

Perancangan Sistem Keamanan Lingkungan Pengenal Suara Kukul dengan Menggunakan Metode Deep Learning

I Putu Adi Yuda¹, I Gusti Ngurah Anom Cahyadi Putra², I Ketut Gede Suhartana³, I Komang Ari Mogi⁴, Made Agung Raharja⁵, Luh Arida Ayu Rahning Putri⁶

¹Teknik Informatika, Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Jimbaran, Badung, Indonesia

adiyuda418@gmail.com
anom.cp@unud.ac.id
ikg.suhartana@unud.ac.id
arimogi@gmail.com
made.agung@unud.ac.id
rahningputri@unud.ac.id

Abstract

Advances in information technology provide benefits for people's lives today. Before the existence of information technology, people still use traditional communication media called kukul. Kukul is a communication tool that is used by hitting. Over time, people began to leave this culture because of the many other information systems used as communication media. However, nowadays people in their teens still do not know what the sound of the kukul means. This is due to the absence of a kukul voice recognition website. The method of data collection in this study was the observation method of a kukul worker located in Denjalan Subvillage, Batubulan, Sukawati, Gianyar. To build a kukul voice recognition system, the author uses a deep learning method. In this system there are 2 processes, namely training and classification. The training is used for the system to learn to recognize the sound of the kukul and the classification to determine the category of the kukul sound. Based on the classification carried out, the results obtained is testing accuracy 85%.

Keywords: Kukul, Deep Learning, Sistem Pengenal Suara Kukul, CNN, MFCC

1. PENDAHULUAN

Kemajuan teknologi informasi memberikan manfaat untuk kehidupan masyarakat saat ini. Teknologi informasi dapat digunakan untuk meningkatkan keamanan masyarakat. Sebelum adanya teknologi informasi, masyarakat masih menggunakan media komunikasi secara tradisional yang disebut kukul. Kukul merupakan alat komunikasi yang digunakan dengan cara dipukul. Cara memukul kukul memberikan pertanda bahwa sedang terjadi sesuatu di lingkungan tersebut. Seiring berjalannya waktu, masyarakat mulai meninggalkan kebudayaan tersebut karena adanya banyak sistem informasi lain yang digunakan sebagai media komunikasi. Namun sekarang bagi kalangan masyarakat di usia remaja masih belum mengenal apa arti suara kukul tersebut. Hal tersebut dapat menghambat dalam upaya menciptakan lingkungan masyarakat yang aman dan damai. Sehingga dibutuhkan suatu sistem yang dapat mengenali arti dari suara kukul agar masyarakat lebih cepat menanggapi suatu kejadian yang sedang terjadi di masyarakat. Belum adanya sistem yang dapat mengklasifikasikan suara kukul untuk melakukan pengenalan suara kukul

Hal tersebut membuat penulis untuk membuat suatu sistem yang digunakan untuk penerjemah suara kukul. Dalam sistem ini akan dibuat berfungsi sebagai mengolah suara hasil rekaman ke suara yang lebih jernih tanpa adanya suara lain selain suara kukul.

Dengan menerapkan sistem keamanan lingkungan pengenal suara kulkul *metode deep learning* diharapkan dapat membantu mengenali suara kulkul sehingga lebih cepat dalam menanggapi suatu pertanda jika terdengar bunyi kulkul. Selain itu, dengan teknologi ini diharapkan melestarikan kebudayaan tradisional yang sudah diwariskan turun-temurun oleh leluhur dalam pengenalan suara kulkul.

2. Metode Penelitian

Metode penelitian ini dilakukan secara sistematis sebagai acuan untuk melakukan penelitian agar memperoleh hasil yang dapat menjadi suatu solusi untuk menyelesaikan permasalahan yang akan diteliti.

