SAIN TEKOM
ISSN 2088-1770 (Print)
ISSN 2503-3247 (Online)
Vol. 16 No.1 (2026) 72-84

JURNAL
SAINTEKOM
Sains, Teknologi, Komputer, dan Manajemen

Pengembangan Model Prediksi Speech Recognition dengan Algoritma Deep Learning Convolutional Neural Network

*Abdul Halim Anshor[1], Aswan Supriyadi Sunge[2]
[1,2)] Teknik Informatika, Universitas Pelita Bangsa
Jalan Inspeksi Kali Malang - Tegal Danas, Cikarang, Cibatu, Cikarang Sel., Bekasi, Jawa Barat
Email: [1]abdulhalimanshor@pelitabangsa.ac.id, [2]aswan.sunge@pelitabangsa.ac.id

ABSTRACT

This study examines the development of an automatic speech recognition (ASR) system in Sundanese, which still faces data limitations. Dialect variations and the lack of labeled data are the main challenges in the speech recognition process. The approach used is a Convolutional Neural Network (CNN) with Mel-Frequency Cepstral Coefficients (MFCC) feature extraction. The data used were 100 voice recordings consisting of West Sundanese and South Sundanese dialects. The processing process was carried out through the stages of pre-emphasis, framing, windowing, Fourier transform, Mel filter bank, and Discrete Cosine Transform to obtain voice features. The data was divided into 80% training data and 20% test data. The CNN model was then trained to recognize the voice patterns of each dialect. Based on the test results, the model achieved an accuracy of 70% with a loss value of 0.60. These results indicate that the approach used can be applied to limited data, although its performance can still be improved in further research.

Keywords : sundanese language; automatic speech recognition; convolutional neural network; mel-frequency cepstral coefficients; language dialect

ABSTRAK

Penelitian ini mengkaji pengembangan sistem pengenalan suara otomatis (Automatic Speech Recognition/ASR) pada bahasa Sunda yang masih memiliki keterbatasan data. Variasi dialek serta minimnya data berlabel menjadi tantangan utama dalam proses pengenalan suara. Pendekatan yang digunakan adalah *Convolutional Neural Network (CNN)* dengan ekstraksi fitur *Mel-Frequency Cepstral Coefficients (MFCC)*. Data yang digunakan sebanyak 100 rekaman suara yang terdiri dari dialek Sunda Barat dan Sunda Selatan. Proses pengolahan dilakukan melalui tahapan pre-emphasis, framing, windowing, transformasi Fourier, *Mel filter bank*, serta *Discrete Cosine Transform* untuk memperoleh fitur suara. Data dibagi menjadi 80% data latih dan 20% data uji. Model CNN kemudian dilatih untuk mengenali pola suara dari masing-masing dialek. Berdasarkan hasil pengujian, model memperoleh akurasi sebesar 70% dengan nilai loss sebesar 0,60. Hasil tersebut menunjukkan bahwa pendekatan yang digunakan sudah dapat diterapkan pada data dengan jumlah terbatas, meskipun performanya masih dapat ditingkatkan pada penelitian selanjutnya.

Kata kunci : bahasa sunda; *automatic speech recognition; convolutional neural network; mel-frequency cepstral coefficients*; dialek bahasa

Diterima Redaksi: 1-02-2026 | Selesai Revisi: 10-04-2026 | Diterbitkan Online: 31-03-2026

1. PENDAHULUAN

Teknologi pengenalan suara berkembang pesat dan telah dimanfaatkan dalam berbagai aplikasi berbasis kecerdasan buatan, seperti asisten virtual dan sistem pembelajaran bahasa (Aminuddin, 2023), (N. Y.-H. Wang et al., 2021). Namun, sebagian besar penelitian dan pengembangan teknologi ini masih berfokus pada bahasa-bahasa utama seperti Inggris, Mandarin, dan Spanyol (Khysru et al., 2022), sementara bahasa daerah seperti Sundanese belum mendapat perhatian yang memadai (Azis et al., 2021).

Bahasa Sundanese, salah satu bahasa daerah terbesar di Indonesia dengan jutaan penutur, memiliki karakteristik fonetik dan linguistik yang unik (Azis et al., 2021). Perbedaan intonasi dan pengucapan antara dialek Sunda Barat dan Sunda Selatan menimbulkan variasi karakteristik akustik yang menjadi tantangan dalam pengenalan suara. (Soekarta et al., 2023). Tantangan ini membuat bahasa Sundanese membutuhkan pendekatan khusus dalam pengembangannya untuk teknologi pengenalan suara, seperti yang telah dilakukan dalam penelitian sebelumnya yang berkaitan dengan dialek lokal (N. Y.-H. Wang et al., 2021), (Khysru et al., 2022).

