

Klasifikasi Pengucapan Huruf Hijaiyah Berbasis Android Menggunakan CNN dengan Fitur Mel-Spectrogram

Fawwaz Ijlal Muqsith*, Endang Supriyati, Tri Listyorini
Teknik Informatika, Universitas Muria Kudus, Kudus, 59327, Indonesia

Info Artikel

Riwayat Artikel:

Received 2024-12-23

Revised 2024-12-31

Accepted 2025-01-01

Corresponding Author:

Fawwaz Ijlal Muqsith

Email:

fawwazijlalmuqsith@gmail.com



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstract – Mastery of Hijaiyah letters is a fundamental basis in learning the Qur'an, but data from the IIQ Community Service Institute 2021/2022 shows that 72.25% of the 3,111 Muslims tested have not been able to read the Qur'an properly. This research aims to develop an Android-based Hijaiyah letter pronunciation classification system using Convolutional Neural Network (CNN) with mel-spectrogram features. The research methodology includes collecting 8,904 voice samples from 53 participants at Pondok Tahfidz Yanbu'ul Qur'an Menawan, preprocessing data using MFCC techniques, developing CNN models, and implementing the system in the form of mobile applications with MVVM architecture. The test results showed promising performance with some classes achieving 100% accuracy and an average overall accuracy of 83.80%, although there were challenges in some classes such as "alif_dommah" and "ghaiin_dommah" which had an accuracy below 40%. The developed system successfully provides an interactive learning platform through the integration of mobile applications with the Flask API, but still requires further development, especially in expanding the dataset to overcome overfitting problems and improve the generalization ability of the model.

Keywords: Android, Audio Classification, CNN, Hijaiyah Letters, Mel-spectrogram.

Abstrak – Penguasaan huruf Hijaiyah merupakan dasar fundamental dalam pembelajaran Al-Qur'an, namun data dari Lembaga Pengabdian kepada Masyarakat IIQ 2021/2022 menunjukkan 72,25% dari 3.111 muslim yang diuji belum mampu membaca Al-Qur'an dengan baik. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi pengucapan huruf Hijaiyah berbasis Android menggunakan Convolutional Neural Network (CNN) dengan fitur mel-spectrogram. Metodologi penelitian meliputi pengumpulan 8.904 sampel suara dari 53 partisipan di Pondok Tahfidz Yanbu'ul Qur'an Menawan, preprocessing data menggunakan teknik MFCC, pengembangan model CNN, dan implementasi sistem dalam bentuk aplikasi mobile dengan arsitektur MVVM. Hasil pengujian menunjukkan performa yang menjanjikan dengan beberapa kelas mencapai akurasi 100% dan rata-rata akurasi keseluruhan 83,80%, meskipun terdapat tantangan pada beberapa kelas seperti "alif_dommah" dan "ghaiin_dommah" yang memiliki akurasi di bawah 40%. Sistem yang dikembangkan berhasil memberikan platform pembelajaran yang interaktif melalui integrasi aplikasi mobile dengan API Flask, namun masih memerlukan pengembangan lebih lanjut terutama dalam perluasan dataset untuk mengatasi masalah overfitting dan meningkatkan kemampuan generalisasi model.

Kata Kunci: Android, CNN, Huruf Hijaiyah, Klasifikasi Audio, Mel-spectrogram.

I. PENDAHULUAN

Penguasaan huruf Hijaiyah merupakan dasar fundamental dalam pembelajaran Al-Qur'an, yang menjadi keterampilan esensial bagi setiap muslim untuk dapat membaca kitab suci dengan benar dan sesuai kaidah. Pengenalan dan penguasaan huruf hijaiyah tidak hanya sebatas mengenal bentuk hurufnya saja, tetapi juga mencakup kemampuan melafalkan setiap huruf dengan tepat sesuai makhraj (tempat keluarnya huruf) dan karakteristik bunyinya ketika diberi harokat yang berbeda (fathah, kasrah, dan dhommah). Ketepatan dalam pengucapan ini sangat krusial karena kesalahan pelafalan dapat mengubah makna dan mengurangi keabsahan bacaan Al-Qur'an.

Berdasarkan data dari Lembaga Pengabdian kepada Masyarakat IIQ 2021/2022, dari 3.111 muslim yang diuji, sebesar 72,25% terkategori belum mampu membaca Al-Qur'an dengan baik [1]. Kondisi yang mengkhawatirkan ini juga ditemukan berdasarkan pengambilan data di MTS dan MA Pondok Tahfidz Yanbu'ul Qur'an Menawan, dimana masih banyak remaja yang kesulitan dalam melafalkan huruf Hijaiyah dengan benar, bahkan dalam lingkungan pondok pesantren sekalipun. Permasalahan ini tidak berhenti pada masa remaja, tetapi terbawa hingga dewasa dan memunculkan rasa malu untuk memulai kembali pembelajaran. Kondisi ini menciptakan siklus yang tidak menguntungkan, di mana kesulitan yang tidak teratasi pada masa remaja berkembang menjadi hambatan psikologis pada usia dewasa, sehingga banyak yang akhirnya enggan untuk memperbaiki kemampuan mereka dalam melafalkan huruf Hijaiyah, padahal kemampuan ini sangat krusial dalam membaca Al-Qur'an.