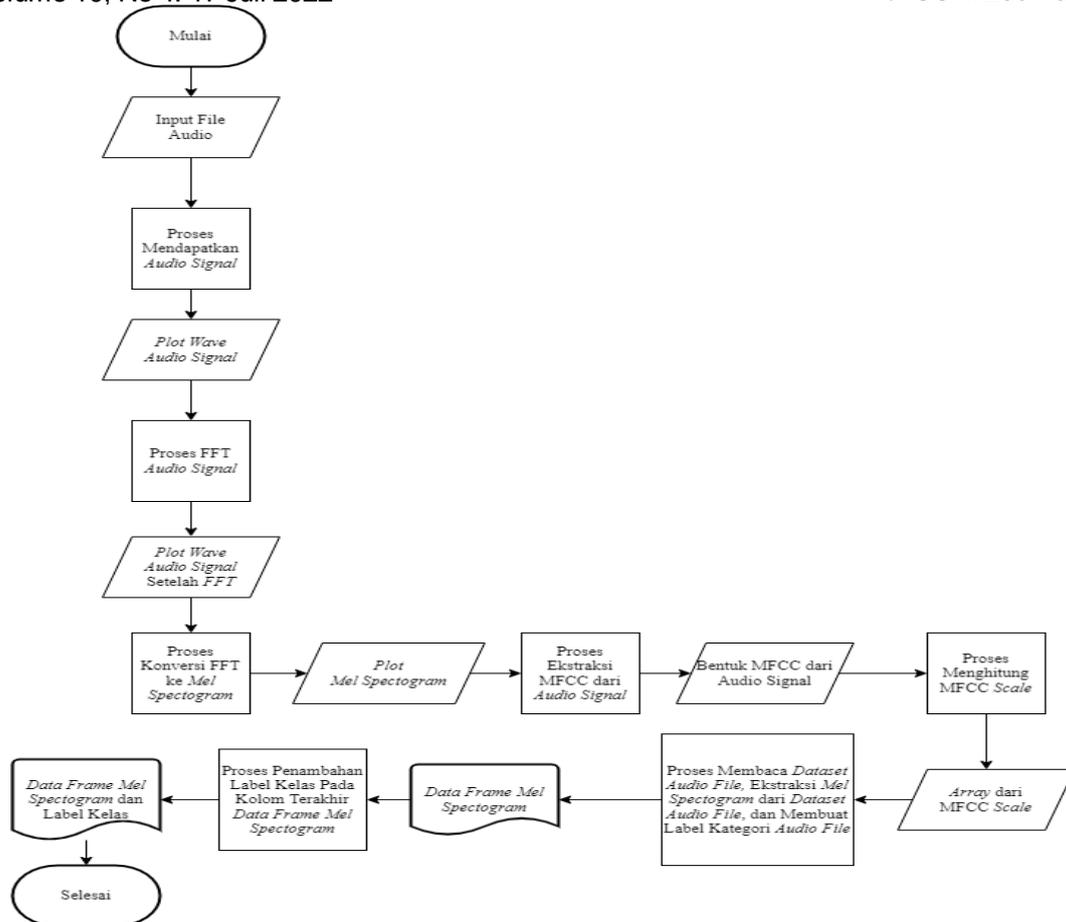
2.1. Pengumpulan Data

Data yang digunakan pada penelitian ini menggunakan data sekunder yang diperoleh dengan metode observasi, dengan bekerja sama dengan pihak kelurahan Banjar untuk memperoleh data *audio* serta informasi suara kulkul tersebut. Data tersebut berupa file *audio* berformat *.wav* dengan jumlah *dataset* suara kulkul berjumlah 50 dengan masing – masing kategori berjumlah 10. Data tersebut diperoleh dari seorang tukang kulkul yang berada di Banjar Denjalan, Batubulan, Sukawati, Gianyar.

2.2. Proses Ekstraksi Ciri

Proses ekstraksi ciri adalah proses untuk mengetahui ciri dari file audio dengan mencari bentuk sinyal audionya. Metode yang digunakan untuk melakukan proses ekstraksi ciri adalah MFCC (*Mel Frequency Cepstral Coefficient*). MFCC adalah salah satu teknik fitur ekstraksi yang dapat menyerupai pendengaran manusia dalam bentuk digital. Keunggulan MFCC adalah kemampuan untuk mengenali suara sangat penting bagi pengenalan suara dan menghasilkan data seminimal mungkin tanpa menghilangkan informasi penting. [1]

Sebelum mendapatkan hasil MFCC dari suatu *audio*, dilakukan proses pengambilan *audio signal*, *Fast Fourier Transform* (FFT), dan *Mel Spectrogram*. Pada hasil ekstraksi fitur yang digunakan pada proses klasifikasi menggunakan CNN adalah *Mel Spectrogram*. Berikut ini merupakan alur dari penelitian yang dilakukan:



Gambar 1. Proses Ekstraksi Ciri

Berdasarkan proses ekstraksi ciri pada gambar 1 digunakan metode MFCC dimana diawali dengan memasukkan file suara kulkul dengan ekstensi .wav. Lalu akan dilakukan proses ekstraksi ciri dengan MFCC. Proses MFCC memiliki 4 proses. Proses tersebut dapat dijelaskan sebagai berikut:

