e-ISSN: 2581-0545 - https://journal.itera.ac.id/index.php/jsat/

Klasifikasi Suara Burung Endemik Sumatera Menggunakan Convolutional Neural Network dan Variasi Teknik Ekstraksi Fitur

Evan Aryaputra 1 [a], Silvina Rizqy Nur Auliya 2 [b], Lion Abdi Marga 3 [c], Ditta Winanda Putri 4 [d], Jihan Putri Yani 5 [e], Ardika Satria M.Si [f], Ade Lailani M.Si [f]

[a,b,c,d,e,f,g] *Program Studi Sains Data, Institut Teknologi Sumatera*

* *E-mail: evan.121450102@student.itera.ac.id*

Abstract: *This study aims to identify endemic bird species on Sumatra Island through the analysis of call and song sounds using technology. By applying Mel-Frequency Cepstral Coefficients (MFCC) and Mel-Spectrogram feature extraction techniques, this research develops a Convolutional Neural Network (CNN) model to classify sounds from the bird species Pitta sordida, Dryocopus javensis, Pnoepyga pusilla, Anthipes solitaris, Caprimulgus macrurus, and Buceros rhinoceros. The CNN architecture used consists of four convolutional layers with filters of sizes 32, 64, 128, 256, and 512, as well as batch normalization to improve training stability. The model is processed with input data in the form of bird sound frequency spectrograms. Evaluation shows that the model with MFCC features achieves an accuracy of 72% on validation data and 62% on test data, while with Mel-Spectrogram, the model reaches 86% accuracy on both validation and test data. The results of the confusion matrix and F1-score analysis indicate that the CNN model can identify bird sounds well, although there are some errors for specific species. This research is expected to contribute to the conservation of Indonesia's biodiversity.*

Keywords: *Biodiversity Conservation, Classification, Convolutional Neural Network, Mel-Frequency Cepstral Coefficients, Mel-Spectrogram, Sumatra Endemic Birds*

Abstrak: Penelitian ini bertujuan untuk mengidentifikasi spesies burung endemik di Pulau Sumatera melalui analisis suara panggilan dan kicauan menggunakan teknologi. Dengan menerapkan teknik ekstraksi fitur *Mel-Frequency Cepstral Coefficients* dan *Mel-Spectrogram,* penelitian ini mengembangkan model *Convolutional Neural Network* untuk mengklasifikasikan suara dari spesies burung *Pitta sordida, Dryocopus javensis, Pnoepyga pusilla, Anthipes solitaris, Caprimulgus macrurus, dan Buceros rhinoceros*. Arsitektur CNN yang digunakan terdiri dari empat lapisan konvolusi dengan filter berukuran 32, 64, 128, 256 dan 512, serta batch normalization untuk meningkatkan stabilitas pelatihan. Model ini diproses dengan data input berupa spektrum frekuensi suara burung. Evaluasi menunjukkan model dengan fitur MFCC mencapai akurasi 72% pada data validasi dan 62% pada data uji, sementara dengan Mel-Spectrogram, model mencapai akurasi 86% pada data validasi dan uji. Hasil analisis confusion matrix dan F1-*score* menunjukkan model CNN mampu mengidentifikasi suara burung dengan baik, meskipun ada beberapa kesalahan pada spesies tertentu. Penelitian ini diharapkan dapat berkontribusi dalam upaya pelestarian keanekaragaman hayati Indonesia.

Kata Kunci : *Burung Endemik Sumatera, Convolutional Neural Network, Mel-Frequency Cepstral Coefficients, Mel-Spectrogram, Klasifikasi, Keanekaragaman Hayati*

Original Article *Journal of Science and Applicative Technology*

Pendahuluan

Latar Belakang

Pulau Sumatera dikenal sebagai salah satu wilayah dengan keanekaragaman hayati yang sangat tinggi, termasuk populasi burung yang beragam. Namun, iklim global dan aktivitas manusia yang semakin meningkat telah menyebabkan berbagai perubahan dalam ekosistem. Hal ini berdampak langsung pada perilaku dan populasi burung di wilayah tersebut. Burung sering kali dianggap sebagai indikator lingkungan yang sensitif karena mereka dapat bereaksi cepat terhadap perubahan ekosistem. Oleh karena itu, analisis suara burung menjadi salah satu cara penting untuk memahami interaksi antara organisme dan lingkungannya di Sumatera.

Selain menjadi indikator perubahan lingkungan, penelitian terbaru menunjukkan bahwa populasi burung di berbagai wilayah termasuk Sumatera, mengalami penurunan yang signifikan. Penurunan ini menjadi ancaman bagi keanekaragaman hayati dan memerlukan perhatian khusus untuk melindungi populasi burung yang ada. Identifikasi spesies burung secara akurat menjadi penting, tidak hanya untuk tujuan konservasi tetapi juga untuk memahami dinamika ekosistem. Metode tradisional untuk mengidentifikasi burung, baik melalui pengamatan langsung maupun analisis audio oleh para ahli, memiliki keterbatasan dalam hal waktu dan efisiensi[1]. Dengan meningkatnya jumlah data yang tersedia, pendekatan manual menjadi tidak praktis dan membutuhkan sumber daya yang mahal. Oleh karena itu, diperlukan pendekatan berbasis teknologi untuk mengatasi tantangan ini[2].

Penelitian ini bertujuan untuk mengembangkan metode berbasis deep learning yang mampu mengidentifikasi jenis burung melalui analisis kicauan mereka. Dengan menggunakan Convolutional Neural Network (CNN) yang memanfaatkan fitur Mel-Spectrogram dan Mel-Frequency Cepstral Coefficients (MFCC). Model ini dipilih karena keunggulan mengubah data audio mentah (waveform) menjadi representasi visual yang dapat dimanfaatkan oleh CNN[3]. Mel-spectrogram mentransformasikan suara ke domain waktu frekuensi berdasarkan skala Mel yang lebih sesuai dengan persepsi pendengaran manusia, sedangkan MFCC merepresentasikan suara yang menangkap informasi spektral dalam bentuk fitur kompak berdasarkan filter skala Mel[4].

Pada penelitian sebelumnya yang telah dilakukan oleh Elsa Ferreira Gomez, dkk (2021) dalam klasifikasi otomatis suara burung menggunakan MFCC dan fitur mel-spectrogram dengan deep learning, menunjukkan hasil akurasi yang lumayan baik pada data latih yaitu 70,74% jika dibandingkan dengan penelitian yang dilakukan oleh Hiatt (2019), yang mana hanya mendapatkan hasil akurasi sebesar 20,44%.