Untuk menangani kompleksitas tersebut, digunakan teknik ekstraksi fitur MFCC yang mampu merepresentasikan informasi spektral suara dalam bentuk numerik untuk diproses oleh model pembelajaran mendalam seperti CNN (N. Y.-H. Wang et al., 2021), (International, 2024). MFCC efektif dalam pengenalan ucapan karena mampu merepresentasikan frekuensi suara untuk membedakan dialek bahasa Sunda (Azis et al., 2021), (Rahman Sya'ban et al., 2022), (Song, 2023).

CNN yang awalnya dikembangkan untuk pemrosesan gambar, kini sering diterapkan dalam pengenalan pola dari data suara yang dikonversi menjadi spektrogram (Aminuddin, 2023), (Akhiril Anwar Harahap et al., 2024). Kemampuan CNN dalam memodelkan pola kompleks pada data berdimensi tinggi membuatnya efektif dalam mengenali karakteristik akustik pada sinyal suara (Rehman et al., 2023a). Dalam pengenalan suara, CNN dapat dilatih menggunakan fitur yang diekstrak oleh MFCC, sehingga mampu mengidentifikasi perbedaan suara yang terkait dengan dialek yang berbeda dengan lebih tepat (Joelianto et al.,

DOI: https://doi.org/10.33020/saintekom.v16i1.1020

2024), (R. Wang et al., 2022). Radford menegaskan pentingnya fitur spektral, termasuk MFCC, dalam pengenalan suara (Radford et al., 2022).

Meskipun CNN terbukti efektif dalam pengenalan suara, sebagian besar penelitian masih berfokus pada bahasa utama seperti Inggris (Aminuddin, 2023). Sementara penelitian yang secara khusus membahas bahasa daerah seperti bahasa Sunda masih jarang (Khysru et al., 2022), (Rendi Nurcahyo & Mohammad Iqbal, 2022). Penelitian ini mengkaji pengenalan suara dialek Sunda Barat dan Selatan menggunakan 100 sampel suara dari dataset lokal Kaggle untuk melatih model CNN.

Selain itu, penelitian sebelumnya umumnya menggunakan dataset dengan jumlah data yang besar tanpa mempertimbangkan variasi dialek tertentu (H. Wang et al., 2022), (Utami et al., 2023). Meskipun menggunakan data terbatas, model CNN dioptimalkan melalui strategi pelatihan dan augmentasi untuk mencapai akurasi yang memadai (Soekarta et al., 2023), (Rahman Sya'ban et al., 2022). Berbeda dengan Soekarta et al. yang menggunakan data besar, penelitian ini menekankan adaptasi metode untuk data terbatas dan kompleks seperti dialek. (Soekarta et al., 2023).

Penelitian sebelumnya menunjukkan bahwa penggunaan algoritma deep learning mampu menangani data dengan kompleksitas tinggi, termasuk pada pemrosesan sinyal suara (Rehman et al., 2023b). Namun, penelitian ini memilih untuk menggunakan CNN karena kemampuannya untuk menangani data yang memiliki berbagai macam frekuensi, yang sangat penting dalam pengenalan dialek Sunda (Khysru et al., 2022) , (Akhiril Anwar Harahap et al., 2024).

Dalam pengenalan suara bahasa Sunda, CNN memanfaatkan fitur MFCC untuk mempelajari pola akustik tiap dialek dan membedakan Sunda Barat serta Selatan meskipun data terbatas. (Joelianto et al., 2024). MFCC menangkap informasi spektral penting untuk membedakan karakteristik suara antar dialek. (Rendi Nurcahyo & Mohammad Iqbal, 2022).

CNN memanfaatkan fitur MFCC untuk mempelajari dan membedakan pola akustik dialek Sunda Barat dan Selatan meskipun data terbatas (Joelianto et al., 2024). Penggunaan MFCC memungkinkan untuk

DOI: https://doi.org/10.33020/saintekom.v16i1.1020

menangkap informasi spektral yang detail, yang sangat penting dalam membedakan karakteristik suara antar dialek (Rendi Nurcahyo & Mohammad Iqbal, 2022).