Penelitian ini bertujuan untuk mengembangkan sebuah sistem pembelajaran mandiri berbasis teknologi yang dapat membantu muslimin dan muslimah dalam mempelajari dan memperbaiki pengucapan huruf hijaiyah secara tepat sesuai dengan harokatnya. Melalui pengembangan aplikasi mobile yang mengintegrasikan teknologi Convolutional Neural Network (CNN) dengan fitur mel-spectrogram, sistem ini dirancang untuk dapat memberikan umpan balik langsung mengenai ketepatan pelafalan huruf hijaiyah. Tujuan utamanya adalah memungkinkan pengguna untuk berlatih secara mandiri, kapan saja dan di mana saja, tanpa terkendala rasa malu atau keterbatasan akses ke guru/pembimbing, sehingga dapat meningkatkan kemampuan membaca Al-Qur'an dengan baik dan benar.

Perkembangan teknologi modern, khususnya dalam bidang kecerdasan buatan, membuka peluang untuk mengatasi permasalahan ini melalui pendekatan pembelajaran yang lebih personal dan mandiri. Teknologi klasifikasi suara menggunakan Convolutional Neural Network (CNN) telah menunjukkan keberhasilan dalam berbagai aplikasi pengenalan suara. Beberapa penelitian terdahulu telah membuktikan efektivitas CNN dalam klasifikasi suara, seperti penelitian Fadhilah Gusti Safinatunnajah [2] yang berhasil mengklasifikasikan suara kucing dengan akurasi yang baik menggunakan kombinasi CNN dan LSTM. Selain itu, Vita Karenina [3] juga berhasil mengembangkan sistem deteksi suara manusia berdasarkan jenis kelamin dan rentang usia dengan tingkat akurasi mencapai 92%.

Meskipun teknologi klasifikasi suara telah banyak diterapkan dalam berbagai bidang, penerapannya dalam pembelajaran huruf Hijaiyah masih terbatas. Sebagian besar aplikasi pembelajaran Al-Qur'an yang ada saat ini masih berfokus pada metode konvensional seperti pengenalan visual dan audio playback, tanpa adanya fitur evaluasi pelafalan secara real-time. Penelitian ini bertujuan untuk mengembangkan aplikasi mobile yang dapat mengklasifikasikan pelafalan huruf Hijaiyah secara otomatis menggunakan model CNN. Sistem ini akan memberikan umpan balik langsung kepada pengguna mengenai ketepatan pelafalan mereka, sehingga memungkinkan proses pembelajaran mandiri yang lebih efektif.

Kebaruan dari penelitian ini terletak pada integrasi teknologi CNN dengan aplikasi mobile untuk pembelajaran huruf Hijaiyah, yang memungkinkan evaluasi pelafalan secara real-time dan pemberian umpan balik langsung kepada pengguna. Penggunaan mel-spectrogram sebagai metode ekstraksi fitur audio, dikombinasikan dengan arsitektur CNN yang dioptimalkan untuk pengenalan huruf Hijaiyah, diharapkan dapat memberikan tingkat akurasi yang tinggi dalam klasifikasi pelafalan. Implementasi sistem ini diharapkan dapat menjadi solusi efektif dalam meningkatkan kemampuan membaca Al-Qur'an di kalangan masyarakat muslim, terutama bagi mereka yang memiliki keterbatasan waktu atau merasa tidak nyaman belajar secara konvensional.

II. METODE

Penelitian ini menggunakan pendekatan Speech Recognition dan Audio Classification dengan memanfaatkan teknik Deep Learning, khususnya Convolutional Neural Network (CNN), untuk mengklasifikasikan pengucapan huruf hijaiyah. Convolutional Neural Network (CNN) merupakan salah satu metode deep learning yang dapat digunakan untuk klasifikasi audio dengan cara memanfaatkan spektrogram mel sebagai input data [4]. Metodologi yang diterapkan mengikuti prinsip-prinsip pemrosesan sinyal digital dan pembelajaran mesin yang telah mapan dalam bidang pengenalan suara.

