PRA-SKRIPSI

KLASIFIKASI DIALEK BAHASA JAWA DARI UCAPAN BERBASIS MEL-SPECTROGRAM DENGAN CONVOLUTIONAL NEURAL NETWORK

DANIKA NAJWA ARDELIA
NPM 22081010103

DOSEN PEMBIMBING
-
-

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2024

BAB I
PENDAHULUAN

1.1. Latar Belakang

Bahasa merupakan sarana utama manusia dalam berkomunikasi dan menyampaikan gagasan kepada orang lain. Dalam konteks kehidupan sosial dan budaya, bahasa juga berperan sebagai identitas suatu kelompok masyarakat. Di Indonesia, bahasa daerah menjadi bagian penting dari identitas budaya lokal yang perlu dilestarikan. Namun, berbagai data menunjukkan bahwa penggunaan bahasa daerah di Indonesia terus mengalami penurunan.

Berdasarkan hasil Sensus Penduduk 2020, sebanyak 74,77 persen penduduk Indonesia berumur lima tahun ke atas menggunakan bahasa daerah untuk berkomunikasi di lingkungan keluarga—menurun dibandingkan 79,64 persen pada SP2010. Tren serupa terlihat di lingkungan masyarakat (72,78 persen pada SP2020). Lebih jauh, penurunan ini bersifat generasional: semakin muda suatu generasi, semakin rendah kecenderungannya menggunakan bahasa daerah di lingkungan keluarga [1].

Fenomena tersebut menunjukkan adanya pergeseran pola komunikasi lintas generasi yang berpotensi memengaruhi vitalitas bahasa daerah, termasuk Bahasa Jawa yang memiliki jumlah penutur terbesar di Indonesia. Berdasarkan data Badan Pusat Statistik (2020), Bahasa Jawa dituturkan oleh sekitar 80 juta jiwa, atau 42,12 persen populasi nasional. Meskipun secara kuantitatif dominan, tren umum penurunan penggunaan bahasa daerah mengindikasikan bahwa Bahasa Jawa pun menghadapi tantangan serupa, khususnya dalam hal transmisi antargenerasi. Kondisi ini menjadi semakin kompleks mengingat Bahasa Jawa memiliki keberagaman dialek yang luas, yang juga perlu diperhatikan dalam upaya pelestarian dan dokumentasi bahasa.

Keberagaman dialek ini menunjukkan kompleksitas internal Bahasa Jawa yang tidak hanya kaya secara budaya, tetapi juga menantang dari perspektif linguistik. Dialek dapat dipahami sebagai variasi bahasa yang digunakan oleh sekelompok penutur di wilayah tertentu, muncul akibat perbedaan sosial, geografis, dan fungsi komunikasi dalam masyarakat [2]. Berdasarkan klasifikasi Uhlenbeck sebagaimana dikutip dalam penelitian [3], Bahasa Jawa terbagi menjadi tiga kelompok dialek utama, yaitu Jawa Ngapak, Jawa Tengahan, dan Jawa Timuran. Perbedaan ini mencerminkan

dinamika sosial dan sejarah persebaran penutur, sekaligus menjadikan dialek sebagai indikator identitas kedaerahan. Fenomena ini menegaskan pentingnya penelitian yang fokus pada variasi dialek, terutama untuk tujuan pelestarian dan dokumentasi bahasa.

Seiring dengan pentingnya pelestarian dialek, dialektologi modern tidak lagi hanya fokus pada studi deskriptif berdasarkan wilayah geografis. Metode komputasional, termasuk analisis *big data* dan *machine learning*, mulai dimanfaatkan untuk mempelajari perbedaan bahasa secara lebih sistematis [4]. Beberapa penelitian terdahulu menunjukkan keberhasilan pendekatan ini. Nahar dkk. menggunakan *machine learning* untuk identifikasi dialek Arab [5], sementara Fauzi dkk. mengekstraksi fitur MFCC untuk klasifikasi dialek di Pulau Jawa menggunakan ANFIS [6]. Pendekatan *deep learning* juga telah diterapkan, seperti penggunaan DNN untuk mengenali dialek di Sumatera Selatan dengan akurasi 73% [7], dan implementasi CNN dengan Mel-Spektogram untuk dialek Sunda yang mencapai akurasi 95% [8]. Keberhasilan metode CNN berbasis Mel-Spektrogram dalam studi-studi sebelumnya menjadi landasan kuat untuk penerapannya dalam penelitian ini.