Berdasarkan penelitian-penelitian sebelumnya, metode CNN dengan fitur Mel-Spectrogram dan MFCC terbukti efektif dalam menghasilkan nilai akurasi yang lebih tinggi, khususnya untuk data suara burung. Oleh karena itu, penelitian ini difokuskan pada pengolahan suara burung di pulau Sumatera tujuan utamanya yaitu untuk mengeksplorasi keandalan pendekatan ini dalam mendukung upaya konservasi burung di wilayah tersebut.

Metode

Data yang digunakan dalam penelitian ini berasal dari data sekunder yang bersumber dari situs Xeno-canto. Data yang digunakan adalah rekaman audio burung khas wilayah Sumatera, yaitu *Pitta sordida, Dryocopus javensis, Pnoepyga pusilla, Anthipes solitaris, Caprimulgus macrurus, dan Buceros rhinoceros.*

Tabel 1. Data Spesies Burung Sumatera

Spesies
Pitta sordida
Dryocopus javensis
Pnoepyga pusilla
Anthipes solitaris
Caprimulgus macrurus
Buceros rhinoceros

Penelitian ini menggunakan metode C*onvolutional Neural Network* (CNN) untuk menganalisis dan mengklasifikasikan suara burung berdasarkan fitur audio yang diekstraksi dari rekaman. Teknik ekstraksi fitur seperti *Mel-Frequency Cepstral Coefficients* (MFCC) dan *Mel-Spectrogram* digunakan untuk mengubah data audio menjadi representasi visual. Representasi ini mempermudah model CNN dalam mengenali pola suara unik dari setiap spesies burung.

Model CNN digunakan karena kemampuannya menangkap pola non-linier yang kompleks dalam data audio. Komponen utama dalam model CNN meliputi lapisan konvolusi untuk

mendeteksi fitur penting, lapisan pooling untuk mereduksi dimensi data, dan lapisan fully connected untuk menghasilkan prediksi. Model ini dioptimalkan untuk mengenali pola spesifik dalam suara burung dan memberikan hasil klasifikasi yang akurat.

Tahapan dalam melakukan metode Convolutional Neural Network dengan Variasi Teknik Ekstraksi Fitur sebagai berikut:

1. Pengumpulan Data

Proses pengambilan data dimulai dengan mengakses halaman pencarian pada platform Xeno-canto. Penelitian ini memfokuskan pada wilayah geografi Pulau Sumatera dengan menerapkan filter pencarian yang spesifik. URL yang digunakan untuk mencapai halaman pencarian pada website tersebut adalah https://www.xeno-canto.org/explore?query=Sumatra, yang memungkinkan peneliti untuk mengakses koleksi suara burung yang relevan dengan lokasi yang ditentukan. Setelah akses ke halaman utama berhasil, pemrosesan lebih lanjut dilakukan dengan menggunakan pustaka *BeautifulSoup*, yang berfungsi untuk menguraikan isi dokumen HTML. Melalui metode ini, peneliti dapat mencari dan mengidentifikasi elemen-elemen penting dari halaman, seperti nama spesies burung, lokasi pengambilan rekaman, tanggal rekaman, dan tautan unduhan untuk file audio yang bersangkutan.

Tahap selanjutnya adalah ekstraksi metadata dari elemen HTML yang relevan. Informasi yang diekstrak mencakup nama spesies *(scientific name)*, lokasi rekaman *(location)*, tanggal rekaman *(date)*, dan tautan untuk mengunduh file audio. Proses ekstraksi ini bertujuan untuk memastikan bahwa data yang diambil tidak hanya akurat tetapi juga representatif dari kekayaan biodiversitas burung di Pulau Sumatera. Setelah metadata diperoleh, tautan yang telah diekstraksi digunakan untuk mengunduh rekaman suara dalam format .wav. File audio ini disimpan dalam struktur direktori yang terorganisir secara lokal, sehingga memudahkan proses pengelolaan dan aksesibilitas data suara burung yang telah diunduh. Sebagai langkah akhir, metadata yang dikumpulkan selama proses scraping disimpan dalam format file CSV. Format ini dipilih karena kemampuannya untuk menyimpan data dalam struktur yang terdefinisi dengan baik.

2. Preprocessing

Langkah pertama dalam analisis dataset suara burung ini adalah melakukan mounting Google Drive ke dalam Google Colab. Proses ini dilakukan dengan menjalankan perintah drive.mount('/content/drive'), yang memungkinkan akses langsung ke folder di Google Drive. Dataset suara burung disini terstruktur dalam sejumlah folder, masing-masing mewakili spesies burung tertentu, di mana setiap folder berisi kumpulan file audio berformat .wav. Persiapan yang benar di tahap ini sangat penting untuk memastikan bahwa data dapat diakses dan diproses lebih lanjut dalam tahapan-tahapan berikutnya. Setelah dataset disiapkan, langkah berikutnya adalah melakukan ekstraksi data dari folder yang ada. Di tahap ini, setiap subfolder yang mewakili spesies burung diidentifikasi, dan file-file audio dengan format .wav dikumpulkan. Proses identifikasi dan pengumpulan file dilakukan menggunakan metode *os.walk*, yang menavigasi melalui direktori dengan sistematis dan mengumpulkan daftar file audio untuk setiap spesies burung.

Setelah file audio terkumpul, tahapan selanjutnya adalah melakukan segmentasi. Setiap file audio .wav dipotong menjadi segmen-segmen yang lebih kecil berdasarkan durasi yang telah ditentukan. Proses segmentasi dilakukan dengan memanfaatkan pustaka *librosa*, di mana file audio dimuat dan segmen ditentukan melalui penghitungan indeks start dan end. Segmen yang dihasilkan kemudian disaring berdasarkan panjangnya, memastikan bahwa hanya segmen yang memenuhi kriteria panjang minimum yang akan diproses lebih lanjut. Segmentasi membantu dalam menyediakan ukuran konsisten dari file yang akan digunakan dalam pembelajaran mesin.