Penelitian ini merumuskan pendekatan untuk mengatasi keterbatasan data pada pengembangan pengenalan suara bahasa daerah melalui penerapan augmentasi data dan strategi pelatihan yang disesuaikan. Pendekatan ini bertujuan menjaga performa model tetap optimal meskipun jumlah sampel terbatas.(Wu et al., 2023).

Penelitian ini mengembangkan model pengenalan suara bahasa Sunda berbasis CNN dan MFCC untuk mendukung penguatan teknologi ASR bahasa daerah serta pelestarian bahasa lokal (Aminuddin, 2023).

Fokus utama penelitian ini bagaimana mengimplementasikan metode CNN yang dikombinasikan dengan fitur MFCC pada dataset bahasa Sunda yang terbatas.

Beberapa Studi sebelumnya telah menggunakan berbagai arsitektur deep learning, seperti RNN, LSTM, dan Transformer untuk pengenalan suara (N. Y.-H. Wang et al., 2021, 2021), (Wu et al., 2023). Metode CNN lebih menjadi pilihan terbaik karena memiliki kemampuan untuk mengenali pola lokal pada representasi spektrum suara dengan kebutuhan komputasi yang relatif rendah (Rehman et al., 2023b). Disamping itu, penelitian ini berfokus pada pengenalan suara dengan sumber daya rendah, yang belum banyak dikaji untuk bahasa daerah seperti Sunda (Khysru et al., 2022). Oleh karena itu, tujuan utama dari penelitian ini adalah untuk mengevaluasi kinerja kombinasi CNN dan MFCC pada dataset terbatas dan mengevaluasi kapasitasnya sebagai model dasar dibandingkan dengan metode lain yang lebih kompleks.

2. METODE

Penelitian ini menerapkan pendekatan kuantitatif eksperimental untuk mengembangkan model pengenalan suara berbasis CNN dengan fitur MFCC.

2.1. Data Penelitian

Dataset terdiri dari 100 rekaman suara bahasa Sunda yang diperoleh dari Kaggle, dengan distribusi seimbang antara dialek Sunda Barat dan Sunda Selatan (masing-masing 50 sampel https://www.kaggle.com/datasets/fitrohamri08/suara-dialek-sunda . Data direkam dalam kondisi lingkungan hening dan bising. Pada kondisi bising, rekaman mengandung noise latar seperti suara

DOI: https://doi.org/10.33020/saintekom.v16i1.1020

lingkungan sekitar. Namun, untuk meminimalkan pengaruh noise terhadap proses pelatihan, dilakukan tahap preprocessing berupa normalisasi sinyal dan ekstraksi fitur MFCC yang mampu mereduksi dampak noise. Seluruh rekaman juga diseragamkan dalam format dan parameter audio sebelum diproses. Seluruh rekaman diseragamkan dalam format dan parameter audio sebelum diproses. Dataset kemudian dibagi menjadi 80% data pelatihan dan 20% data pengujian

2.2. Ekstraksi Fitur MFCC

Ekstraksi fitur dilakukan untuk merepresentasikan karakteristik spektral sinyal suara dalam bentuk numerik. Tahap awal adalah pre-emphasis yang dihitung menggunakan Persamaan (1):

$$y[n] = s[n] - \propto . s[n-1] \quad (1)$$

Dimana:
y[n] = sinyal hasil pre-emphasis
s[n] = sinyal asli
α = koefisien pre-emphasis
n = indeks sampel

Transformasi ke domain frekuensi dilakukan menggunakan FFT sebagaimana pada Persamaan (2):

$$(n) = xi(n)w(n), n = 0,1,1, \dots, N \quad (2)$$

Dimana:
S(k) = spektrum frekuensi
x(n) = sinyal pada domain waktu
k = indeks frekuensi

N = jumlah sampel per frame
j = bilangan imajiner

Konversi ke skala Mel dilakukan menggunakan Persamaan (3):

$$\text{Mel(f)} = 2595\log10(1 + \text{f}/700\text{f}) \quad (3)$$

Dimana:
Mel(f) = frekuensi pada skala Mel
f = frekuensi dalam Hertz

Koefisien MFCC diperoleh melalui DCT menggunakan Persamaan (4):

$$C(k) = 2\sum_{n=0}^{N-1} x(n) cos \frac{\pi(2n+1)k}{2N} \quad (4)$$

Dimana:
C(k) = koefisien MFCC ke-k
x(n) = energi hasil filter Mel
k = indeks koefisien
N = jumlah filter

Sebanyak 13 koefisien MFCC digunakan sebagai fitur masukan model.