Proses penelitian melibatkan beberapa tahapan utama yang sistematis, dimulai dari pengumpulan data suara secara langsung dari para santri. Data audio yang direkam kemudian melalui tahap preprocessing untuk mengoptimalkan kualitas sinyal. Ekstraksi fitur dilakukan menggunakan Mel-frequency Cepstral Coefficients (MFCC), sebuah teknik ekstraksi fitur yang meniru sistem pendengaran manusia dengan mengubah sinyal audio menjadi koefisien-koefisien yang merepresentasikan karakteristik spektral suara melalui serangkaian proses transformasi, termasuk Fourier Transform dan filter mel-scale yang telah terbukti efektif dalam aplikasi speech recognition dan telah banyak digunakan dalam penelitian-penelitian sebelumnya [5].

Arsitektur CNN yang digunakan dalam penelitian ini dirancang khusus untuk menangani karakteristik spektral dari data audio. Model ini dilatih menggunakan dataset yang telah dipersiapkan, dengan pembagian data menjadi set pelatihan dan pengujian. Evaluasi performa model dilakukan dengan menggunakan berbagai metrik standar seperti akurasi, presisi, recall, dan F1-score.

Bagian berikut akan menjelaskan secara detail setiap tahapan dalam metodologi penelitian, mencakup proses pengumpulan data, preprocessing, ekstraksi fitur, arsitektur model, proses pelatihan, evaluasi, hingga pembuatan aplikasi. Penjelasan akan dilengkapi dengan diagram alur kerja untuk memberikan gambaran yang komprehensif tentang keseluruhan sistem yang dikembangkan.



Gambar 1. Langkah - langkah pengembangan model klasifikasi suara huruf hijaiyah

A. Pengumpulan Data

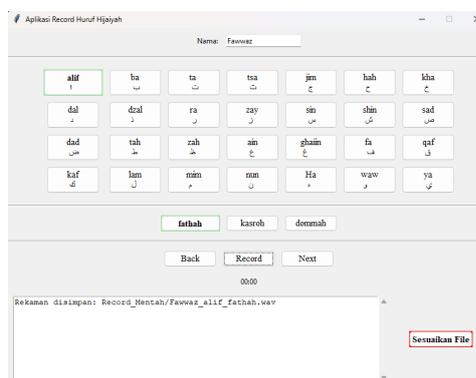
Tahap pengumpulan data menjadi langkah krusial dalam penelitian ini. Proses ini dilaksanakan di Pondok Tahfidz Yanbuul Quran Menawan yang berlokasi di Kudus. Untuk memastikan kualitas data yang diperoleh, telah dilakukan proses seleksi terhadap 53 orang berdasarkan ketepatan pelafalan makhorijul huruf mereka. Hanya individu-individu dengan kemampuan pelafalan yang baik yang dipilih untuk berpartisipasi dalam tahap pengumpulan data.

Proses perekaman data dilakukan dua kali untuk setiap partisipan. Setiap individu diminta untuk memberikan sampel suara yang mencakup 84 jenis pelafalan huruf hijaiyah, dimulai dari alif_fathah hingga ya_dommah. Dengan demikian, setiap partisipan menghasilkan total 168 rekaman suara yang siap diolah lebih lanjut dalam penelitian ini.

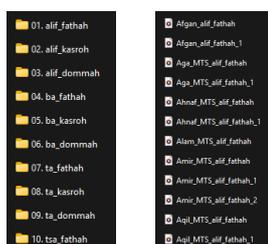
Untuk memastikan konsistensi dan kualitas data audio yang dihasilkan, digunakan beberapa settingan teknis yang seragam dalam proses perekaman. Format audio yang dipilih adalah pyaudio.paInt16, yang merupakan format integer 16-bit. Format ini memberikan kualitas audio yang baik dengan ukuran file yang masih terkendali. Selain itu, perekaman dilakukan dengan menggunakan mono channel (single channel) yang cocok untuk pengambilan suara atau ucapan tanpa memerlukan efek stereo.

Ukuran buffer yang digunakan adalah 1024 sampel per frame, yang merupakan nilai umum untuk memberikan latency yang baik dalam perekaman. Sample rate yang digunakan adalah 16000 Hz, Dalam penelitian [6], sample rate 16000Hz terbukti cukup baik dan dapat diadaptasi untuk mengaplikasikan model CNN pada klasifikasi audio di beragam domain seperti musik, suara, dan lingkungan akustik. Terakhir, format file output yang dipilih adalah WAV (.wav), dengan memanfaatkan modul wave untuk menyimpan file. Sample width disesuaikan dengan format input yaitu 16-bit, sehingga konsisten dengan settingan awal perekaman.