Berdasarkan fakta-fakta dan penelitian terdahulu di atas, penelitian ini akan memfokuskan pada klasifikasi dialek Bahasa Jawa menggunakan Convolutional Neural Network berbasis fitur Mel-Spektogram. Pendekatan ini diharapkan tidak hanya memberikan akurasi tinggi dalam pengenalan dialek, tetapi juga mendukung pelestarian dan dokumentasi bahasa Jawa secara sistematis.

1.2. Rumusan Masalah

Berdasarkan latar belakang penelitian, penelitian ini difokuskan pada klasifikasi dialek Bahasa Jawa menggunakan metode komputasional berbasis Convolutional Neural Network (CNN) dan fitur Mel-Spektrogram. Rumusan masalah penelitian ini dirumuskan dalam bentuk pertanyaan penelitian sebagai berikut:

1. Bagaimana penerapan model Convolutional Neural Network (CNN) berbasis fitur Mel-Spektrogram untuk menyelesaikan masalah klasifikasi dialek Bahasa Jawa?
2. Bagaimana performa model Convolutional Neural Network (CNN) yang dihasilkan, diukur berdasarkan tingkat akurasi dan metrik evaluasi lainnya dalam mengklasifikasikan dialek Bahasa Jawa?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditentukan, maka tujuan dari penelitian ini adalah:

1. Mengimplementasikan model Convolutional Neural Network (CNN) yang memanfaatkan fitur Mel-Spektrogram untuk dapat menyelesaikan masalah klasifikasi dialek Bahasa Jawa.
2. Mengevaluasi dan menganalisis performa model yang telah dibangun dalam mengklasifikasikan dialek Bahasa Jawa, berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score.

1.4. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan serangkaian manfaat signifikan, baik dari sisi keilmuan maupun penerapan praktis, yang meliputi:

1. Memberikan kontribusi pada bidang dialektologi komputasional dengan menyajikan sebuah metode yang objektif dan terukur untuk analisis serta klasifikasi dialek, yang dapat mendukung studi pemetaan bahasa modern.
2. Menjadi referensi dan studi kasus bagi pengembangan ilmu komputer, khususnya dalam penerapan deep learning untuk tugas klasifikasi suara pada domain bahasa daerah di Indonesia.
3. Menghasilkan model klasifikasi yang berfungsi sebagai bentuk dokumentasi digital atas kekayaan dialek Bahasa Jawa, sehingga dapat dimanfaatkan sebagai alat bantu oleh lembaga kebudayaan dalam upaya pelestarian bahasa.
4. Menyediakan dasar fundamental untuk pengembangan teknologi berbasis suara di masa depan yang lebih sadar-dialek (dialect-aware), seperti aplikasi edukasi, layanan terjemahan, atau sistem speech-to-text yang lebih akurat untuk penutur lokal.

1.5. Batasan Masalah

Untuk menjaga agar penelitian ini tetap fokus dan terarah pada tujuan yang telah ditetapkan, maka ruang lingkup penelitian ini dibatasi oleh beberapa hal sebagai berikut:

1. Penelitian ini hanya berfokus pada klasifikasi tiga kelompok dialek utama Bahasa Jawa, yaitu dialek Jawa Ngapak (Banyumas-Cilacap), Jawa Tengahan (Solo), dan Jawa Timuran (Surabaya-Sidoarjo).

2. Data suara yang digunakan merupakan data primer yang diperoleh melalui proses perekaman langsung terhadap penutur asli dari setiap perwakilan dialek.
3. Data dibatasi pada rekaman suara dari penutur native usia 15-60 tahun dalam kondisi lingkungan tenang, dengan durasi ucapan per sampel sekitar 3-5 detik.
4. Ekstraksi fitur terbatas pada Mel Spectrogram tanpa kombinasi dengan fitur lain seperti MFCC atau prosodik lanjutan, sementara arsitektur CNN menggunakan konfigurasi sederhana (misalnya, 3-4 lapisan konvolusi) tanpa eksplorasi model hybrid seperti Transformer.