2.3 Pelatihan Model CNN.

Penelitian ini mengembangkan model pengenalan suara berbasis CNN

DOI: https://doi.org/10.33020/saintekom.v16i1.1020

dan MFCC melalui pendekatan kuantitatif eksperimental (5) :

$$\theta_{t+1=}\theta_t \, n.\frac{m_t}{\sqrt{v_t+\epsilon}} \quad (5)$$

Dimana:
θ_t = parameter pada iterasi ke-t
η = learning rate
m_t = estimasi rata-rata gradien
v_t = estimasi rata-rata kuadrat gradien
ε = konstanta stabilitas numerik

Fungsi loss yang digunakan adalah categorical cross-entropy pada Persamaan (6):

$$C(k) = 2\sum_{n=0}^{N-1} x(n)cos\frac{\pi(2n+1)k}{2N} \quad (6)$$

Dimana:
y_i = label aktual
ŷ_i = probabilitas prediksi
N = jumlah sampel

2.4. Evaluasi Model

Kinerja model dievaluasi menggunakan beberapa matrik yaitu Akurasi, presicion, recall, dan F1-Score. Matrik akurasi merepresentasikan rasio prediksi yang benar (baik positif maupun negatif) terhadap total seluruh sampel uji. Nilai akurasi yang dihitung berdasarkan Persamaan (7):

$$\boldsymbol{Akurasi} = \frac{\boldsymbol{Jumlah\ Prediksi\ Benar}}{\boldsymbol{Jumlah\ Sampel}} \boldsymbol{X\ 100\%} \quad (7)$$

Dimana:
Jumlah Prediksi Benar = total klasifikasi sesuai label
Jumlah Sampel = total data pengujian

Confusion matrix digunakan untuk menganalisis distribusi kesalahan klasifikasi antar dialek.

Disamping itu evaluasi dilakukan dengan matrik Precision menunjukkan tingkat ketepatan antara data yang diminta dengan hasil prediksi yang diberikan oleh model persamaan precisi ditunjukkan oleh persamaan 8.

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

Matrik Recall digunakan untuk mengukur keberhasilan model dalam menemukan kembali sebuah informasi (kelas) dari keseluruhan data yang tersedia, untuk menghitung nilai recall ditunjukkan oleh persamaan 9.

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

Matrik F1 score bertujuan untuk mengintegrasikan aspek ketepatan dan sensitivitas model dalam membedakan

DOI: https://doi.org/10.33020/saintekom.v16i1.1020

variasi dialek Persamaan yang digunakan untuk menghitung nilai F1 score ditunjukkan oleh persamaan 10

$$F1 - Score = 2\ \frac{Precision.Recall}{Precision+REcall} (10)$$

3. HASIL DAN PEMBAHASAN

Ekstraksi fitur MFCC dilakukan untuk mendapatkan informasi penting dari sinyal suara, Adapun proses ekstraksi fitur MFCC terdiri dari beberapa tahapan. Tahap yang pertama pre-emphasis, tahapan ini bertujuan untuk meningkatkan komponen frekuensi tinggi sehingga sinyal suara menjadi representasi numerik yang dapat diproses oleh model CNN. Selanjutnya, teknik framing digunakan untuk membagi sinyal menjadi beberapa frame. Frame ini memiliki ukuran frame 25 milidetik dan jarak frame 10 milidetik, yang menghasilkan overlap sebesar 15 milidetik untuk menjaga kontinuitas sinyal.

Tahapan berikutnya windowing fungsi Hamming, dimana tahapan akan memproses setiap frame untuk mengurangi efek diskontinuitas pada tepi frame. Selanjutnya, sinyal akan diubah ke domain frekuensi menggunakan transformasi Fourier (FFT) dengan jumlah titik FFT 512 untuk menghasilkan spektrum daya.

Pada tahap berikutnya, dilakukan pemetaan ke skala Mel menggunakan filter bank sebanyak 40 filter untuk meniru persepsi pendengaran manusia terhadap frekuensi suara. Proses ini menghasilkan representasi Mel-spectrogram yang menekankan frekuensi rendah dibandingkan frekuensi tinggi.

Selanjutnya, hasil Mel-spectrogram dikompresi menggunakan transformasi discrete cosine (DCT) untuk menghasilkan tiga belas koefisien MFCC yang menunjukkan fitur utama sinyal suara. Koefisien ini dimasukkan ke dalam model CNN untuk proses klasifikasi dialek.