3. Mel-Spectrogram (MelSpec)

Setelah segmentasi, proses selanjutnya adalah preprocessing Mel-Spectrogram. Mel-Spectrogram merupakan representasi dari energi spektral sinyal audio yang dihitung dengan menggunakan Fast Fourier Transform (FFT)[5]. Dalam tahapan ini, spektrum audio diubah menjadi skala Mel, yang lebih sesuai dengan cara manusia mendengar. Selanjutnya, energi dalam skala Mel divisualisasikan sebagai gambar dengan bantuan librosa.display.specshow, dan gambar tersebut disimpan dalam format .png. Mel-Spectrogram merupakan representasi dari energi spektral sinyal audio yang dihitung dengan menggunakan Fast Fourier Transform (FFT). FFT adalah algoritma cepat dari Discrete Fourier Transform (DFT) yang berguna untuk mengubah setiap frame dengan sampel

N dari domain waktu menjadi domain frekuensi[7]. Dalam tahapan ini, spektrum audio diubah menjadi skala Mel, yang lebih sesuai dengan cara manusia mendengar. Hasilnya adalah spektrum frekuensi $S(f)$, yang menunjukkan kekuatan atau amplitudo sinyal pada frekuensi tertentu. Persamaan untuk *mel- spectrogram* dapat diperoleh dari persamaan berikut :

Persamaan (1) fast fourier transform (FFT) pada sinyal audio.

$$S(f) = FFT(x(t)) \quad (1)$$

Keterangan :

$x(t)$: sinyal audio dalam domain waktu

$S(f)$: spektrum frekuensi hasil dari transformasi fourier

Persaman (2) filter bank mel yang terdiri dari sejumlah filter yang membentuk spektrum pada skala mel.

$$f_{mel} = \left\{2595.\ log_{10}\ (1 + \frac{Fhz}{700}\right\} \quad (2)$$

Keterangan :

f_{mel} : frekuensi dalam skala mel

FHz : frekuensi dalam skala linier (Hz)

Persamaan (3) penerapan filter bank mel ke spektrum frekuensi.

$$M_i = \sum_j H_i\,(f_j)\,S(f_j) \quad (3)$$

Keterangan :

$H_i(f_j)$: respon dari filter mel ke-j pada frekuensi f_j

$S(f)$: spektrum frekuensi hasil dari transformasi fourier

M_i : energi yang terkumpul pada mel ke-i

Persamaan (4) logaritma dari spektrum mel.

$$log(M_i) = log\left(\sum_j Hi\,(fj)\,S(fj)\right) \quad (4)$$

Keterangan :

$x(t)$: sinyal audio dalam domain waktu

$S(f)$: spektrum frekuensi hasil dari transformasi fourier

Persamaan (5) interpretasi mel spectrogram secara keseluruhan.

$$M_{mel}(t,\ f) = log\left(\sum_i Hi(f)\,S(j)\right) \quad (5)$$

Keterangan :

$M_{mel}(t,\ f)$: nilai pada waktu t dan frekuensi mel f

$H_i(f)$: filter mel untuk i pada frekuensi f

$S(f)$: spektrum frekuensi dari waktu t

Setelah mendapatkan spektrum frekuensi, langkah berikutnya yaitu mengubahnya ke dalam skala mel, dirumuskan sebagai berikut :

$$f_{mel} = \left\{2595.\ log_{10}\ (1 + \frac{Fhz}{700}\ , jika\ Fhz > 1000\right\} \quad (6)$$

4. Mel Frequency Cepstral Coefficient (MFCC)

Selanjutnya, fitur MFCC (Mel-Frequency Cepstral Coefficients) diekstraksi dari sinyal audio untuk menghasilkan representasi yang lebih ringkas dari data audio[6]. Fitur MFCC dihitung menggunakan metode *librosa.feature.mfcc*, dengan berbagai parameter, seperti jumlah koefisien yang diinginkan (misalnya 40) dan ukuran hop. Visualisasi dari fitur MFCC juga dilakukan dengan *librosa.display.specshow*, dan hasil visualisasi tersebut disimpan sebagai gambar berformat .png. Skala frekuensi mel adalah frekuensi rendah yang linier di bawah 1000 Hz dan frekuensi tinggi yang logaritmik di atas 1000 Hz[8]. Persamaan berikut menunjukkan hubungan skala mel dengan frekuensi dalam Hz

$$f_{mel} = \left\{2595.\ log_{10}\ (1 + \frac{Fhz}{700}\ , jika\ Fhz > 1000\right\} \quad (7)$$

Keterangan :

f : adalah frekuensi dalam satuan mel.

Fhz : adalah frekuensi dalam hertz (satuan frekuensi).

700 : adalah konstanta yang digunakan dalam skala mel untuk menyesuaikan frekuensi agar sesuai dengan persepsi pendengaran manusia.

log_{10} : adalah fungsi logaritma dengan basis 10.

5. Convolutional Neural Network

Convolutional Neural Network (CNN) adalah jenis jaringan syaraf tiruan yang digunakan untuk memproses data, dengan tujuan mendeteksi dan mengenali objek [9]. CNN dirancang untuk bekerja dengan data dalam bentuk *array* dan terinspirasi oleh cara kerja korteks visual otak manusia. CNN digunakan dalam pembelajaran mendalam di bidang visi komputer untuk mendeteksi dan mengklasifikasikan fitur gambar.

Convolutional Layer adalah sebuah tahapan ketika seluruh data menyentuh lapisan *convolutiona*l yang sudah melalui proses konvolusi lalu dilakukan filter yang akhirnya akan menghasilkan *activation map* [10]. Pada model CNN penetapan *stride* atau *step width* biasanya sebesar 1 dengan *zero padding* yang dapat dilihat pada persamaan (8).

$$P = \frac{(F-1)}{2} \qquad (8)$$

Keterangan:

P : Ukuran padding

F : Tingkat ukuran filter

Untuk ukuran pada input layer dapat dihitung berdasarkan lebar, tinggi dan juga jumlah kanal yang dapat dilihat pada persamaan (9)

$$w_1 x h_1 x d_1 \qquad (9)$$

Keterangan:

w_1 : Ukuran *padding*

h_1 : Tingkat ukuran filter

d_1 : Tingkat ukuran filter

Untuk *output* dapat dihitung dengan menggunakan persamaan (10)

$$w_2 x h_2 x d_2 \qquad (10)$$

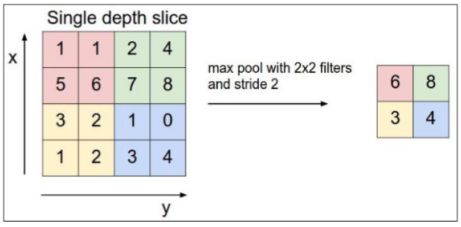
w_2, h_2, d_2 dapat dihitung menggunakan persamaan (11), persamaan (12), dan persamaan (13).