BAB II
TINJAUN PUSTAKA

2.1. Penelitian Terdahulu

Berbagai pendekatan komputasional telah diterapkan untuk mengatasi tantangan dalam identifikasi dialek. Namun, tidak semua pendekatan memberikan hasil yang memuaskan, terutama untuk dialek-dialek di Pulau Jawa. Sebuah studi oleh Fauzi et al. [6] mencoba mengklasifikasikan bahasa yang ada di Pulau Jawa, meliputi Bahasa Betawi, Sunda, Banyumasan, dan Suroboyoan menggunakan metode *Adaptive Network-based Fuzzy Inference System* (ANFIS) dengan fitur MFCC. Hasilnya menunjukkan akurasi yang sangat rendah, yaitu hanya 32,5%, yang disebabkan oleh keterbatasan data dan metode yang kurang optimal. Hal ini mengindikasikan bahwa klasifikasi dialek Jawa memerlukan pendekatan yang lebih canggih.

Peningkatan performa yang signifikan terlihat pada penelitian yang menerapkan metode berbasis jaringan saraf. Putra et al. [7] menggunakan *Deep Neural Network* (DNN) untuk mengenali dialek di Sumatera Selatan dan menemukan bahwa fitur Mel-Spectrogram efektif, dengan capaian akurasi tertinggi sekitar 72,7%. Hasil ini setara dengan penelitian di konteks internasional, seperti studi oleh Nahar et al. [5] pada 17 dialek Arab, di mana metode K-Nearest Neighbor (KNN) mampu mencapai akurasi 76%. Bahkan, dengan metode yang lebih kompleks seperti *Ensemble Support Vector Machine* (ESVM), Chittaragi & Koolagudi [9] berhasil mencapai akurasi 86,25% untuk klasifikasi dialek bahasa Kannada di India. Studi-studi ini menunjukkan bahwa akurasi di atas 70-85% adalah target yang realistis dengan metode yang tepat.

Terobosan paling signifikan dalam konteks bahasa daerah di Indonesia ditunjukkan oleh Setianingrum et al. [8]. Penelitian mereka menerapkan Convolutional Neural Network (CNN) dengan ekstraksi fitur Mel-Spectrogram untuk membedakan dialek Sunda. Dengan mengubah sinyal suara menjadi representasi gambar (spektrogram) dan mengolahnya dengan arsitektur CNN yang memang dirancang untuk data visual, penelitian ini berhasil mencapai akurasi pengujian yang luar biasa, yaitu 95,00%. Keberhasilan ini membuktikan bahwa kombinasi CNN dan Mel-Spectrogram adalah pendekatan termutakhir dengan performa paling unggul untuk tugas klasifikasi dialek di Indonesia.

Dari perbandingan tersebut, terlihat sebuah celah penelitian yang sangat jelas.

Di satu sisi, pendekatan dengan metode ANFIS terbukti tidak berhasil pada dialek Jawa. Di sisi lain, pendekatan superior menggunakan CNN dan Mel-Spectrogram telah terbukti sangat sukses pada dialek Sunda yang serumpun. Hingga saat ini, metode dengan performa tertinggi tersebut belum pernah diterapkan secara spesifik untuk mengatasi masalah klasifikasi dialek Bahasa Jawa. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan mengimplementasikan dan mengevaluasi model CNN berbasis fitur Mel-Spectrogram untuk klasifikasi dialek Jawa Ngapak, Tengahan, dan Timuran, guna melihat apakah keberhasilan serupa dapat dicapai.

2.2. Landasan Teori

2.2.1. Bahasa dan Dialek

Bahasa merupakan salah satu aspek yang tidak dapat dipisahkan dari kehidupan manusia sebagai makhluk sosial. Melalui bahasa, manusia dapat saling menyampaikan informasi dan menjalin komunikasi dalam kehidupan sehari-hari [10]. Di Indonesia, selain bahasa nasional, terdapat keragaman bahasa daerah yang sangat kaya, dengan total 718 bahasa daerah yang telah teridentifikasi, tidak termasuk dialek dan subdialeknya [11].

Salah satu bahasa daerah dengan jumlah penutur terbesar di Indonesia adalah bahasa Jawa [1]. Menurut Zulaeha dalam Dicta et al. [10], bahasa ini utamanya digunakan sebagai bahasa ibu oleh suku Jawa yang mendiami wilayah Provinsi Jawa Tengah, Daerah Istimewa Yogyakarta (DIY), dan Jawa Timur. Dalam praktiknya, penggunaan bahasa Jawa di setiap wilayah tersebut memiliki variasi khas yang disebut dengan dialek.