Hasil menunjukkan bahwa pola spektral yang relatif konsisten dalam satu dialek dihasilkan oleh setiap rekaman suara, tetapi ada variasi antar dialek yang menjadi dasar model untuk klasifikasi. Hal ini menunjukkan bahwa fitur MFCC dapat mengidentifikasi perbedaan akustik antara dialek Sunda Banten dan Sunda Selatan.

DOI: https://doi.org/10.33020/saintekom.v16i1.1020

Tabel 1. Hasil Proses MFCC

MFCC Feature Voice Recording 1	Fitur MFCC 1	Fitur MFCC 2	Fitur MFCC 3	…….	Fitur MFCC 13	Dialect
1	-30.592.072	19.048.987	-11.232.471	…….	17.895.751	West
2	-3.225.896	18.630.025	-1.644.849	…….	0.190.044	West
3	-3.247.438	19.591.693	-5.007.981	…….	1.577.572	West
4	-33.368.457	17.907.031	13.588.304	…….	-20.210.373	West
5	-3.354.729	19.246.504	-7.352.086	…….	39.016.023	West
6	-31.121.466	18.178.156	30.424.871	…….	0.487.063	West
7	-33.910.632	18.296.077	-9.557.931	…….	2.497.453	West
8	-3.064.007	18.064.787	8.289.159	…….	0.419.405	West
9	-32.516.586	18.623.717	0.095.455	…….	17.717.685	West
10	-26.932.578	18.869.408	2.234.495	…….	-0.561.695	West
…….	…….	…….	…….	…….	…….	…….
100	-38.785.645	14.822.371	7.441.543	…….	-19.110.253	South

Tabel 1 merupakan hasil proses MFCC. Vektor fitur MFCC yang terdiri dari 13 koefisien pada 100 sampel rekaman suara bahasa Sunda. Setiap baris merepresentasikan satu sampel suara dengan 13 nilai MFCC yang menangkap karakteristik spektral suara. Koefisien MFCC 1–13 mencerminkan distribusi energi pada berbagai pita frekuensi; koefisien rendah (MFCC 1–3) merepresentasikan informasi umum dan energi keseluruhan, sedangkan koefisien yang lebih tinggi (MFCC 8–13) menangkap detail variasi spektral yang lebih halus. Nilai negatif dan positif menunjukkan fase dan intensitas relatif pada setiap pita frekuensi. Adanya perbedaan pola nilai antar sampel dialek West dan South menjadi dasar bagi model klasifikasi untuk membedakan kedua dialek tersebut.

3.1 Hasil Ekstraksi Fitur MFCC

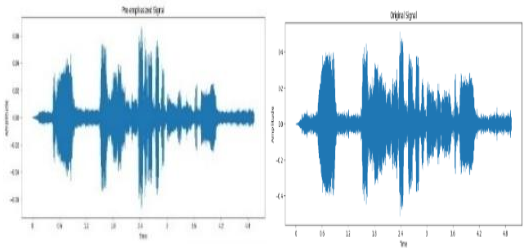
a. Hasil Pre-emphasis

proses pre-emphasis menghasilkan nilai sinyal sebesar −8.1078242. Nilai negatif yang cukup besar ini menunjukkan terjadinya penguatan amplitudo yang sangat signifikan pada komponen frekuensi tinggi. Sebelum pre-emphasis, amplitudo frekuensi tinggi cenderung lemah dan mudah tertutup oleh energi frekuensi rendah, setelah proses ini, komponen frekuensi tinggi menjadi jauh lebih dominan. Perubahan ini sangat penting karena informasi fonetik yang paling tinggi berada pada rentang

DOI: https://doi.org/10.33020/saintekom.v16i1.1020

frekuensi tinggi. Dengan demikian, pre-emphasis berhasil meningkatkan kualitas representasi spektral sinyal, sehingga tahap ekstraksi fitur MFCC selanjutnya dapat menangkap detail akustik yang lebih tajam dan relevan untuk membedakan dialek bahasa Sunda.

Visualisasi hasil pre-emphasis ditunjukkan pada Gambar 1.