$$w_2 = \frac{(w_1 - F + 2P)}{S} + 1 \qquad (11)$$

$$h_2 = \frac{(h_1 - F + 2P)}{S} + 1 \qquad (12)$$

$$d_2 = K \qquad (13)$$

Pooling Layer pada CNN berfungsi untuk menjaga volume data saat proses konvolusi dengan mengurangi jumlah sampel. Lapisan ini ditempatkan secara teratur di antara lapisan konvolusi. Salah satu bentuk *Pooling layer* yang paling sering digunakan adalah 2x2 [11]. Proses *pooling* yang digunakan yaitu *max pooling* yang dapat dilihat pada **Gambar 1.**



Gambar 1. Proses *Pooling* [12]

Evaluasi Model

Untuk mengevaluasi model digunakan metrik akurasi, presisi, recall dan F1-Score [13]. Akurasi mengukur sejauh mana model mampu mengklasifikasikan data dengan benar secara keseluruhan [14]. Presisi mengukur ketepatan prediksi positif yang dihasilkan oleh model, yaitu sejauh mana prediksi positif tersebut sesuai dengan kelas sebenarnya [15]. Recall mengukur kemampuan model dalam mendeteksi kelas positif dengan benar. F1-*Score* adalah metrik yang menggabungkan *precision* dan *recall.* Persamaan metrik tersebut dapat dilihat pada persamaan (14), persamaan (15), persamaan (16), dan persamaan (17).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (14)$$

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

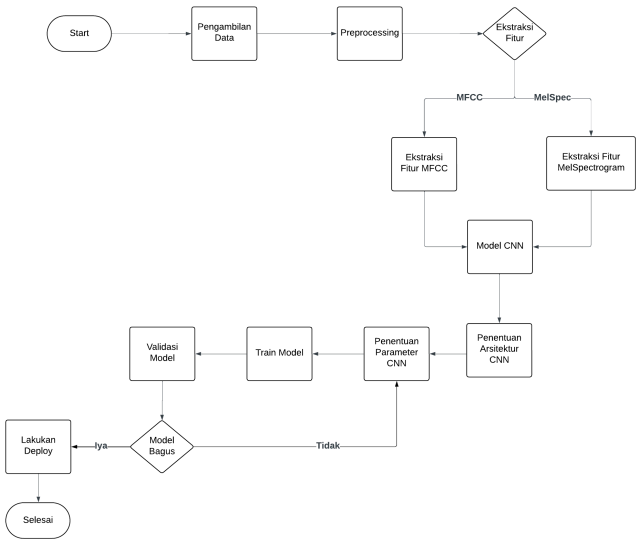
$$Recall = \frac{TP}{TP + FN} \quad (16)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + recall} \quad (17)$$

Keterangan:

TP : True positive
TN : True negative
FP : False positive
FN : False Negative

Flowchart



Gambar 2. *Flowchart* Penelitian

Pada **Gambar 2.** menunjukkan proses klasifikasi suara burung dengan CNN dimulai dengan pengumpulan data dari Xeno-Canto, meliputi spesies seperti *Pitta sordida* dan *Dryocopus javensis*. Rekaman suara diproses melalui tahap preprocessing, seperti normalisasi dan penghapusan noise. Ekstraksi fitur dilakukan dengan *Mel Frequency Cepstral Coefficients* (MFCC) untuk menangkap karakteristik suara berdasarkan persepsi manusia, diikuti oleh *Mel-Spectrogram* (MelSpec) untuk visualisasi distribusi energi pada spektrum frekuensi. Lalu dilakukan pembentukan model dan evaluasi model. Jika model yang dihasilkan sudah bagus maka akan dilakukan proses deploy.

Hasil dan Pembahasan

1. Proses Pembagian Data

Proses pembagian data dilakukan untuk memastikan setiap dataset memiliki distribusi yang seimbang, baik untuk pelatihan (training), validasi (validation), maupun pengujian (testing). Data yang digunakan dalam penelitian ini terdiri dari file audio **.wav** dari berbagai jenis burung. Proses ini melibatkan pembatasan jumlah data per kelas, dan pembagian data berdasarkan proporsi tertentu. Jumlah pembagian data dapat dilihat pada **Tabel 2.**

Tabel 2. Pembagian Data

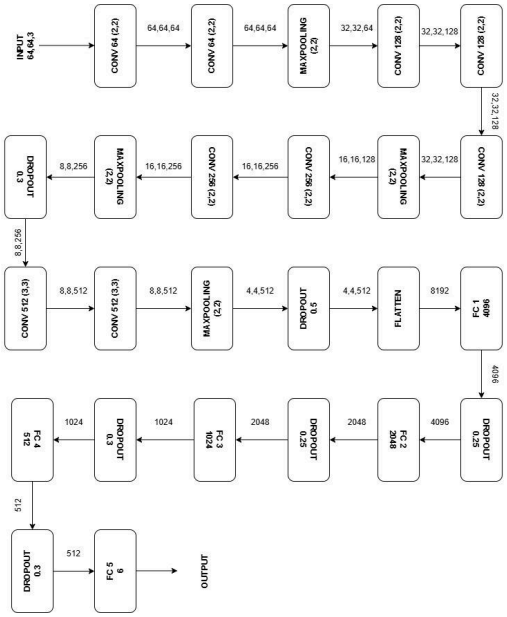
Pembagian	Proporsi
Training	70%
Validation	15%
Testing	15%

Pada **Tabel 2.** menunjukkan pembagian data penelitian yang terdiri dari subset training, validation, dan testing. Data yang digunakan berupa file audio berformat .wav dari berbagai jenis burung, yang dibagi dengan proporsi 70% untuk training, 15% untuk validation dan 15% untuk testing.