Dengan menerapkan settingan teknis yang tepat dan konsisten ini, diharapkan data audio yang dihasilkan memiliki kualitas yang baik dan seragam. Hal ini akan mendukung keberhasilan tahapan-tahapan selanjutnya dalam penelitian klasifikasi pelafalan huruf hijaiyah ini menggunakan pendekatan CNN. Aplikasi yang disiapkan untuk melakukan record dan otomatis akan melakukan pelabelan kelasnya ditunjukkan pada Gambar 3, sedangkan untuk contoh label kelas dan isinya ditunjukkan pada Gambar 4(a) dan Gambar 4(b).



Gambar 2. Aplikasi Record



Gambar 3. Pelabelan Otomatis (a) Nama Label per Kelas, (b) Hasil Record per Kelas

B. Pembagian Data

Proses pengumpulan data yang dilakukan menghasilkan total 8.904 sampel audio yang terdiri dari 84 kelas pelafalan huruf hijaiyah berbeda (dari alif_fathah hingga ya_dommah), dimana setiap kelas memiliki 106 sampel suara. Data ini diperoleh dari hasil perekaman 53 partisipan yang masing-masing melafalkan dua kali untuk setiap jenis huruf. Untuk memastikan kualitas pembelajaran model, dataset kemudian dibagi menjadi dua bagian utama dengan menggunakan fungsi `train_test_split` dari library `sklearn` dengan pemisahan data dengan rasio 80:20 untuk data latih dan data uji merupakan pendekatan yang umum dan efektif dalam proses pemodelan machine learning, dimana 80% data digunakan untuk proses pelatihan model dan 20% sisanya digunakan untuk menguji performa model [7], serta `random_state=2` untuk menjamin konsistensi pembagian data.

C. Pembuatan Model

Dalam tahap pembuatan model, arsitektur Convolutional Neural Network (CNN) dirancang khusus untuk mengklasifikasikan pelafalan huruf hijaiyah berdasarkan data spektrogram mel. Model CNN terdiri dari beberapa lapisan yang disusun berurutan, di mana setiap lapisan memiliki peran dan fungsi berbeda dalam ekstraksi fitur dan klasifikasi.

Input Layer menerima data spektrogram mel dengan bentuk (128, 128, 1) dan menentukan dimensi data yang akan diproses. Convolutional Layer 1, 2, dan 3 mengekstraksi fitur-fitur dari spektrogram mel dengan jumlah filter yang meningkat, yaitu 32, 64, dan 128. Filter-filter ini melakukan operasi konvolusi pada input dan menghasilkan peta fitur yang merepresentasikan keberadaan fitur-fitur tertentu. Fungsi aktivasi ReLU digunakan untuk menambahkan non-linearitas pada output konvolusi.

Max Pooling Layer 1, 2, dan 3 diterapkan setelah setiap Convolutional Layer untuk mengurangi dimensi spasial dari output dan mempertahankan fitur-fitur penting. Flatten Layer mengubah output dari Max Pooling Layer 3 menjadi vektor 1 dimensi agar dapat diproses oleh Dense Layer. Dense Layer 1 menggabungkan fitur-fitur yang telah diekstraksi dengan 64 unit neuron dan fungsi aktivasi ReLU. Output Layer menghasilkan prediksi probabilitas pelafalan huruf hijaiyah dengan 1 unit neuron dan fungsi aktivasi sigmoid.

Model CNN ini mampu mengekstraksi fitur-fitur penting dari spektrogram mel dalam dataset yang digunakan dan menggunakannya untuk mengklasifikasikan audio dengan baik [8]. Arsitektur model dirancang secara khusus untuk menangani data audio dan memberikan hasil yang optimal dalam konteks penelitian ini.

Model dikompilasi dengan optimizer Adam, fungsi loss binary cross-entropy, dan metrik akurasi untuk memantau performa selama pelatihan dan evaluasi. Dengan arsitektur model CNN yang telah dibangun, penelitian dapat dilanjutkan ke tahapan pelatihan model menggunakan dataset yang telah disiapkan dan evaluasi performa model dalam mengklasifikasikan pelafalan huruf hijaiyah.

D. Ekstraksi Fitur MFCC

MFCC merupakan algoritma ekstraksi ciri suara yang dirancang untuk mengubah karakteristik suara menjadi serangkaian koefisien yang merepresentasikan ciri suara [9]. Dalam implementasinya pada penelitian ini, sistem menggunakan fungsi `librosa` untuk mengekstrak fitur mel-spectrogram dari setiap sampel audio huruf hijaiyah yang direkam. Proses ekstraksi menggunakan parameter `n_mels=128` untuk mendapatkan resolusi frekuensi yang optimal, `n_fft=2048` untuk ukuran jendela analisis Fast Fourier Transform, dan `hop_length=512` untuk menentukan jumlah sampel antara frame yang berurutan. Hasil ekstraksi kemudian dinormalisasi ke dalam rentang nilai yang konsisten menggunakan konversi `power-to-db` untuk menyesuaikan dengan karakteristik pendengaran manusia.