1. Audio Signal
Audio Signal adalah bentuk konversi file *audio* dalam bentuk gelombang. Pada gelombang terdapat amplitudo dan waktu.
2. Fast Fourier Transform (FFT)
Fast Fourier Transform (FFT) adalah pemrosesan *audio signal* dimana akan dilakukan konversi dari waktu ke frekuensi.
3. Mel Spectrogram
Mel Spectrogram adalah proses untuk mengkonversikan *audio signal* dari amplitudo ke *decibel* (dB). Nilai *decibel* ditentukan berdasarkan warna pada *mel spectrogram*.
4. MFCC
MFCC merupakan proses terakhir yang digunakan pada proses ekstraksi ciri setelah dilakukan proses untuk mendapatkan *mel spectrogram*. Setelah didapatkan bentuk MFCC dari audio signal, maka akan dilakukan untuk menghitung MFCC Scale dan akan mendapatkan keluaran dalam bentuk array. Lalu, dilakukan proses untuk membaca dataset dan melakukan proses ekstraksi mel spectrogram dan hasil data frame akan diekspor ke dokumen dalam format .xml. Setelah dilakukan proses ekspor, akan dilakukan proses penambahan label kelas pada kolom terakhir yang akan diekspor lagi ke dokumen .xml sebelumnya.

2.3. Kecerdasan Buatan

Kecerdasan Buatan merupakan salah satu bidang dalam ilmu komputer yang ditujukan pada pembuatan *software* dan *hardware* yang dapat berfungsi sebagai sesuatu yang dapat berpikir seperti manusia. Kecerdasan buatan banyak digunakan untuk memecahkan berbagai masalah seperti bisnis, robotika, bahasa alami, matematika, game, persepsi, diagnosis medis, teknik, analisis keuangan, analisis sains, dan penalaran.

Pada awal diciptakannya, komputer hanya difungsikan sebagai alat hitung saja. Namun seiring dengan perkembangan jaman, maka peran komputer semakin mendominasi kehidupan umat manusia. Komputer tidak lagi hanya digunakan sebagai alat hitung, lebih dari itu, komputer diharapkan untuk dapat digunakan untuk mengerjakan sesuatu yang seperti halnya yang dilakukan oleh manusia pada umumnya. Manusia menjadi pandai dalam menyelesaikan permasalahan yang ada pada hidupnya karena manusia mempunyai pengetahuan dan pengalaman pengetahuan diperoleh dari belajar. Semakin banyak pengetahuan dan pengalaman yang di dapat oleh manusia, maka manusia lebih mampu dalam menyelesaikan permasalahan. Namun bekal pengetahuan saja tidak cukup, manusia harus menggunakan akal nya untuk melakukan penalaran, mengambil kesimpulan dan keputusan berdasarkan pengetahuan dan pengalaman yang mereka miliki. Tanpa memiliki kemampuan untuk menalar dengan baik, manusia dengan segudang pengalaman dan pengetahuan tidak akan dapat menyelesaikan masalah dengan baik. Demikian pula, dengan kemampuan menalar yang sangat baik, namun tanpa bekal pengetahuan dan pengalaman yang memadai, manusia juga tidak akan bisa menyelesaikan masalah dengan baik. [2]

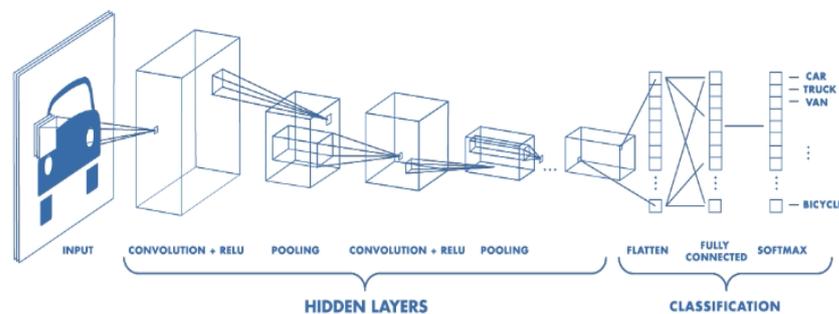
2.4. Deep Learning

Deep Learning adalah salah satu bagian dari *machine learning*. *Deep Learning* menerapkan jaringan saraf tiruan dimana memiliki fungsi layaknya pada sistem saraf pada manusia. Dalam metode ini mesin mempelajari *dataset* yang didapat sebagai pengalaman sistem. [3]

Deep Learning digunakan untuk mengembangkan aplikasi seperti halnya pengenalan *genre music*. Dimana sistem mempelajari secara mendalam dataset yang ada dengan melakukan melakukan proses ekstraksi data dan analisa data. Dalam melakukan pengenalan dibutuhkan data uji seperti data musik untuk nantinya diketahui *genrenya*. Dalam klasifikasi data uji menggunakan suatu probabilitas dimana data tersebut lebih memungkinkan ke arah *genre* apa musik tersebut.