2. Model CNN

Arsitektur model CNN yang dibuat dimulai dengan empat blok konvolusional, masing-masing terdiri dari dua lapisan konvolusi berturut-turut dengan filter yang semakin besar 64, 128, 256, 512 diikuti oleh operasi pooling Max Pooling untuk mengurangi dimensi dan mengekstraksi fitur dari input gambar. Setiap lapisan konvolusi menggunakan fungsi aktivasi ReLU untuk memperkenalkan non-linearitas, dan padding 'same' digunakan agar dimensi output tetap sama dengan input pada setiap lapisan. Setelah itu, model menerapkan dropout untuk mengurangi risiko *overfitting*. Setelah lapisan konvolusional, gambar yang telah diproses dipipihkan menggunakan layer *Flatten*, kemudian dilanjutkan dengan beberapa lapisan *fully connected Dense* dengan jumlah unit 4096, 2048, 1024, 512 untuk menangkap hubungan yang lebih kompleks antar fitur yang telah diekstraksi. Regularisasi L2 diterapkan di setiap layer fully connected untuk mencegah *overfitting*. Terakhir, model menghasilkan output berupa probabilitas untuk setiap kelas menggunakan fungsi aktivasi *softmax*, yang cocok untuk masalah klasifikasi multi kelas. Optimisasi dilakukan

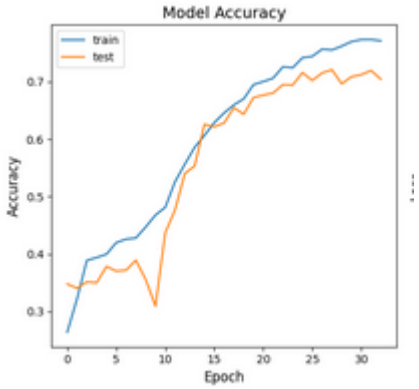
menggunakan Adam *optimizer* dengan *learning rate* yang sangat kecil, dan ada dua callback yang diterapkan, yaitu *Learning Rate Scheduler* untuk menurunkan *learning rate* setelah *epoch* tertentu dan *Early Stopping* untuk menghentikan pelatihan jika tidak ada perbaikan dalam beberapa *epoch* berturut-turut.



Gambar 5.Arsitektur CNN

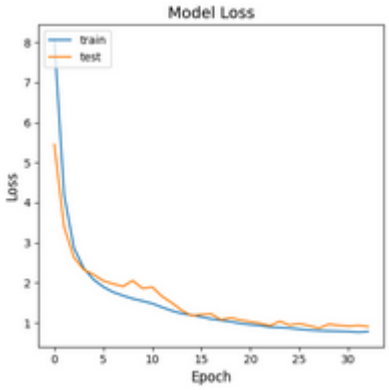
MFCC

Pada **Gambar 6.** menunjukkan hasil grafik akurasi model CNN yang menggunakan ekstraksi fitur MFCC untuk klasifikasi suara burung. Pada bagian kiri grafik akurasi, terlihat bahwa akurasi model meningkat secara signifikan pada iterasi awal dan terus stabil seiring berjalannya waktu, dengan fluktuasi kecil yang menunjukkan proses pembelajaran yang efektif. Pencapaian akurasi tertinggi tercatat pada epoch 30, yang mengindikasikan bahwa model berhasil mengenali pola audio dengan baik pada data pelatihan. Fluktuasi kecil pada grafik menunjukkan adanya tantangan dalam generalisasi data, meskipun secara keseluruhan akurasi model cukup tinggi.



Gambar 6. Grafik *Accuracy* MFCC

Pada **Gambar 7.** terlihat penurunan yang konsisten pada nilai loss seiring dengan bertambahnya epoch, menunjukkan bahwa model terus belajar dan meminimalkan kesalahan prediksi. Penurunan loss yang stabil mengindikasikan bahwa model mampu mengurangi perbedaan antara prediksi dan label aktual. Setelah beberapa epoch, loss mencapai titik stabil yang menunjukkan bahwa model sudah hampir mencapai konvergensi, dan tidak ada penurunan signifikan lagi. Grafik ini memberikan gambaran tentang kinerja model dalam mengoptimalkan akurasi dan mengurangi loss selama proses pelatihan, yang menunjukkan kemampuan model CNN dengan MFCC untuk mengklasifikasikan suara burung secara efektif.

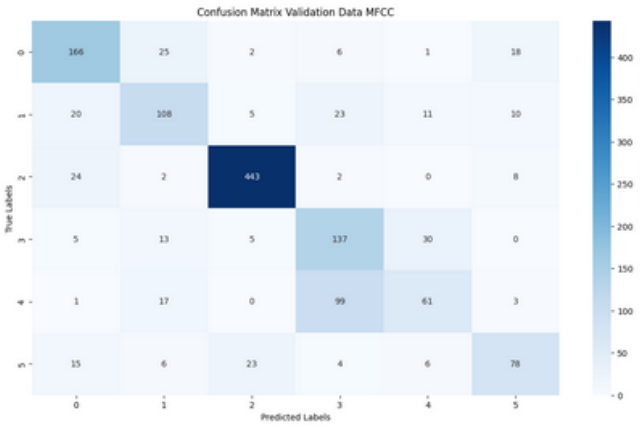


Gambar 7. Grafik *Loss* MFCC

Pada **Gambar 8.** menggambarkan jumlah prediksi benar di sepanjang diagonal utama, sedangkan kesalahan klasifikasi ditunjukkan oleh nilai di luar diagonal. Secara keseluruhan, model memiliki performa yang sangat baik pada kelas 2, di mana 443 sampel diprediksi dengan benar dan hanya sedikit kesalahan ke kelas lain. Ini menunjukkan bahwa fitur MFCC efektif dalam menangkap karakteristik suara burung di kelas tersebut. Di sisi lain, kelas 3 dan kelas 4 memiliki tingkat kebingungan yang cukup tinggi, terlihat

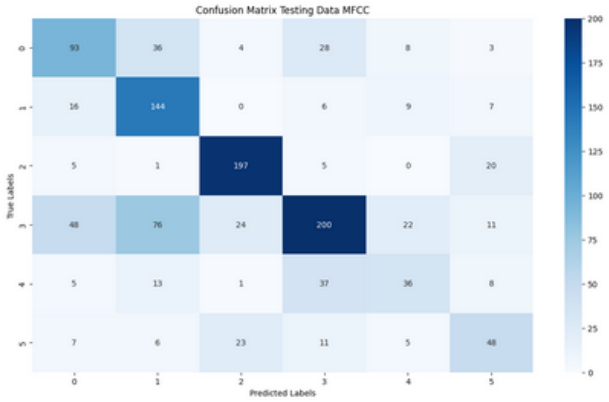
dari banyaknya kesalahan prediksi antara kedua kelas tersebut, seperti 61 sampel kelas 4 diprediksi menjadi kelas 3, dan 30 sampel kelas 3 salah diprediksi ke kelas 4. Hal ini bisa mengindikasikan bahwa karakteristik suara burung di kedua kelas tersebut cukup mirip, sehingga sulit dibedakan oleh model. Selain itu, kelas 5 juga menunjukkan hasil prediksi yang tersebar, dengan 23 sampel salah diprediksi ke kelas 3 dan hanya 78 sampel yang benar.