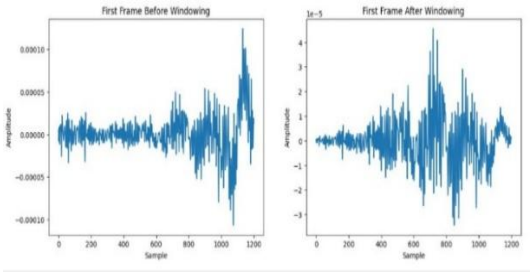


Gambar 1 Perbandingan Sinyal Sebelum dan Sesudah Pra-penekanan

Gambar 1 memperlihatkan peningkatan amplitudo pada frekuensi tinggi setelah pre-emphasis, yang memperjelas informasi fonetik untuk proses ekstraksi selanjutnya.

b. Windowing

Setelah pre-emphasis, sinyal dibagi menjadi frame-frame pendek dan dikalikan dengan fungsi jendela Hamming. Pada frame pertama sampel yang sama, diperoleh nilai w(0) = 0,08 w yang menghasilkan x(0) = 5,71032054 . Nilai ini mencerminkan penyesuaian amplitudo yang halus sesuai karakteristik jendela Hamming. Proses windowing ini efektif meredam amplitudo pada tepi frame, sehingga mengurangi efek diskontinuitas dan kebocoran spektral saat transformasi ke domain frekuensi. Visualisasi proses tersebut ditampilkan pada Gambar 2.



Gambar 2 Efek Proses Windowing pada Frame Pertama

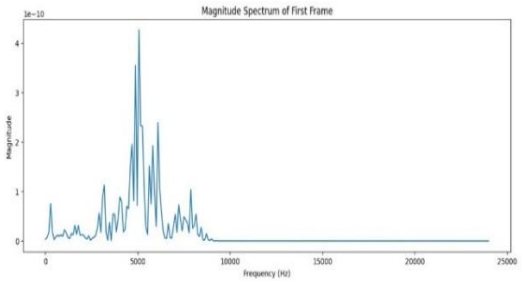
Grafik menunjukkan peredaman amplitudo pada tepi frame setelah windowing, yang mengurangi diskontinuitas dan kebocoran spektral saat transformasi Fourier.

c. Fast Fourier Transform (FFT)

Sinyal yang telah di-window kemudian ditransformasikan ke domain frekuensi menggunakan FFT dengan panjang N = 512 . Pada frame pertama diperoleh nilai spektrum awal S(0) = 7.08910173. Nilai ini merepresentasikan komponen frekuensi dasar sinyal. Spektrum magnitudo hasil FFT ditunjukkan pada Gambar 3. Puncak-puncak frekuensi dominan pada grafik tersebut menunjukkan resonansi vokal

DOI: https://doi.org/10.33020/saintekom.v16i1.1020

yang khas dan menjadi dasar pemetaan ke skala Mel.



Gambar 3 Spektrum Magnitudo Bingkai Pertama Setelah FFT

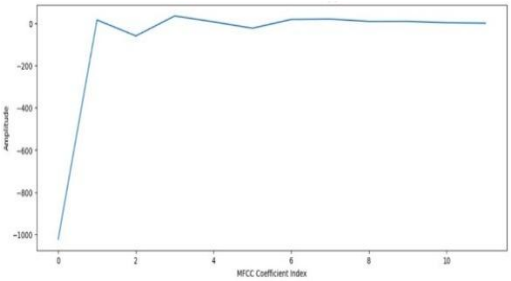
Puncak frekuensi dominan pada grafik mencerminkan resonansi suara dan menjadi dasar pemetaan ke skala Mel, menunjukkan keberhasilan FFT dalam merepresentasikan sinyal pada domain frekuensi.

d. Mel Filter Bank

Pemetaan ke skala Mel menghasilkan rentang nilai 0–236,4045 melalui Mel Filter Bank. Hasil Mel Filter Bank akan menjadi input DCT untuk pengelompokan energi spektral sebelum kompresi fitur.

e. Discrete Cosine e. Transform (DCT)

Pada tahapan ini diperoleh koefisien MFCC, dengan nilai koefisien pertama C(0) = -3,24568825. Koefisien ini mencerminkan energi rata-rata keseluruhan sinyal, sedangkan koefisien berikutnya menggambarkan variasi spektral yang lebih halus. Distribusi keseluruhan koefisien MFCC ditunjukkan pada Gambar 4. Perbedaan amplitudo antar koefisien secara jelas merepresentasikan karakteristik akustik yang membedakan dialek West dan South, sehingga menjadi fitur input yang optimal bagi model CNN.