E. Pelatihan Model

Pelatihan model CNN dilakukan dengan membagi dataset menjadi data latih (80%) dan data uji (20%). Data latih digunakan untuk melatih model agar dapat mempelajari pola dari setiap kelas huruf hijaiyah, sedangkan data uji dipakai untuk mengevaluasi performa model pada data baru.

Selama pelatihan, diterapkan teknik `early stopping` menggunakan callback `CustomEarlyStopping`. Tujuannya adalah mencegah `overfitting` dengan menghentikan pelatihan saat nilai kerugian validasi mencapai 0.15 dan akurasi validasi melampaui 96%. Penerapan `early stopping` seperti ini membantu model menjaga kemampuan generalisasinya saat berhadapan dengan data baru [10].

F. Evaluasi Model

Untuk mengevaluasi kinerja dan kehandalan model CNN yang dikembangkan, kami menggunakan beberapa metrik standar seperti akurasi, presisi, recall, dan F1-score [11]. Selain itu, visualisasi kurva akurasi/loss selama pelatihan juga digunakan untuk mendapatkan wawasan yang lebih dalam tentang perilaku model dan mengidentifikasi area-area yang perlu perbaikan lebih lanjut.

G. Pengujian Model

Setelah model CNN selesai dilatih, tahap selanjutnya adalah melakukan pengujian untuk mengevaluasi kemampuan model dalam mengklasifikasikan pengucapan huruf hijaiyah menggunakan data uji yang telah disiapkan sebanyak 20% dari total dataset. Proses pengujian ini bertujuan untuk mengukur performa model ketika berhadapan dengan data baru yang belum pernah dilihat sebelumnya. Dalam konteks klasifikasi pengucapan huruf hijaiyah, nilai akurasi menjadi indikator penting yang menunjukkan tingkat ketepatan model dalam mengenali dan membedakan berbagai variasi pengucapan. Semakin tinggi akurasi yang dicapai, semakin baik kemampuan model dalam melakukan klasifikasi pengucapan huruf hijaiyah dengan benar [12]. Evaluasi performa tidak hanya dilakukan melalui pengukuran akurasi, tetapi juga mempertimbangkan nilai loss untuk mendapatkan penilaian yang lebih komprehensif tentang kualitas model.

H. Pembuatan API

API (Application Programming Interface) merupakan protokol yang memungkinkan aplikasi perangkat lunak untuk saling berkomunikasi dan berbagi data, layaknya sebuah jembatan yang menghubungkan satu program dengan program lainnya untuk saling bertukar informasi dan layanan [13]. Dalam konteks penelitian yang akan dikembangkan, penggunaan Flask sebagai REST API merupakan pilihan yang tepat untuk mengimplementasikan komunikasi antara model Machine Learning dan aplikasi mobile, dimana Flask merupakan web framework Python yang ringan dan fleksibel yang sangat cocok untuk membuat REST API [14] yang akan melayani permintaan dari aplikasi mobile. Dengan Flask, dapat dibuat endpoint API yang akan menerima input audio dari aplikasi mobile, memproses input menggunakan model ML yang telah dilatih, dan mengirimkan hasil klasifikasi kembali ke aplikasi mobile. Arsitektur ini memungkinkan model ML tetap berada di server sehingga mengurangi beban pada perangkat mobile, pembaruan model dapat dilakukan di sisi server tanpa perlu mengupdate aplikasi mobile, dan skalabilitas yang lebih baik karena pemrosesan dilakukan di server. Dalam kasus klasifikasi pengucapan huruf hijaiyah, Flask API dapat menerima file audio dari aplikasi mobile, mengekstrak fitur mel-spectrogram, melakukan prediksi menggunakan model CNN yang telah dilatih, dan mengembalikan hasil klasifikasi ke aplikasi mobile. Pendekatan ini sejalan dengan prinsip microservice yang disebutkan dalam jurnal [15], di mana setiap komponen (aplikasi mobile, API, dan model ML) memiliki tanggung jawab yang terpisah namun dapat berkomunikasi secara efektif melalui API.