Dari bentuk matriks tersebut, performa model bisa dikatakan cukup baik namun belum optimal. Hal ini terlihat dari dominasi diagonal utama, yang menunjukkan prediksi benar, tetapi masih ada kesalahan yang cukup signifikan pada beberapa kelas tertentu.



Gambar 8. Matriks *Validation* MFCC

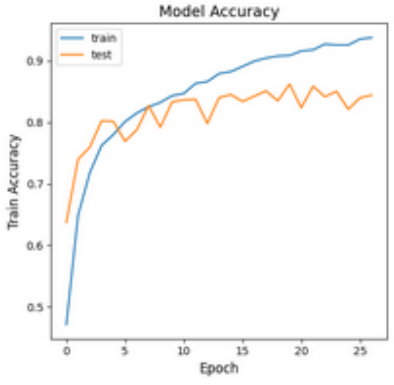
Berdasarkan *confusion matrix* pada **Gambar 9.** untuk data uji dengan fitur MFCC, performa model CNN menunjukkan hasil yang cukup baik, terutama pada kelas 2 dan kelas 3, dengan jumlah prediksi benar yang tinggi, yakni 197 sampel dan 200 sampel. Namun, terdapat beberapa kebingungan antar kelas, misalnya kelas 3 yang sering salah diprediksi ke kelas 1 dan kelas 2, serta kelas 0 yang memiliki 36 kesalahan ke kelas 1 dan 28 kesalahan ke kelas 4. Meskipun kelas 1 memiliki performa yang baik dengan 144 prediksi benar, model masih kesulitan mengenali suara di kelas 5, dengan hanya 48 sampel benar dan kesalahan signifikan ke kelas 2 (23 sampel). Kesalahan ini menunjukkan adanya tumpang tindih karakteristik suara antar kelas tertentu, yang memerlukan perbaikan seperti augmentasi data, penambahan fitur ekstraksi, atau fine-tuning model CNN. Secara keseluruhan, model sudah mampu mengenali pola suara burung dengan baik, tetapi masih perlu dioptimalkan untuk meminimalkan kebingungan antar kelas.



Gambar 9. Matriks *Testing* MFCC

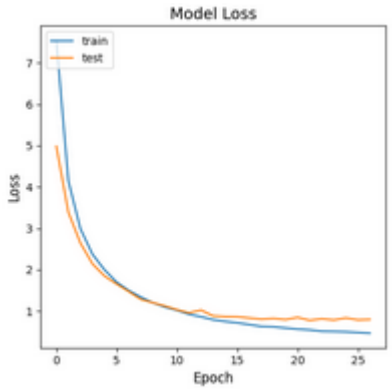
Mel-Spec

Pada **Gambar 10** menunjukkan hasil grafik akurasi model CNN yang menggunakan ekstraksi fitur *Mel-Spectrogram* untuk klasifikasi suara burung. Pada grafik ini, terlihat bahwa akurasi model meningkat dengan cepat pada iterasi awal, terutama pada data pelatihan, dan terus mengalami peningkatan yang stabil seiring bertambahnya epoch. Akurasi pelatihan mencapai titik tertinggi mendekati 0.95 pada akhir epoch, menunjukkan bahwa model dapat belajar pola dari data pelatihan dengan baik. Sementara itu, akurasi pengujian terlihat berfluktuasi di sekitar angka 0.8, menunjukkan bahwa model memiliki tantangan dalam generalisasi terhadap data uji. Meskipun fluktuasi terjadi, perbedaan antara akurasi pelatihan dan pengujian tetap cukup konsisten, yang mengindikasikan bahwa *overfitting* tidak terlalu signifikan. Secara keseluruhan, grafik ini menunjukkan bahwa model CNN dengan fitur Mel-Spec memiliki performa yang cukup baik.



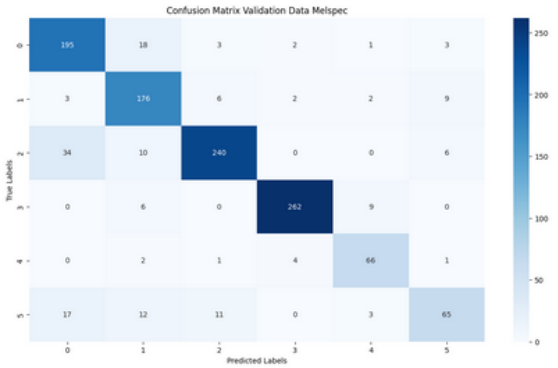
Gambar 7. Grafik *Accuracy* Mel-Spec

Pada **Gambar 11.** terlihat penurunan *loss* yang konsisten seiring dengan bertambahnya *epoch*, menunjukkan bahwa model CNN dengan *Mel-Spectrogram* sebagai fitur input berhasil belajar dan meminimalkan kesalahan prediksi secara bertahap. Pada epoch awal, terjadi penurunan *loss* yang cepat baik pada data *training* maupun *testing*, yang menandakan model sedang mengenali pola penting dari data. Setelah mencapai sekitar epoch ke-15, nilai *loss* mulai stabil di sekitar 0.5-1, menunjukkan bahwa model telah mencapai titik konvergensi. Stabilitas ini juga mengindikasikan bahwa model tidak mengalami *overfitting*, sehingga mampu melakukan generalisasi dengan baik pada data uji. Grafik ini menunjukkan bahwa model berhasil mengoptimalkan performa klasifikasi suara burung dengan mempelajari pola dari *Mel-Spectrogram* secara efektif.