Gambar 4 Distribusi Koefisien MFCC Setelah Transformasi Kosinus Diskrit

3.2 Pelatihan Model CNN

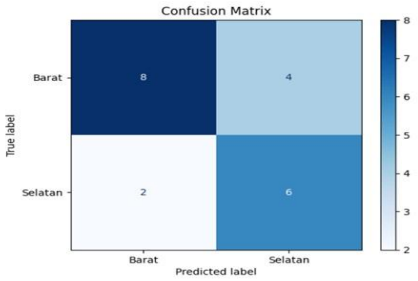
Model CNN dilatih pada fitur MFCC dari 100 rekaman dengan skema 80% pelatihan dan 20% pengujian. Model menggunakan lapisan konvolusi, pooling, dan fully connected dengan ReLU–Softmax, dioptimasi oleh Adam dan loss cross-entropy. Model dilatih hingga loss konvergen dan selanjutnya dievaluasi pada data pengujian.

3.3 Evaluasi Model

Model dievaluasi menggunakan data uji sebanyak 20 sampel (20% dari total dataset). Hasil evaluasi

DOI: https://doi.org/10.33020/saintekom.v16i1.1020

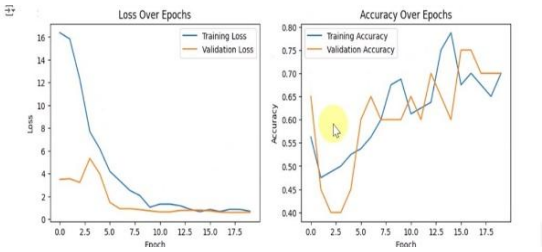
menunjukkan akurasi keseluruhan sebesar 70% (14/20). Confusion matrix pada Gambar 5 menggambarkan distribusi prediksi model secara rinci: model berhasil mengklasifikasikan 8 sampel dialek Barat dan 6 sampel dialek Selatan dengan benar, sementara terdapat 4 sampel Barat yang salah diklasifikasikan sebagai Selatan dan 2 sampel Selatan yang salah diklasifikasikan sebagai Barat.



Gambar 5 Confusion Matrix

Evaluasi matrik lainnya dihitung melalui confusion matrix tersebut. Dari hasil eksperimen menunjukkan performa nilai macro average untuk precision sebesar 0,70, nilai recall sebesar 0,71, dan nilai F1-score sebesar 0,70. Hasil ini menunjukkan bahwa model memiliki kemampuan prediktif yang stabil pada setiap kategori yang diuji.

Nilai Loss yang dihasilkan sebesar 0,6 hal ini menunjukkan pembelajaran pola akustik telah terbentuk, namun belum optimal. Gambar 6 menunjukkan penurunan loss dan stabilisasi pada akhir pelatihan.



Gambar 6 Perkembangan Nilai Loss dan Akurasi Selama Pelatihan

Secara umum, model menunjukkan performa yang cukup baik dalam mengidentifikasi dialek bahasa Sunda pada dataset terbatas.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa kombinasi CNN dan MFCC efektif untuk pengenalan suara bahasa Sunda dengan data terbatas, ditunjukkan oleh akurasi 70% dan loss 0,6. Studi ini berkontribusi pada pengembangan ASR untuk bahasa daerah dengan sumber daya rendah. Keterbatasan penelitian terletak pada jumlah data dan kompleksitas perbedaan fonem antar dialek. Penelitian selanjutnya disarankan memperluas dataset dan mengeksplorasi model yang lebih kompleks guna meningkatkan kinerja sistem.

DOI: https://doi.org/10.33020/saintekom.v16i1.1020

5. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Pelita Bangsa atas dukungan, fasilitas, dan lingkungan akademik yang telah diberikan sehingga penelitian ini dapat diselesaikan dengan baik**.**

DAFTAR PUSTAKA

Akhiril Anwar Harahap, Novita, R., Ahsyar, T. K., & Zarnelly, Z. (2024). Classification of Beef and Pork with Deep Learning Approach. *Jurnal Sistem Cerdas*, *7*(1), 55–65. https://doi.org/10.37396/jsc.v7i1.393

Aminuddin, M. (2023). English Spoken Digit Recognition using Convolutional Neural Network (CNN). *Jurnal EEICT (Electric Electronic Instrumentation Control Telecommunication)*, *6*(2). https://doi.org/10.31602/eeict.v6i2.11877

Azis, N., Herwanto, H., & Ramadhani, F. (2021). Implementasi Speech Recognition Pada Aplikasi E-Prescribing Menggunakan Algoritme Convolutional Neural Network. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *5*(2), 460. https://doi.org/10.30865/mib.v5i2.2841