2.5. Convolutional Neural Network (CNN)

Convolutional Neural Network merupakan salah satu jenis *neural network* yang biasanya digunakan dalam pengolahan data *image* maupun *audio*. *Convolutional* berfungsi melakukan *filter* pada gambar maupun *audio*. *Convolutional Neural Network* memiliki beberapa *layer* yang difungsikan untuk melakukan *filter* pada setiap prosesnya. Pada proses *training* terdapat 3 tahapan yaitu *Convolutional layer*, *Pooling layer*, dan *Fully connected layer*.



Gambar 2. Proses *Convolutional Neural Network*

Pada gambar 2 merupakan arsitektur dari CNN. Pada tahapan pertama, yaitu *convolutional layer*. Pada *convolutional layer*, seluruh data yang menyentuh lapisan konvolusional akan mengalami proses konvolusi. lapisan akan mengkonversi setiap *filter* ke seluruh bagian data masukan dan menghasilkan sebuah *activation map* atau *feature map 2D*. *Filter* yang terdapat pada *Convolutional Layer* memiliki panjang, tinggi(*pixels*) dan tebal sesuai dengan *channel* data masukan. Setiap *filter* akan mengalami pergeseran dan operasi “*dot*” antara data masukan dan nilai dari *filter*. Lapisan konvolusional secara signifikan mengalami kompleksitas model melalui optimalisasi *outputnya*.

Pooling Layer merupakan tahap setelah *Convolutional Layer*. *Pooling Layer* terdiri dari sebuah *filter*

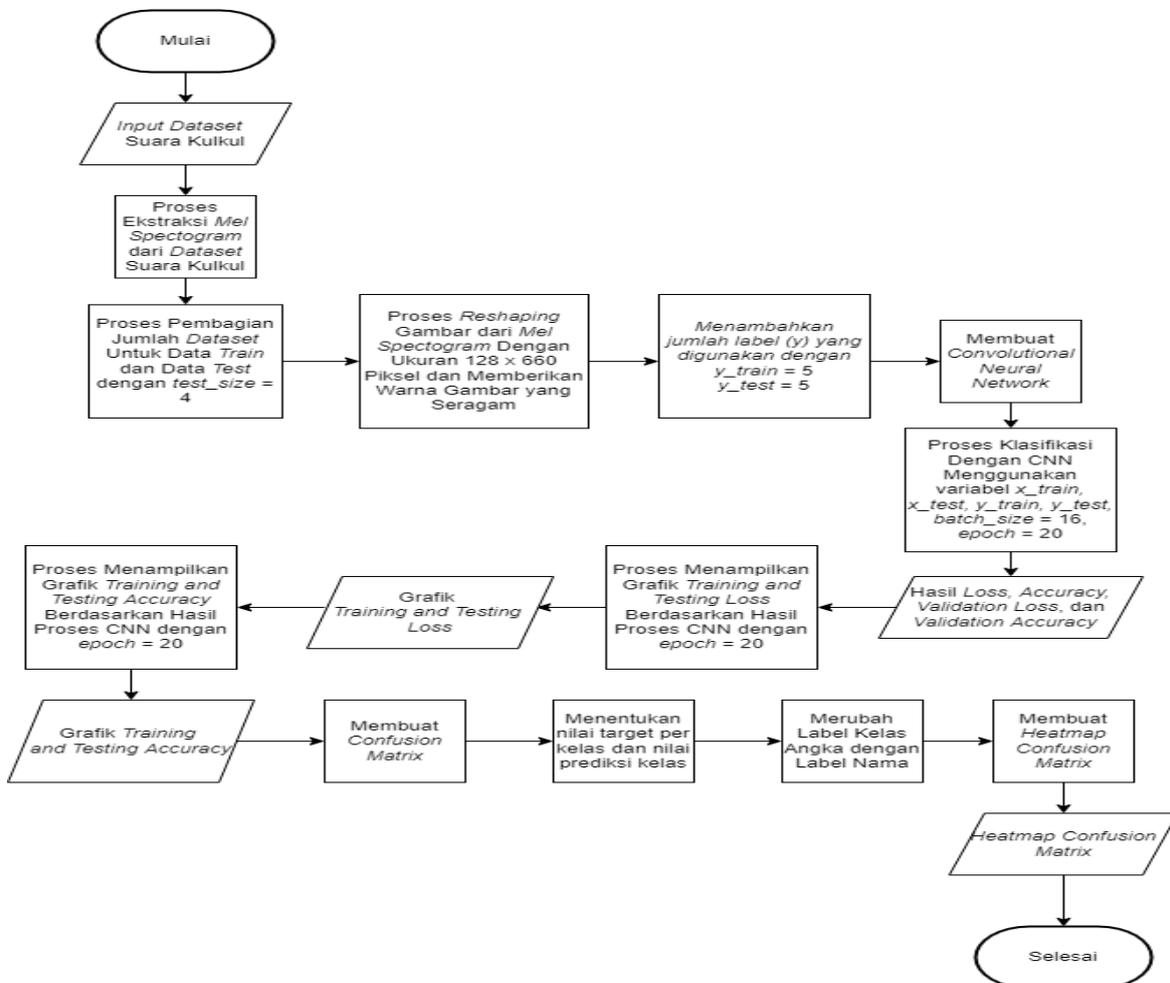
dengan ukuran dan *stride* tertentu. Setiap pergeseran akan ditentukan oleh jumlah *stride* yang akan digeser pada seluruh *area feature map* atau *activation map*. Dalam penerapannya, *pooling Layer* yang biasa digunakan adalah *Max Pooling* dan *Average Pooling*.