Dialek adalah variasi bahasa yang berbeda-beda menurut pemakainya, misalnya bahasa dari suatu daerah tertentu, kelompok sosial, atau kurun waktu tertentu [414]. Menurut Weijnen dalam Dicta et al. [10], dialek juga dapat didefinisikan sebagai sistem kebahasaan yang dipakai oleh suatu masyarakat untuk membedakannya dari masyarakat lain yang berdekatan, meskipun memiliki hubungan yang erat. Variasi ini bisa muncul karena faktor geografis maupun sosial [12].

Bahasa Jawa sendiri memiliki beragam dialek. Berdasarkan Nothofer (1975), sebagaimana dikutip oleh Hasisah [3], bahasa Jawa dibagi menjadi tiga kelompok dialek utama. Kelompok pertama adalah dialek Jawa daerah barat, yang meliputi

daerah Tegal dan Banyumas. Kelompok kedua merupakan bahasa Jawa daerah tengah, mencakup dialek Yogyakarta, Surakarta, Rembang, Jepara, Semarang, Bagelan, Kedu. Kelompok ketiga berada di wilayah Jawa Timur yang dikenal sebagai dialek Jawa Timuran.

2.2.2. Representasi Audio Digital

Suara di alam adalah sinyal analog, yaitu besaran fisis yang bersifat kontinu dan dapat diukur seiring perubahan waktu atau ruang. Namun, komputer memproses data dalam bentuk digital yang bersifat diskrit [13]. Agar dapat disimpan dan dimanipulasi oleh komputer, sinyal audio analog harus diubah menjadi representasi digital [14]Proses fundamental ini dikenal sebagai Analog-to-Digital Conversion (ADC), yang secara umum terdiri dari dua tahapan utama: sampling dan kuantisasi.

Tahap pertama adalah sampling, atau yang lebih formal disebut diskritisasi (*discretization*). Proses ini bertujuan membagi sinyal kontinu menjadi interval-interval waktu yang sama, di mana setiap interval diwakili oleh sebuah nilai amplitudo yang terukur. Seberapa sering cuplikan ini diambil disebut sebagai frekuensi sampling (*sampling rate*), yang diukur dalam Hertz (Hz). Sebagai contoh, frekuensi sampling 44.000 Hz berarti nilai sinyal analog dibaca dan disimpan setiap interval 0.00002 detik [14].

Tahap kedua adalah kuantisasi (*quantization*). Dalam tahap ini, setiap nilai amplitudo hasil sampling yang bersifat kontinu didekati (diaproksimasi) oleh sebuah nilai dari himpunan diskrit yang terbatas. Proses ini mirip dengan pembulatan angka riil menjadi bilangan bulat [14]. Akurasi dari proses ini ditentukan oleh kedalaman bit (*bit depth*).

Secara ringkas, *sampling* mengubah sinyal dari kontinu menjadi diskrit pada sumbu waktu, sementara kuantisasi melakukan hal yang sama pada sumbu amplitudo. Hasil akhir dari kedua proses ini adalah aliran data biner yang siap untuk diproses oleh sistem digital.

2.2.3. Mel Spectrogram

Mel Spectrogram adalah sebuah fitur yang digunakan dalam pemrosesan audio di mana klip audio diubah menjadi bentuk gambar spektrogram. Secara spesifik, Mel Spectrogram merupakan representasi visual dari spektrum frekuensi sinyal audio [15]. Ini adalah representasi waktu-frekuensi dari suara yang dihasilkan untuk meniru sistem

pendengaran biologis manusia [16].

Proses pembuatannya melibatkan konversi sinyal dari domain waktu ke domain frekuensi. Frekuensi audio (f) kemudian ditransfer ke skala Mel (M(f)) menggunakan transformasi nonlinier [16]. Rumus matematis untuk konversi ini adalah sebagai berikut [17]:

$$M(f) = 2595 \log_{10}(1 + \frac{f}{700}) \quad (2.1)$$

Hasilnya adalah sebuah gambar di mana komponen magnitudo sinyal diuraikan sesuai dengan frekuensi pada skala Mel. Gambar-gambar ini kemudian dapat digunakan sebagai input untuk model klasifikasi seperti Convolutional Neural Networks (CNN).