International, O. T. (2024). Retracted: A Classroom Emotion Recognition Model Based on a Convolutional Neural Network Speech Emotion Algorithm. *Occupational Therapy International*, *2024*(1), 9825450. https://doi.org/10.1155/2024/9825450

Joelianto, E., Mandasari, M. I., Marpaung, D. B., Hafizhan, N. D., Heryono, T., Prasetyo, M. E., Dani, Tjahjani, S., Anggraeni, T., & Ahmad, I. (2024). Convolutional neural network-based real-time mosquito genus identification using wingbeat frequency: A binary and multiclass classification approach. *Ecological Informatics*, *80*, 102495. https://doi.org/10.1016/j.ecoinf.2024.102495

Khysru, K., Wei, J., & Dang, J. (2022). Research on Tibetan Speech Recognition Based on the Am-do Dialect. *Computers, Materials & Continua*, *73*(3), 4897–4907. https://doi.org/10.32604/cmc.2022.027591

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust Speech Recognition via Large-Scale Weak Supervision* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2212.04356

Rahman Sya'ban, D., Hamzah, A., & Susanti, E. (2022). Klasifikasi Buah Segar Dan Busuk Menggunakan Algoritma Convolutional Neural Network Dengan Tflite Sebagai Media Penerapan Model Machine Learning. *PROSIDING SNAST*, F7-16. https://doi.org/10.34151/prosidingsnast.v8i1.4180

Rehman, A., Kim, D., & Paul, A. (2023a). Convolutional Neural Network Model for Fire Detection in Real-Time Environment. *Computers, Materials & Continua*, *77*(2), 2289–2307.

DOI: https://doi.org/10.33020/saintekom.v16i1.1020

https://doi.org/10.32604/cmc.2023.036435

Rehman, A., Kim, D., & Paul, A. (2023b). Convolutional Neural Network Model for Fire Detection in Real-Time Environment. *Computers, Materials & Continua*, *77*(2), 2289–2307. https://doi.org/10.32604/cmc.2023.036435

Rendi Nurcahyo & Mohammad Iqbal. (2022). Pengenalan Emosi Pembicara Menggunakan Convolutional Neural Networks. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *6*(1), 115–122. https://doi.org/10.29207/resti.v6i1.3726

Soekarta, R., Nurdjan, N., & Syah, A. (2023). Klasifikasi Penyakit Tanaman Tomat Menggunakan Metode Convolutional Neural Network (CNN). *Insect (Informatics and Security): Jurnal Teknik Informatika*, *8*(2), 143–151. https://doi.org/10.33506/insect.v8i2.2356

Song, Y. (2023). Chinese Speech Recognition System Based on Neural Network Acoustic Network Model. *Procedia Computer Science*, *228*, 144–154. https://doi.org/10.1016/j.procs.2023.11.018

Utami, N. W., I Nyoman Purnama, & I Putu Restu Prajna. (2023). Klasifikasi Tanaman Upakara Adat Hindu Di Kebun Raya Eka Karya Bali Menggunakan Algoritma Convolutional Neural Network. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, *5*(4), 671–678. https://doi.org/10.51401/jinteks.v5i4.3416

Wang, H., Zhang, W.-Q., Suo, H., & Wan, Y. (2022). Multilingual Zero Resource Speech Recognition Base on Self-Supervise Pre-Trained Acoustic Models. *2022 13th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 11–15. https://doi.org/10.1109/ISCSLP57327.2022.10037938

Wang, N. Y.-H., Wang, H.-L. S., Wang, T.-W., Fu, S.-W., Lu, X., Wang, H.-M., & Tsao, Y. (2021). Improving the Intelligibility of Speech for Simulated Electric and Acoustic Stimulation Using Fully Convolutional Neural Networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 184–195. https://doi.org/10.1109/TNSRE.2020.3042655

Wang, R., Lei, Z., Zhang, Z., & Gao, S. (2022). Dendritic Convolutional Neural Network. *IEEJ Transactions on Electrical and Electronic Engineering*, *17*(2), 302–304. https://doi.org/10.1002/tee.23513

Wu, J., Gaur, Y., Chen, Z., Zhou, L., Zhu, Y., Wang, T., Li, J., Liu, S., Ren, B., Liu, L., & Wu, Y. (2023). On Decoder-Only Architecture For Speech-to-Text and Large Language Model Integration. *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 1–8. https://doi.org/10.1109/ASRU57964.2023.10389705

DOI: https://doi.org/10.33020/saintekom.v16i1.1020