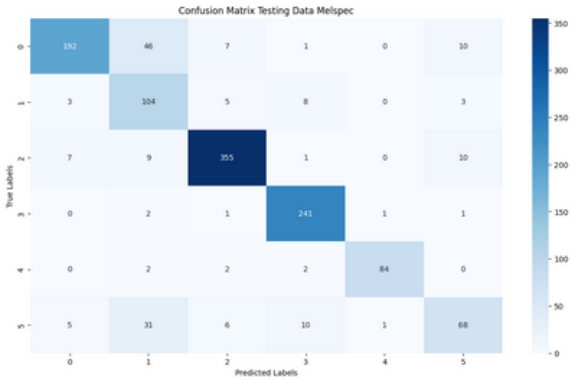
Gambar 11. Grafik *Loss* Mel-Spec

Pada **Gambar 12**, jumlah prediksi benar ditunjukkan oleh nilai di sepanjang diagonal utama, sedangkan kesalahan klasifikasi terlihat pada nilai di luar diagonal. Model memiliki performa terbaik pada kelas 3 dengan 262 sampel diprediksi benar dan hanya sedikit kesalahan, menunjukkan efektivitas fitur *Mel-Spectrogram* dalam menangkap karakteristik suara burung di kelas tersebut. Namun, terdapat kebingungan pada kelas 2, di mana 34 sampel salah diklasifikasikan ke kelas 0, yang mengindikasikan adanya kemiripan pola antara kedua kelas tersebut. Selain itu, kelas 5 memiliki hasil prediksi yang tersebar dengan hanya 65 sampel diprediksi benar, sementara 17 sampel salah diklasifikasikan ke kelas 0 dan 12 sampel ke kelas 1. Dominasi nilai pada diagonal utama menunjukkan bahwa model sudah bekerja cukup baik, tetapi kesalahan yang signifikan pada kelas 2 dan kelas 5 menunjukkan bahwa performa model masih bisa ditingkatkan.



Gambar 12. Matriks *Validation* Mel-Spec

Pada **Gambar 13.** model menunjukkan performa terbaik pada kelas 2 dengan 355 sampel diprediksi benar dan hanya sedikit kesalahan ke kelas lain, menandakan bahwa fitur Mel-Spectrogram berhasil mengenali karakteristik suara burung pada kelas tersebut. Di sisi lain, kelas 0 memiliki tingkat kebingungan yang cukup tinggi dengan 46 sampel salah diklasifikasikan ke kelas 1, yang mengindikasikan adanya kemiripan fitur antara kedua kelas. Sementara itu, kelas 5 memperlihatkan prediksi yang tersebar, di mana 31 sampel salah diklasifikasikan ke kelas 1 dan hanya 68 sampel diprediksi dengan benar. Kelas 4 memiliki kinerja yang cukup baik dengan 84 prediksi benar dan kesalahan minimal ke kelas lainnya. Secara keseluruhan, model mampu memprediksi sebagian besar kelas dengan baik, seperti terlihat dari dominasi nilai pada diagonal utama. Namun, masih terdapat beberapa kesalahan signifikan, terutama pada kelas 0 dan kelas 5, yang menunjukkan perlunya peningkatan lebih lanjut dalam pemisahan karakteristik antar kelas.



Gambar 13. Matriks *Testing* Mel-Spec

3. Evaluasi Model

Tabel 3. menunjukkan hasil evaluasi model CNN berdasarkan data validasi. Model memiliki performa terbaik pada kelas **2Capri**, dengan nilai F1-score sebesar 0.93, yang menunjukkan bahwa model sangat baik dalam mengenali pola pada kelas ini. Sebaliknya, kelas **4Anthi** memiliki performa terendah dengan nilai F1-score sebesar 0.42, disebabkan oleh rendahnya recall sebesar 0.34. Kelas lainnya, seperti **3Pnoep** dan **1Dryoc**, menunjukkan nilai F1-score yang lebih rendah, masing-masing 0.59 dan 0.62. Secara keseluruhan, akurasi model pada data validasi adalah 72%, dengan nilai macro average F1-score sebesar 0.65, menunjukkan performa yang cukup variatif antar kelas.

Tabel 3. Evaluasi Model MFCC Data *Validation*

	precision	recall	f1-score	support
0Pitta	0.72	0.76	0.74	218
1Dryoc	0.63	0.61	0.62	177
2Capri	0.93	0.92	0.93	479
3Pnoep	0.51	0.72	0.59	190
4Anthi	0.56	0.34	0.42	181
5Buer	0.67	0.59	0.63	132
accuracy			0.72	1377
macro avg	0.67	0.67	0.65	1377
weighted avg	0.72	0.72	0.72	1377

Berdasarkan laporan klasifikasi pada data pengujian **Tabel 4.** menghasilkan akurasi sebesar 62%. Kelas 2Capri menunjukkan performa terbaik dengan nilai F1-score 0.83, menunjukkan pengenalan pola yang baik pada kelas ini. Namun, kelas 4Anthi memiliki kinerja yang sangat rendah, dengan F1-score sebesar 0.40, disebabkan oleh rendahnya precision dan recall. Kelas 1Dryoc memiliki F1-score 0.63 dengan recall yang tinggi, tetapi precision yang rendah, sedangkan kelas 3Pnoep menunjukkan F1-score 0.60, mencerminkan ketidakseimbangan antara precision dan recall. Dengan macro average F1-score sebesar 0.58 dan weighted average F1-score sebesar 0.61, model menunjukkan variasi kinerja antar kelas. Perbaikan diperlukan, terutama untuk kelas dengan performa rendah, agar model dapat mengklasifikasikan data dengan lebih baik.

Tabel 4. Evaluasi Model MFCC Data *Test*

	precision	recall	f1-score	support
0Pitta	0.53	0.54	0.54	172
1Dryoc	0.52	0.79	0.63	182
2Capri	0.79	0.86	0.83	228
3Pnoep	0.70	0.52	0.60	381
4Anthi	0.45	0.36	0.40	100
5Buer	0.49	0.48	0.49	100
accuracy			0.62	1163
macro avg	0.58	0.59	0.58	1163
weighted avg	0.63	0.62	0.61	1163

Tabel 5. menunjukkan hasil evaluasi model CNN berdasarkan data validasi. Dari laporan klasifikasi, terlihat bahwa model memiliki performa terbaik pada kelas 3Pnoep dengan nilai F1-score sebesar 0.96, diikuti oleh kelas 4Anthi dengan F1-score sebesar 0.85. Sebaliknya, performa terendah ditemukan pada kelas 5Bucer dengan F1-score sebesar 0.68, terutama disebabkan oleh rendahnya nilai recall sebesar 0.60. Secara keseluruhan, akurasi model pada data validasi mencapai 86%, dengan macro average F1-score sebesar 0.84, yang menunjukkan performa model yang cukup stabil di berbagai kelas. Akan tetapi, diperlukan peningkatan pada kelas dengan performa rendah untuk memperbaiki kemampuan generalisasi model.