2.2.4. *Convolutional Neural Network* (CNN)

Convolutional Neural Network (CNN) adalah bagian dari jaringan saraf yang sangat efektif untuk klasifikasi gambar [16]. Dalam analisis audio, CNN bekerja dengan mengolah spektrogram—representasi visual dari suara—sebagai sebuah gambar [18]. Proses kerjanya dimulai dengan mengubah sinyal audio menjadi gambar spektrogram. Gambar ini kemudian dimasukkan ke model CNN yang telah dilatih untuk mengekstraksi fitur-fitur penting secara otomatis. Kinerja model dievaluasi menggunakan data uji untuk mengukur akurasinya dalam melakukan klasifikasi.

Arsitektur CNN secara hierarkis memproses gambar spektrogram melalui beberapa lapisan inti. Setelah Lapisan Input menerima gambar , Lapisan Konvolusi (CONV) menggunakan serangkaian filter (kernel) yang dapat dipelajari untuk memindai gambar dan membuat peta fitur (feature maps), yang secara efektif menarik keluar pola-pola penting dari gambar. Setiap lapisan konvolusi diikuti oleh Lapisan Aktivasi non-linear (seperti ReLU) untuk meningkatkan kemampuan belajar jaringan. Selanjutnya, Lapisan Pooling (misalnya max pooling) mengurangi ukuran spasial data untuk menekan kompleksitas komputasi dan risiko overfitting. Terakhir, Lapisan Terhubung Penuh (FC) memanfaatkan fitur-fitur yang telah diekstraksi untuk melakukan klasifikasi akhir , dengan Lapisan Output (umumnya menggunakan Softmax) yang menghasilkan probabilitas untuk setiap kelas [16][18].

2.2.5. Metrik Evaluasi Kinerja

Tahap evaluasi kinerja model klasifikasi bertujuan untuk mengukur kualitas prediksi sebuah model secara kuantitatif. Metrik-metrik evaluasi ini umumnya

dihitung berdasarkan empat kemungkinan hasil prediksi yang dirangkum dalam sebuah Confusion Matrix. Bagian ini akan menguraikan Confusion Matrix sebagai dasar, diikuti dengan cara perhitungan dan interpretasi metrik Akurasi, Presisi, Recall, dan F1-Score.

Confusion matrix adalah sebuah tabel yang merangkum hasil prediksi, dimana setiap sampel akan masuk ke dalam salah satu dari empat kategori berikut:

- True Positive (TP): Jumlah data yang kondisi aktualnya positif dan diprediksi dengan benar sebagai positif oleh model.
- True Negative (TN): Jumlah data yang kondisi aktualnya negatif dan diprediksi dengan benar sebagai negatif oleh model.
- False Positive (FP): Jumlah data yang kondisi aktualnya negatif namun salah diprediksi sebagai positif. Kesalahan ini juga dikenal sebagai "Type I Error".
- False Negative (FN): Jumlah data yang kondisi aktualnya positif namun salah diprediksi sebagai negatif. Kesalahan ini dikenal sebagai "Type II Error" [19].

Keempat komponen dasar dari confusion matrix ini—yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN)—kemudian menjadi dasar untuk menghitung metrik evaluasi kinerja yang lebih spesifik. Metrik-metrik seperti Akurasi, Presisi, Recall, dan F1-Score diturunkan dari nilai-nilai ini untuk memberikan gambaran yang lebih komprehensif tentang performa model.

1. Akurasi (Accuracy)

   Definisi: Akurasi merepresentasikan rasio antara jumlah instans yang diprediksi dengan benar (TP + TN) terhadap jumlah keseluruhan instans dalam dataset [19][20].

   Cara Menghitung:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Presisi (Precision)

   Definisi: Presisi dihitung sebagai jumlah prediksi positif yang benar (TP) dibagi dengan jumlah total prediksi positif (TP + FP) [20].

   Cara Menghitung:

$$Presisi = \frac{TP}{TP + FP}$$

3. Recall (Sensitivity atau True Positive Rate)

Definisi: Recall dihitung sebagai jumlah prediksi positif yang benar (TP) dibagi dengan jumlah total data yang sebenarnya positif (TP + FN). Metrik ini juga dikenal dengan nama Sensitivity atau TP Rate [20].

Cara Menghitung:

$$Recall = \frac{TP}{TP + FN}$$

4. F1-Score (F-Measure)

Definisi: F1-Score adalah rata-rata harmonik (harmonic mean) dari Presisi dan Recall, yang berfungsi sebagai ukuran akurasi sebuah pengujian. Metrik ini menggabungkan Presisi dan Recall menjadi satu angka tunggal [20].