Feature map yang dihasilkan oleh tahap sebelumnya berbentuk *multidimensional array*. Sehingga, Sebelum masuk pada tahap *Fully Connected Layer*, *Feature Map* tersebut akan melalui proses “*flatten*” atau *reshape*. Proses *flatten* menghasilkan sebuah vektor yang akan digunakan sebagai *input* dari *Fully Connected Layer*. *Fully Connected Layer* memiliki beberapa *Hidden Layer*, *Action Function*, *Output Layer* dan *Loss Function*.

Dropout merupakan salah satu usaha untuk mencegah terjadinya *overfitting* dan juga mempercepat proses *learning* pada sistem klasifikasi *audio*. *Overfitting* adalah kondisi dimana hampir semua data yang telah melalui proses *training* mencapai persentase yang baik, tetapi terjadi prediksi yang tidak sesuai. Dalam sistem kerjanya, *Dropout* menghilangkan sementara suatu *neuron* yang berupa *Hidden Layer* maupun *Visible Layer* yang berada didalam jaringan. (Santoso & Ariyanto, 2018)

Dalam *convolutional neural network*, menggunakan *Epoch* pada proses *training* maupun *testing*. *Epoch* adalah ketika seluruh dataset sudah melalui proses *training* pada *Neural Network* sampai dikembalikan ke awal dalam satu putaran. Dalam *Neural Network* satu *epoch* itu terlalu besar dalam proses pelatihan karena seluruh data diikutkan kedalam proses *training* sehingga akan membutuhkan waktu cukup lama. Agar mempermudah dan mempercepat proses *training*, biasanya data *rate* dibagi *per batch* (*Batch Size*). Penentuan nilai dari *batch size* biasanya tergantung banyak sampel. [4]

CNN digunakan untuk proses klasifikasi dari jenis audio. Jenis masukan yang akan digunakan pada proses CNN adalah *mel spectrogram* dari audio. Pola *audio signal* pada *mel spectrogram* akan digunakan sebagai pembanding jenis audio. Berikut merupakan alur proses CNN.



Gambar 3. Flowchart Proses Klasifikasi Menggunakan CNN

Pada gambar 3 merupakan flowchart dari proses klasifikasi menggunakan CNN. Pertama,

memasukan *dataset* suara kulkul. Lalu, akan dilakukan proses untuk mengekstrak *mel spectrogram* dari *dataset* suara kulkul berdasarkan *data frame* yang tersimpan pada dokumen berekstensi .xml. Setelah itu, dilakukan proses pembagian dataset menjadi 2 jenis data, yaitu *data train* dan *data test* dimana *data test* yang digunakan pada masing – masing kelas adalah berjumlah 4. Lalu, dilakukan proses *reshaping mel spectrogram* dengan warna yang seragam. Setelah dilakukan *reshaping*, dilakukan proses untuk membuat kelas dengan $y_{train} = 5$ dan $y_{test} = 5$ dimana y_{train} adalah label *training* dan y_{test} adalah label *testing*. Lalu, membuat CNN dengan variable x_{train} , y_{train} , x_{test} , y_{test} , $batch_size = 16$ dan $epoch = 20$. Setelah membuat CNN, akan dilakukan proses klasifikasi untuk mendapatkan keluaran berupa nilai *loss*, *accuracy*, *validation_loss*, dan *validation accuracy*. Nilai *loss* dan *accuracy* merupakan hasil *training*. Sedangkan *validation_loss* dan *validation_accuracy* merupakan hasil *testing*. Hasil *training* dan *testing* akan dipresentasikan 2 jenis grafik, yaitu grafik *training and testing loss* dan *training and testing accuracy*. Setelah dilakukan proses membuat grafik, akan dilakukan pengujian model dengan *confusion matrix* dengan membuat target per kelas dan merubah nama label menjadi label sesuai jenis suara kulkul. Setelah itu, dilakukan proses untuk membuat dan menampilkan *heatmap confusion matrix* sebagai keluaran terakhir dari proses klasifikasi.