Tabel 5. Evaluasi Model Mel-Spec Data *Validation*

	precision	recall	f1-score	support
0Pitta	0.78	0.88	0.83	222
1Dryoc	0.79	0.89	0.83	198
2Capri	0.92	0.83	0.87	290
3Pnoep	0.97	0.95	0.96	277
4Anthi	0.81	0.89	0.85	74
5Buer	0.77	0.60	0.68	108
accuracy			0.86	1169
macro avg	0.84	0.84	0.84	1169
weighted avg	0.86	0.86	0.86	1169

Tabel 6. Evaluasi Model Mel-Spec Data *Test*

	precision	recall	f1-score	support
0Pitta	0.93	0.75	0.83	256
1Dryoc	0.54	0.85	0.66	123
2Capri	0.94	0.93	0.94	382
3Pnoep	0.92	0.98	0.95	246
4Anthi	0.98	0.93	0.95	90
5Buer	0.74	0.56	0.64	121
accuracy			0.86	1218
macro avg	0.84	0.83	0.83	1218
weighted avg	0.88	0.86	0.86	1218

Tabel 6. menunjukkan hasil evaluasi model CNN berdasarkan data pengujian. Model menunjukkan performa terbaik pada kelas 3Pnoep dan 4Anthi, masing-masing dengan nilai F1-score sebesar 0.95, mencerminkan kemampuan model dalam mengenali pola data uji untuk kelas-kelas ini dengan baik. Sebaliknya, performa terendah ditemukan pada kelas 1Dryoc dengan F1-score sebesar 0.66 dan kelas 5Bucer dengan F1-score sebesar 0.64, yang disebabkan oleh precision dan recall yang rendah. Secara keseluruhan, akurasi model pada data pengujian mencapai 86%, dengan weighted average F1-score sebesar 0.86, menunjukkan kinerja yang cukup baik. Namun, diperlukan perhatian lebih pada kelas dengan jumlah data kecil untuk meningkatkan kemampuan deteksi pada kelas-kelas tersebut.

Kesimpulan

Model yang dievaluasi menggunakan fitur MFCC dan Mel-Spec menunjukkan hasil yang bervariasi dalam hal akurasi dan performa antar kelas. Pada evaluasi menggunakan MFCC, model mencapai akurasi 72% pada data validasi dengan macro average F1-score sebesar 0.65, yang menunjukkan adanya variasi yang signifikan dalam kemampuan model untuk mengenali pola suara dari masing-masing kelas. Kelas 2Capri menunjukkan performa terbaik dengan F1-score 0.93, sementara kelas 4Anthi memiliki F1-score terendah sebesar 0.42. Sementara itu, pada evaluasi menggunakan Mel-Spec, model menunjukkan peningkatan yang signifikan dengan akurasi 86% pada data validasi dan pengujian, serta macro average F1-score sebesar 0.84 dan weighted average F1-score sebesar 0.86. Mel-Spec menghasilkan performa yang lebih stabil dan lebih baik, dengan F1-score tertinggi tercatat pada kelas 3Pnoep (0.96) dan 4Anthi (0.85). Secara keseluruhan, Mel-Spec memberikan hasil yang lebih baik dibandingkan MFCC, meskipun perbaikan masih diperlukan pada kelas-kelas dengan data terbatas, seperti 5Buer.

References

[1] S. Carvalho and E. F. Gomes, "Automatic Classification of Bird Sounds: Using MFCC and Mel Spectrogram Features with Deep Learning," *Vietnam J. Comput. Sci.*, vol. 10, no. 1, pp. 39–54, 2023, doi: 10.1142/S2196888822500300.

[2] Goni, A. W. ., Salaki, D. T. ., & Latumakulita , L. A. . (2021). Identifikasi Suara Burung Menggunakan Mel-Frequency Cepstral Coefficients Dan Backpropagation Neural Network. *Proceeding KONIK (Konferensi Nasional Ilmu Komputer)*, *5*(1), 9–13.

[3] Y. Tang, C. Liu, and X. Yuan, "Recognition of bird species with birdsong records using machine learning methods," *PLoS One*, vol. 19, no. 2 February, pp. 1–11, 2024, doi: 10.1371/journal.pone.0297988.

[4] J. Xie and M. Zhu, "Acoustic Classification of Bird Species Using an Early Fusion of Deep Features," *Birds*, vol. 4, no. 1, pp. 138–147, 2023, doi: 10.3390/birds4010011.

[5] H. R. Paleva and B. H. Prasetio, "Penerapan Short Time Fourier Transform pada MFCC untuk Sistem Pengenalan Ucapan Tingkat Stres," vol. 1, no. 1, pp. 1–10, 2024.

[6] S. Santoso, R. Hartayu, C. Anam, and Dimas Abdul Aziz, "Simulasi Simulasi Ekstraksi Fitur Suara menggunakan Mel-Frequency Cepstrum Coefficient," *J. Sains dan Inform.*, vol. 8, no. 1, pp. 80–87, 2022, doi: 10.34128/jsi.v8i1.357.

[7] H. Wang, Y. Xu, Y. Yu, Y. Lin, and J. Ran, "An Efficient Model for a Vast Number of Bird Species Identification Based on Acoustic Features," *Animals*, vol. 12, no. 18, 2022, doi: 10.3390/ani12182434.

[8] T. Nasution, "Metoda Mel Frequency Cepstrum Coefficients (MFCC) untuk Mengenali Ucapan pada Bahasa Indonesia," *SATIN - Sains dan Teknol. Inf.*, vol. 1, pp. 22–31, 2012.

[9] I. B. L. M. Suta, R. S. Hartati, and Y. Divayana, "Diagnosa Tumor Otak Berdasarkan Citra MRI (Magnetic Resonance Imaging).," JTE 18, 149–154, 2019.

[10] P. A. Nugroho, I. Fenriana, and R. Arijanto, "Implementasi Deep Learning Menggunakan Convolutional Neural Network (Cnn) Pada Ekspresi Manusia," Algor, vol. 2, no. 1, pp. 12–21, 2020.

[11] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," ISPRS J. Photogramm. Remote Sens., vol. 173, no. November 2020, pp. 24–49, 2021, doi:10.1016/j.isprsjprs.2020.12.010.

[12] S. Y. Yusdiantoro and T. B. Sasongko, "Implementasi Algoritma MFCC dan CNN dalam Klasifikasi Makna Tangisan Bayi," *Indones. J. Comput. Sci.*, vol. 12, no. 4, pp. 1957–1968, 2023, doi: 10.33022/ijcs.v12i4.3243.