC). In *Seminar Nasional Inovasi Teknologi e-ISSN* (Vol. 2549, No. 7952, p. 7952).
- Nurhikmat, T. 2018. Implementasi *Deep Learning* Untuk Image Classification Menggunakan Algoritma *Convolutional Neural network* (CNN) Pada Citra Wayang Golek.
- Putra PN, C. S. 2020. Deteksi Penyakit Tanaman Padi Dengan Menggunakan *Convolutional Neural network* (Doctoral dissertation, UPN Veteran Jawa Timur).
- Sena, S. 2017. Pengenalan *Deep Learning* Part 7 *Convolutional Neural network* CNN. <https://medium.com/@samuelsena/pengenalan-deep-learning-part-7-convolutional-neural-network-cnn-b003b477dc94> (Diakses pada 12 April 2021).
- Swedia, E. R., Mutiara, A. B., & Subali, M. 2018. Deep learning long-short term memory (LSTM) for Indonesian speech digit recognition using LPC and MFCC Feature. In *2018 Third International Conference on Informatics and Computing (ICIC)* (pp. 1-5). IEEE.
- Triwijoyo, B. K. 2019. Model Fast Tansfer Learning pada Jaringan Syaraf Tiruan Konvolusional untuk Klasifikasi Gender Berdasarkan Citra Wajah. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 18(2), 211-221.
- Tomchuk, K. K. 2018. Spectral Masking in MFCC Calculation for Noisy Speech. In *2018 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)* (pp. 1-4). IEEE.