2.6 Confusion Matrix

Confusion matrix sering digunakan sebagai metode untuk melakukan evaluasi model yang digunakan untuk melakukan klasifikasi data. *Confusion matrix* merupakan *matrix* yang berisikan nilai aktual dan prediksi dari proses klasifikasi. [5] Dalam *confusion matrix*, terdapat 4 istilah, yaitu *true positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). *True Positive* adalah data *positive* yang terprediksi benar. *True Negative* adalah data *negative* yang terprediksi benar. *False Positive* adalah data *negative* tetapi terprediksi sebagai data *positive*. *False Negative* adalah data *positive* yang terprediksi *negative*. Dibawah ini merupakan tabel dari *confusion matrix*.

Tabel 1. Confusion Matrix

	Prediksi (+)	Prediksi (-)
Aktual (+)	TP	FN
Aktual (-)	FP	FN

Pada *confusion matrix* terdapat perhitungan *precision*, *recall*, dan akurasi. Berikut merupakan beberapa persamaan yang digunakan dalam perhitungan *precision*, *recall*, dan akurasi.

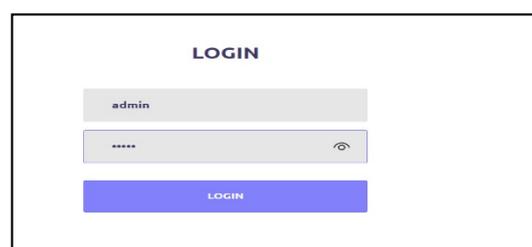
$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (1)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

$$\text{Akurasi} = \text{TP} / \text{Jumlah Data Test} \quad (3)$$

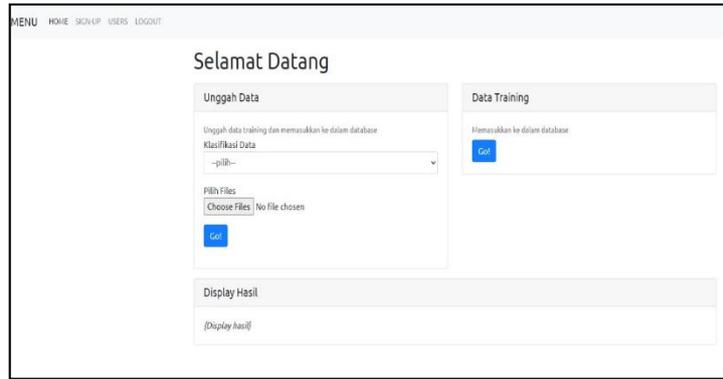
3. Hasil dan Pembahasan

3.1 Implementasi Antarmuka Sistem



Gambar 4. Tampilan Antarmuka Login

Pada gambar 4 di atas merupakan tampilan menu *login*. Pada menu tersebut disediakan kotak *input username* dan *password*. Untuk melakukan *login* pada menu tersebut, pastikan *user* telah memiliki akun untuk mengakses sistem pengenal suara kulkul.



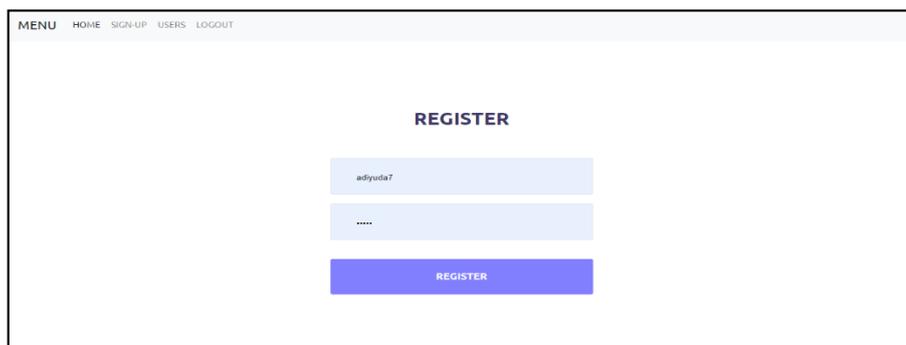
Gambar 5. Tampilan Antarmuka Menu Utama Admin

Gambar 5. merupakan tampilan menu utama dari sistem pengenalan suara kulkul. Pada menu utama terdapat fitur klasifikasi data untuk memilih kategori suara kulkul. Setelah memilih kategori, lalu pilih *file* dengan mengklik tombol “Choose File”. Setelah *file* suara kulkul di *upload*, terdapat 2 fitur yang bisa digunakan, yaitu *training* dan *classify*. Fitur *training* digunakan untuk sistem dapat menganalisa file yang diupload dan akan ditampilkan hasil *training dataset* suara kulkul pada kotak “Display Hasil”. Fitur *classify* digunakan untuk melakukan klasifikasi terhadap *file* yang dipilih. Hasil tersebut berupa kategori suara kulkul.