I. Pembuatan Aplikasi

Pengembangan aplikasi dilakukan menggunakan Jetpack Compose dengan bahasa pemrograman Kotlin sebagai teknologi utama untuk membangun aplikasi Android secara native [16]. Aplikasi dirancang menggunakan arsitektur MVVM (Model-View-ViewModel) dan Clean Architecture untuk memastikan pemisahan concerns yang baik, dengan tiga layer utama yaitu presentation layer untuk UI menggunakan Compose, domain layer untuk business logic, dan data layer untuk komunikasi dengan API [17]. Fitur utama aplikasi meliputi sistem perekaman audio menggunakan Android AudioRecord API yang diimplementasikan dengan Kotlin Coroutines untuk operasi asynchronous, komunikasi jaringan menggunakan Retrofit untuk integrasi dengan REST API, dan manajemen dependensi menggunakan Hilt. Interface pengguna diimplementasikan dengan mempertimbangkan prinsip Material Design, terdiri dari beberapa screen utama seperti recording screen dengan visualisasi audio untuk feedback visual saat perekaman, dan result screen untuk menampilkan hasil klasifikasi dari API. Flow aplikasi dirancang sederhana dimana pengguna dapat memilih huruf hijaiyah yang akan dilatih, melakukan perekaman suara, dan sistem secara otomatis mengirimkan rekaman ke API service untuk mendapatkan hasil klasifikasi pengucapan.

III. HASIL DAN PEMBAHASAN

Pada penelitian ini, metode Convolutional Neural Network (CNN) digunakan untuk mengklasifikasikan pengucapan huruf hijaiyah dengan memanfaatkan fitur mel-spectrogram dari data suara. Model CNN yang dikembangkan dilatih menggunakan dataset sebanyak 8.904 sampel suara, mencakup 84 kelas pelafalan huruf hijaiyah yang berbeda.

A. CNN Model

Proses pengembangan model ini meliputi beberapa tahapan, mulai dari preprocessing dataset, pelatihan dan evaluasi model, hingga pembuatan API menggunakan Flask untuk integrasi dengan aplikasi mobile.

1) *Preprocessing* : Dalam penelitian ini, tahap pertama dalam preprocessing adalah pembagian dataset menjadi set pelatihan dan set pengujian dengan rasio 80:20, bisa dilihat pada Tabel 1. Pembagian ini dilakukan untuk memastikan bahwa model dapat dilatih pada sebagian besar data yang tersedia, sementara tetap memiliki set terpisah untuk mengevaluasi performa model pada data yang belum pernah dilihat sebelumnya. Fungsi `train_test_split` dari library `sklearn` digunakan untuk melakukan pembagian dataset secara acak dengan `random_state=2` untuk menjamin konsistensi pembagian data.

TABEL 1
JUMLAH DATA

	Jumlah Data
Pelatihan	7.123
Pengujian	1.781
Jumlah Data	8.904

Setelah pembagian dataset, tahap selanjutnya adalah ekstraksi fitur menggunakan teknik Mel-frequency Cepstral Coefficients (MFCC). MFCC merupakan metode yang umum digunakan dalam pemrosesan sinyal suara untuk mengekstraksi fitur yang merepresentasikan karakteristik spektral suara. Dalam penelitian ini, library librosa digunakan untuk menghitung MFCC dari setiap sampel suara dalam dataset. Parameter yang digunakan dalam ekstraksi MFCC meliputi $n_mels=128$ untuk menentukan jumlah filter mel, $n_fft=2048$ untuk menentukan ukuran jendela analisis Fast Fourier Transform, dan $hop_length=512$ untuk menentukan jumlah sampel antara frame yang berurutan. Hasil ekstraksi MFCC kemudian dinormalisasi menggunakan konversi power-to-db untuk menyesuaikan dengan skala logaritmik yang lebih sesuai dengan persepsi pendengaran manusia. Fitur MFCC yang diekstraksi akan menjadi input untuk model CNN dalam tahap pelatihan dan evaluasi.

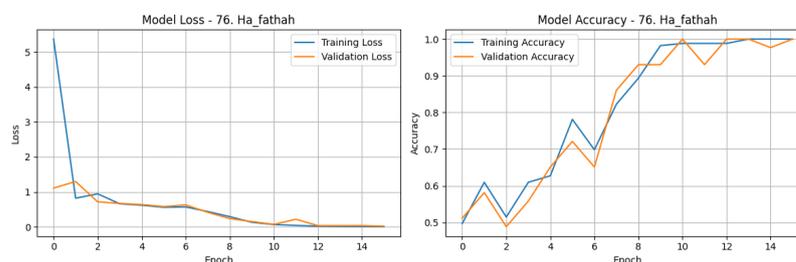
2) *Pelatihan Model* : Proses pelatihan dilakukan dengan menggunakan optimizer Adam dan fungsi loss binary cross-entropy. Untuk mengoptimalkan proses pelatihan dan mencegah overfitting, diterapkan mekanisme early stopping yang dikustomisasi dengan parameter $val_loss_threshold$ sebesar 0.15 dan $val_accuracy_threshold$ sebesar 0.96. Mendapatkan hasil pelatihan yang bisa dilihat pada Tabel 2.