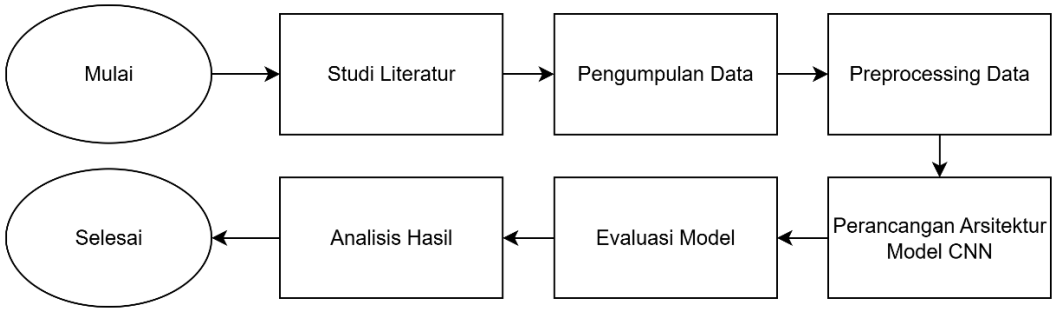
Cara Menghitung:

$$F1 - Score = 2\,x\,\frac{Presisi\ x\ Recall}{Presisi + Recall}$$

BAB III
DESAIN DAN IMPLEMENTASI SISTEM

3.1. Tahapan Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan yang terstruktur untuk memastikan alur kerja yang logis dan sistematis. Secara garis besar, tahapan-tahapan tersebut digambarkan dalam diagram alur pada Gambar 3.1.



Gambar 3. 1 Tahapan Penelitian

3.2. Pengumpulan Data

Kualitas dan kuantitas data merupakan fondasi utama dalam penelitian berbasis machine learning. Berikut adalah rincian mengenai data yang digunakan:

3.2.1. Sumber Data

Data audio yang digunakan dalam penelitian ini merupakan rekaman suara dari penutur asli dialek Jawa yang bersumber dari data primer yang dikumpulkan melalui orang terdekat.

3.2.2. Subjek Penelitian

Dialek yang Digunakan: Penelitian ini berfokus pada klasifikasi tiga dialek utama di Jawa, yaitu:

- Dialek Arekan (representasi wilayah Surabaya dan sekitarnya).
- Dialek Mataraman (representasi wilayah Solo).
- Dialek Banyumasan (representasi wilayah Cilacap).

3.2.3. Karakteristik Penutur

Penutur merupakan penutur asli (native speaker) dari masing-masing dialek untuk memastikan keaslian data. Data diambil dari penutur dengan rentang usia 15-60 tahun dengan distribusi jenis kelamin yang seimbang.

3.3. Preprocessing Data

Data audio mentah mengandung informasi yang kompleks dan tidak terstruktur sehingga tidak dapat langsung diolah oleh model Convolutional Neural Network (CNN). Oleh karena itu, serangkaian tahapan pra-pemrosesan dan ekstraksi fitur diperlukan untuk mentransformasi data audio menjadi representasi yang informatif dan siap dianalisis.

Proses awal yang dilakukan adalah segmentasi, di mana setiap rekaman audio dipotong menjadi klip-klip yang lebih pendek dengan durasi seragam. Standardisasi durasi ini krusial untuk memastikan setiap sampel data memiliki dimensi yang konsisten. Selanjutnya, setiap klip audio dinormalisasi untuk menyamakan tingkat amplitudonya, sehingga model tidak bias terhadap variasi volume rekaman. Bagian audio yang tidak mengandung ucapan (hening) di awal dan akhir klip juga dihilangkan untuk memfokuskan analisis hanya pada sinyal suara yang relevan.

Untuk menangkap karakteristik akustik yang paling membedakan antar dialek, penelitian ini memanfaatkan Mel Spectrogram sebagai fitur utama. Teknik ini dipilih karena kemampuannya yang unggul dalam merepresentasikan suara. Tidak seperti spectrogram linear biasa, skala frekuensi Mel meniru cara kerja sistem pendengaran manusia yang lebih sensitif terhadap perubahan frekuensi rendah. Hal ini menjadikannya sangat efektif untuk mengekstrak fitur-fitur penting dalam ucapan manusia. Proses ekstraksi ini menghasilkan sebuah "gambar" suara dua dimensi untuk setiap klip audio.