Gambar 6. Tampilan Antarmuka Kelola Akun

Gambar 6 diatas merupakan tampilan dari Kelola akun. Pada menu tersebut ditampilkan *username* dan *password*. Untuk melakukan perubahan klik tombol “Del” untuk menghapus akun.



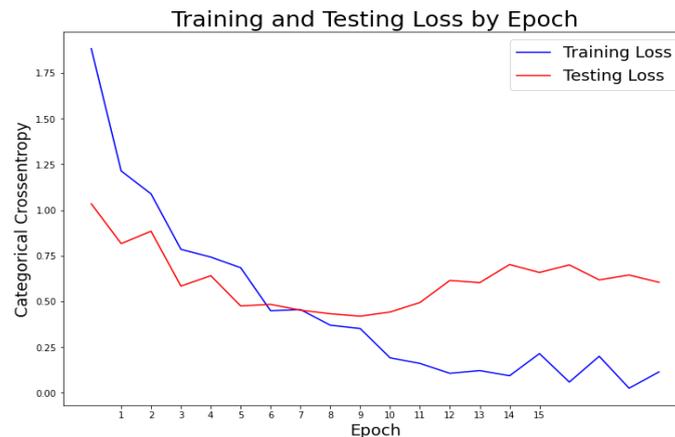
Gambar 7. Tampilan Antarmuka Daftar Akun

Gambar 7 merupakan tampilan daftar akun. Pada menu tersebut disediakan kotak untuk memasukkan *username* dan *password*. Setelah memasukkan *username* dan *password*, *user*

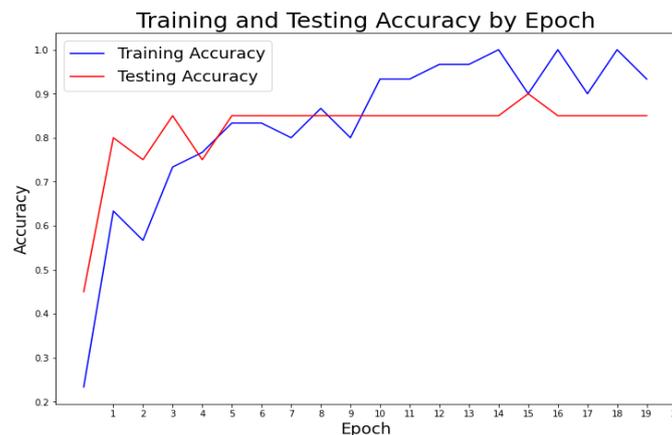
dapat melakukan klik pada tombol “*Register*”.

3.2 Klasifikasi

Pada bagian ini berisikan tentang hasil dari proses klasifikasi pengenalan suara kukul. Terdapat 5 kategori suara kukul, yaitu kukul bulus, kukul kemalingan, kukul kematian, kukul orang mengamuk, dan kukul pura. Jumlah data yang digunakan adalah 50 dengan masing – masing kategori berjumlah 10. Untuk data *train* digunakan 6 file setiap kelas dan data *test* digunakan 4 file setiap kelas. Proses klasifikasi terdiri dari *training* dan *testing*. Metode yang digunakan untuk proses klasifikasi adalah *Convolutional Neural Network (CNN)* dengan *epoch* yang digunakan adalah 20. Hasil dari proses klasifikasi menggunakan CNN dapat diperhatikan pada gambar berikut.