TABEL 2
HASIL PELATIHAN MODEL

Class	Accuracy	Precision	Recall	F1-Score
alif_fathah	72.09%	75.00%	68.18%	71.43%
alif_kasroh	65.12%	64.00%	72.73%	68.09%
alif_dommah	34.88%	35.00%	31.82%	33.33%
ba_fathah	97.67%	100.00%	95.45%	97.67%
ba_kasroh	69.77%	68.00%	77.27%	72.34%
ba_dommah	100.00%	100.00%	100.00%	100.00%
ta_fathah	67.44%	66.67%	72.73%	69.57%
ta_kasroh	46.51%	47.37%	40.91%	43.90%
ta_dommah	79.07%	74.07%	90.91%	81.63%
tsa_fathah	83.72%	85.71%	81.82%	83.72%
tsa_kasroh	67.44%	61.76%	95.45%	75.00%
tsa_dommah	60.47%	66.67%	45.45%	54.05%
jim_fathah	97.67%	100.00%	95.45%	97.67%
jim_kasroh	81.40%	79.17%	86.36%	82.61%
jim_dommah	97.67%	100.00%	95.45%	97.67%
hah_fathah	88.37%	84.00%	95.45%	89.36%
hah_kasroh	83.72%	89.47%	77.27%	82.93%
hah_dommah	72.09%	77.78%	63.64%	70.00%
kha_fathah	90.70%	95.00%	86.36%	90.48%
kha_kasroh	88.37%	94.74%	81.82%	87.80%
kha_dommah	60.47%	77.78%	31.82%	45.16%
dal_fathah	100.00%	100.00%	100.00%	100.00%
dal_kasroh	81.40%	75.00%	95.45%	84.00%
dal_dommah	95.35%	95.45%	95.45%	95.45%
dzal_fathah	100.00%	100.00%	100.00%	100.00%
dzal_kasroh	83.72%	82.61%	86.36%	84.44%
dzal_dommah	97.67%	100.00%	95.45%	97.67%
ra_fathah	97.67%	95.65%	100.00%	97.78%
ra_kasroh	69.77%	68.00%	77.27%	72.34%
ra_dommah	72.09%	75.00%	68.18%	71.43%
zay_fathah	93.02%	100.00%	86.36%	92.68%
zay_kasroh	97.67%	100.00%	95.45%	97.67%
zay_dommah	97.67%	100.00%	95.45%	97.67%
sin_fathah	88.37%	94.74%	81.82%	87.80%
sin_kasroh	93.02%	88.00%	100.00%	93.62%
sin_dommah	97.67%	100.00%	95.45%	97.67%
shin_fathah	100.00%	100.00%	100.00%	100.00%
shin_kasroh	97.67%	95.65%	100.00%	97.78%
shin_dommah	100.00%	100.00%	100.00%	100.00%
sad_fathah	97.67%	95.65%	100.00%	97.78%
sad_kasroh	58.14%	60.00%	54.55%	57.14%
sad_dommah	97.67%	100.00%	95.45%	97.67%
dad_fathah	97.67%	95.65%	100.00%	97.78%
dad_kasroh	74.42%	70.37%	86.36%	77.55%

dad_dommah	97.67%	100.00%	95.45%	97.67%
tah_fathah	95.35%	95.45%	95.45%	95.45%
tah_kasroh	88.37%	84.00%	95.45%	89.36%
tah_dommah	72.09%	77.78%	63.64%	70.00%
zah_fathah	97.67%	95.65%	100.00%	97.78%
zah_kasroh	86.05%	90.00%	81.82%	85.71%
zah_dommah	97.67%	100.00%	95.45%	97.67%
ain_fathah	88.37%	94.74%	81.82%	87.80%
ain_kasroh	97.67%	100.00%	95.45%	97.67%
ain_dommah	58.14%	83.33%	22.73%	35.71%
ghaiin_fathah	97.67%	95.65%	100.00%	97.78%
ghaiin_kasroh	79.07%	84.21%	72.73%	78.05%
ghaiin_dommah	39.53%	35.71%	22.73%	27.78%
fa_fathah	83.72%	89.47%	77.27%	82.93%
fa_kasroh	83.72%	77.78%	95.45%	85.71%
fa_dommah	67.44%	75.00%	54.55%	63.16%
qaf_fathah	97.67%	95.65%	100.00%	97.78%
qaf_kasroh	69.77%	80.00%	54.55%	64.86%
qaf_dommah	55.81%	55.56%	68.18%	61.22%
kaf_fathah	88.37%	100.00%	77.27%	87.18%
kaf_kasroh	69.77%	73.68%	63.64%	68.29%
kaf_dommah	58.14%	56.25%	81.82%	66.67%
lam_fathah	90.70%	87.50%	95.45%	91.30%
lam_kasroh	79.07%	78.26%	81.82%	80.00%
lam_dommah	88.37%	81.48%	100.00%	89.80%
mim_fathah	83.72%	85.71%	81.82%	83.72%
mim_kasroh	83.72%	85.71%	81.82%	83.72%
mim_dommah	81.40%	88.89%	72.73%	80.00%
nun_fathah	88.37%	84.00%	95.45%	89.36%
nun_kasroh	79.07%	78.26%	81.82%	80.00%
nun_dommah	93.02%	88.00%	100.00%	93.62%
Ha_fathah	100.00%	100.00%	100.00%	100.00%
Ha_kasroh	79.07%	76.00%	86.36%	80.85%
Ha_dommah	97.67%	95.65%	100.00%	97.78%
waw_fathah	83.72%	80.00%	90.91%	85.11%
waw_kasroh	88.37%	86.96%	90.91%	88.89%
waw_dommah	60.47%	60.00%	68.18%	63.83%
ya_fathah	90.70%	100.00%	81.82%	90.00%
ya_kasroh	55.81%	60.00%	40.91%	48.65%
ya_dommah	88.37%	84.00%	95.45%	89.36%