3.4. Arsitektur Model Convolutional Neural Network (CNN)

Arsitektur model yang diusulkan akan dibangun berdasarkan prinsip-prinsip desain CNN yang umum untuk tugas klasifikasi gambar. Secara konseptual, model akan terdiri dari dua bagian utama: blok ekstraksi fitur dan blok klasifikasi.

1. Blok Ekstraksi Fitur: Bagian ini akan terdiri dari beberapa tumpukan lapisan konvolusi (Conv2D) dan pooling (MaxPooling2D).
    - Lapisan Konvolusi: Bertugas sebagai detektor fitur utama. Lapisan-lapisan ini akan menggunakan serangkaian filter (kernel) untuk memindai Mel Spectrogram dan mengidentifikasi pola-pola lokal seperti bentuk formant, tekstur, dan gradien frekuensi. Jumlah filter pada lapisan yang lebih dalam akan cenderung ditingkatkan untuk menangkap fitur yang lebih kompleks dan abstrak.

- Lapisan Pooling: Ditempatkan setelah lapisan konvolusi untuk mereduksi dimensi spasial dari peta fitur. Proses ini membantu membuat representasi fitur lebih robust terhadap variasi posisi dan mengurangi beban komputasi.
- Untuk menjaga agar model tidak mengalami *overfitting*, teknik regularisasi seperti Dropout akan disisipkan di antara lapisan-lapisan ini.

2. Blok Klasifikasi: Setelah fitur-fitur penting diekstraksi, peta fitur akhir akan diratakan (Flatten) menjadi sebuah vektor satu dimensi. Vektor ini kemudian akan menjadi input bagi blok klasifikasi yang terdiri dari satu atau lebih lapisan *fully connected* (Dense). Lapisan-lapisan ini berfungsi untuk memetakan fitur-fitur yang telah diekstraksi ke kelas-kelas dialek yang ada. Lapisan output terakhir akan menggunakan fungsi aktivasi Softmax untuk menghasilkan probabilitas keanggotaan setiap sampel data pada masing-masing kelas dialek.

Desain arsitektur yang lebih detail, termasuk penentuan jumlah lapisan, jumlah filter, dan *hyperparameter* lainnya, akan disempurnakan melalui serangkaian eksperimen selama tahap implementasi penelitian untuk mendapatkan performa yang optimal.

3.5. Metrik Evaluasi

Akurasi digunakan sebagai metrik utama untuk melihat persentase prediksi yang benar secara keseluruhan. Namun, untuk memahami performa model pada setiap kelas dialek secara lebih mendalam, metrik Presisi, Recall, dan F1-Score juga dihitung. Metrik-metrik ini memberikan wawasan tentang seberapa baik model dapat menghindari prediksi positif palsu (presisi) dan seberapa baik ia dapat mengidentifikasi semua sampel positif yang sebenarnya (recall). Lebih lanjut, Confusion Matrix akan dibuat untuk memvisualisasikan secara detail kesalahan klasifikasi yang dilakukan model, sehingga dapat dianalisis dialek mana yang paling sering tertukar.

DAFTAR PUSTAKA

[1] "PROFIL SUKU DAN KERAGAMAN BAHASA DAERAH HASIL LONG FORM SENSUS PENDUDUK 2020."

[2] RIZAL HAYADI, "PENGARUH DIALEK BAHASA SERAWAI TERHADAP PEMAHAMAN SISWA DALAM PEMBELAJARAN IPS DI SD NEGERI 31 BENGKULU SELATAN," UNIVERSITAS ISLAM NEGERI FATMAWATI SUKARNO BENGKULU, 2022.

[3] S. N. Hasisah and M. Suryadi, "VARIASI PEMAKAIAN BAHASA JAWA DIALEK REMBANG PADA MASYARAKAT PEDESAAN: KAJIAN SOSIODIALEKTOLOGI," *MEDAN MAKNA: Jurnal Ilmu Kebahasaan dan Kesastraan*, vol. 20, no. 1, p. 24, Oct. 2022, doi: 10.26499/mm.v20i1.3912.

[4] O. VINTONIAK, M. HNATYUK, R. MINIAILO, O. TURYSHEVA, and V. KOTVYTSKA, "DIALECTOLOGY IN MODERN LINGUISTIC RESEARCH: THEORETICAL APPROACHES AND METHODS," *AD ALTA: Journal of Interdisciplinary Research*, vol. 14, no. 1, Jan. 2024, doi: 10.33543/1401393944.