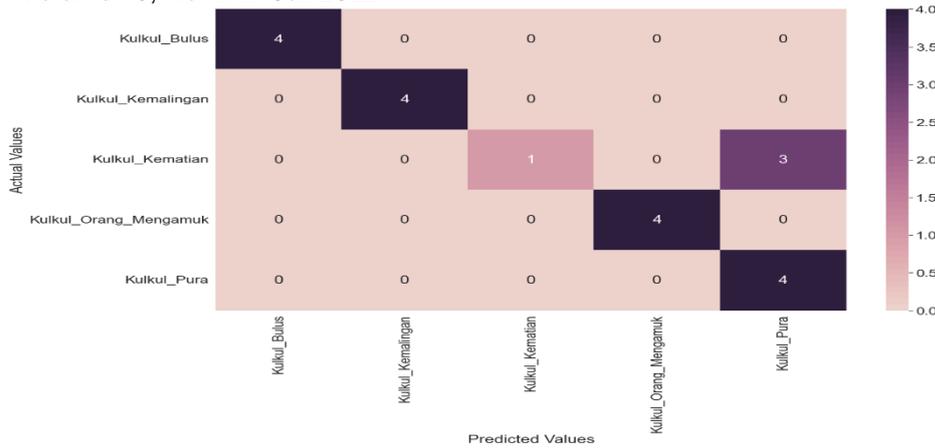
Gambar 8. Grafik *Training and Testing Loss*



Gambar 9. Grafik *Training and Testing Accuracy*

Berdasarkan grafik pada gambar 8, hasil dari *training loss* adalah 11,48% dan hasil *testing loss* adalah 60,51%. Pada gambar 9, hasil dari *training accuracy* adalah 93,33% dan hasil *testing accuracy* adalah 85%.

Pengujian model CNN dilakukan menggunakan *confusion matrix*. Hasil dari *confusion matrix* berisikan nilai *precision*, *recall*, dan *accuracy*. *Precision* dan *recall* dihitung pada masing – masing kelasnya. Sedangkan untuk *accuracy* dihitung berdasarkan jumlah data *test* yang benar dari keseluruhan data *test*. Berikut merupakan hasil pengujian model CNN dengan *confusion matrix*.



Gambar 10. Confusion Matrix

Berdasarkan *confusion matrix* pada gambar 10, diperoleh data yang termasuk *True Positive* (TP) atau jumlah data benar adalah kukul bulus berjumlah 4, kukul kemalingan berjumlah 4, kukul kematian berjumlah 1, kukul orang mengamuk berjumlah 4, dan kukul pura berjumlah 4. Data test yang digunakan adalah 20. Hasil *accuracy* dapat dihitung sebagai berikut.

$$\begin{aligned}
 \text{Accuracy} &= \text{jumlah data benar (TP)} / \text{jumlah data test} \\
 &= (4 + 4 + 1 + 4 + 4) / 20 \\
 &= 17 / 20 \\
 &= 85\%
 \end{aligned}$$

Tabel 2. Hasil *Precision* dan *Recall* dari Pengujian Model CNN

Kategori Suara Kukul	Precision	Recall
Kukul Bulus	100%	100%
Kukul Kemalingan	100%	100%
Kukul Kematian	100%	33,33%
Kukul Orang Mengamuk	100%	100%
Kukul Pura	57,14%	100%

Pada tabel 2, terdapat hasil *precision* dan *recall* dari masing – masing kategori suara kukul. Hasil dari rata – rata *precision* adalah 91,42% dan rata – rata *recall* adalah 86,6%.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan sebagai berikut.

1. Untuk membangun sistem pengenalan suara kukul berbasis web diperlukan data suara kukul, kebutuhan fungsional dan non – fungsional, program untuk melakukan ekstraksi ciri, program untuk melakukan proses klasifikasi menggunakan metode *Convolutional Neural Network* (CNN), dan program untuk tampilan antarmuka yang berbasis *web*.
2. Berdasarkan hasil klasifikasi suara kukul menggunakan metode CNN, diperoleh hasil *accuracy* 85% dan tergolong baik dalam mengenali suara kukul.

DAFTAR PUSTAKA

- [1] Heriyanto, H., Hartati, S., & Putra, A. E. (2018). EKSTRAKSI CIRI MEL FREQUENCY CEPSTRAL COEFFICIENT (MFCC) DAN RERATA COEFFICIENT UNTUK PENGECEKAN BACAAN AL-QUR'AN. *Telematika*, 15(2).
- [2] Jaya, H., Sabran, Idris, M. M., Djawad, Y. A., Iham, & Ahmar, A. S. (2018). Kecerdasan Buatan.
- [3] Santoso, A., & Ariyanto, G. (2018). *Implementasi Deep Learning Berbasis Keras Untuk Pengenalan Wajah*, 18, `15–21.
- [4] Thohari, A., & Hertantyo, G. B. (2018). *Implementasi Convolutional Neural Network untuk Klasifikasi Pembalap MotoGP Berbasis GPU. Proceedings on Conference on Electrical Engineering, Telematics, Industrial Technology, and Creative Media*, 50–55.

- [5] Candra, D. P. (2021). *Klasifikasi Suara Dengan Ekstraksi Ciri Mel Frequency Cepstral Coefficients Menggunakan Machine Learning*. JAKARTA.