[5] K. M. O. Nahar, O. M. Al-Hazaimeh, A. Abu-Ein, and M. A. Al-Betar, "Arabic Dialect Identification Using Different Machine Learning Methods," Jun. 2022, doi: 10.21203/rs.3.rs-1726491/v1.

[6] F. M. Fauzi, L. Hayat, D. Nova, and K. Hardani, "Pengenalan Dialek Bahasa Daerah di Pulau Jawa menggunakan Metode Mel-Frequency Cepstral Coefficients dan Adaptive Network-based Fuzzy Inference System," *JURNAL RISET REKAYASA ELEKTRO*, vol. 4, no. 2, pp. 39–50, 2022, [Online]. Available: http://jurnalnasional.ump.ac.id/index.php/JRRE

[7] M. Rizki Putra, B. Yudho Suprapto, and dan Suci Dwijayanti, "PENGENALAN DIALEK DI SUMATERA SELATAN MENGGUNAKAN ALGORITMA DEEP NEURAL NETWORK," *Applicable Innovation of Engineering and Science Research (AVoER)*, pp. 210–217, 2021.

[8] A. H. Setianingrum, K. Hulliyah, and M. F. Amrilla, "Speech Recognition of Sundanese Dialect Using Convolutional Neural Network Method with Mel-Spectrogram Feature Extraction," in *2023 11th International Conference on*

Cyber and IT Service Management, CITSM 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/CITSM60085.2023.10455447.

[9] N. B. Chittaragi and S. G. Koolagudi, "Automatic dialect identification system for Kannada language using single and ensemble SVM algorithms," *Lang Resour Eval*, vol. 54, no. 2, pp. 553–585, Jun. 2020, doi: 10.1007/s10579-019-09481-5.

[10] P. N. Dicta, Z. Rafli, and S. Ansoriyah, "Perbandingan Leksikon Bahasa Jawa Dialek Malang dan Bahasa Jawa Dialek Blitar," *Jurnal Bastrindo*, vol. 2, no. 2, 2021.

[11] D. Sugianto and I. Mufidah, "STATISTIK KEBAHASAAN DAN KESASTRAAN 2024," 2024.

[12] F. Simanjuntak, "VARIASI BAHASA DIALEK MELAYU DI KECAMATAN PANAI HILLIR DAN KECAMATAN PANAI TENGAH," *Pediaqu: Jurnal Pendidikan Sosial dan Humaniora*, vol. 2, no. 3, 2023.

[13] N. Najah Ulfah, E. Saragih, and R. Samuel Fransisco Sinaga, "Pengolahan dan Pemrosesan Sinyal Digital," 2025.

[14] R. Dastres and M. Soori, "A Review in Advanced Digital Signal Processing Systems." [Online]. Available: https://hal.science/hal-03183633v1

[15] Z. Mushtaq, S. F. Su, and Q. V. Tran, "Spectral images based environmental sound classification using CNN with meaningful data augmentation," *Applied Acoustics*, vol. 172, Jan. 2021, doi: 10.1016/j.apacoust.2020.107581.

[16] D. Joshi, J. Pareek, and P. Ambatkar, "Comparative Study of Mfcc and Mel Spectrogram for Raga Classification Using CNN," *Indian J Sci Technol*, vol. 16, no. 11, pp. 816–822, Mar. 2023, doi: 10.17485/IJST/v16i11.1809.

[17] M. Lesnichaia, V. Mikhailava, N. Bogach, I. Lezhenin, J. Blake, and E. Pyshkin, "Classification of Accented English Using CNN Model Trained on Amplitude Mel-Spectrograms," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, International Speech Communication Association, 2022, pp. 3669–3673. doi: 10.21437/Interspeech.2022-462.

[18] L. Nanni, G. Maguolo, S. Brahnam, and M. Paci, "An ensemble of convolutional neural networks for audio classification," *Applied Sciences (Switzerland)*, vol. 11, no. 13, Jul. 2021, doi: 10.3390/app11135796.

[19] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” *BMC Genomics*, vol. 21, no. 1, Jan. 2020, doi: 10.1186/s12864-019-6413-7.

[20] Ž. Vujović, “Classification Model Evaluation Metrics,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.