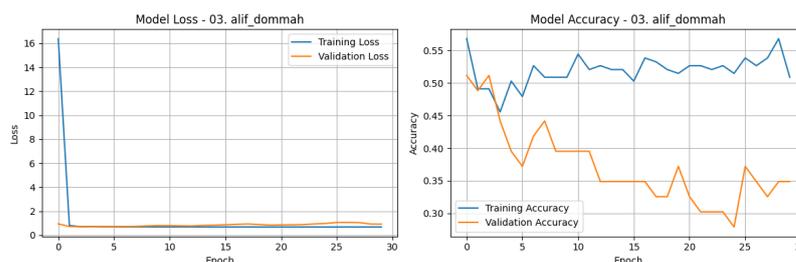
Hasil evaluasi menunjukkan performa yang bervariasi di antara 84 kelas huruf hijaiyah. Beberapa kelas mencapai akurasi sempurna 100%, termasuk "ba_dommah", "dal_fathah", "dzal_fathah", "shin_fathah", "shin_dommah", dan "Ha_fathah". Model menunjukkan presisi dan recall yang seimbang untuk kelas-kelas ini, menghasilkan F1-Score yang juga mencapai 100%. Di sisi lain, beberapa kelas menunjukkan performa yang lebih rendah, seperti "alif_dommah" dengan akurasi 34.88% (presisi 35.00%, recall 31.82%) dan "ghaiin_dommah" dengan akurasi 39.53% (presisi 35.71%, recall 22.73%).



Gambar 4. Grafik Accuracy dan Loss Model 76. Ha_fathah

Analisis grafik pelatihan model untuk kelas "Ha_fathah" bisa dilihat pada Gambar 4. Training loss yang awalnya tinggi (sekitar 5.0) mengalami penurunan drastis dalam 2 epoch pertama, diikuti penurunan gradual hingga konvergen di sekitar 0.1. Akurasi model mengalami peningkatan signifikan setelah epoch ke-6, meningkat dari 70% menjadi 90%, dan akhirnya mencapai akurasi mendekati 100% setelah epoch ke-10. Kedua kurva loss dan accuracy berjalan berdekatan setelah epoch ke-4, menunjukkan model tidak mengalami overfitting yang signifikan dan memiliki kemampuan generalisasi yang baik. Model berhenti pada epoch ke-15 karena mencapai kriteria early stopping dengan validation loss di bawah threshold 0.15 dan validation accuracy di atas 96%. Penghentian ini terjadi pada waktu yang tepat karena kurva akurasi dan loss sudah menunjukkan stabilitas, memastikan model

mencapai performa optimal sambil mencegah overfitting. Pattern pembelajaran ini mengindikasikan bahwa model berhasil mempelajari fitur-fitur penting dari data kelas "Ha_fathah" dengan sangat baik. Tetapi ada juga model yang kurang bagus seperti pada model "alif_dommah", grafik pelatihannya dapat dilihat pada Gambar 5.



Gambar 5. Grafik Accuracy dan Loss Model 03. alif_dommah

Analisis grafik model untuk kelas "alif_dommah" menunjukkan tantangan dalam proses pembelajaran. Training loss mengalami penurunan drastis pada epoch pertama dan stabil di nilai 1.0, namun validation loss tetap lebih tinggi. Pada grafik akurasi, training accuracy berfluktuasi di kisaran 50-55%, sementara validation accuracy menurun dari 50% menjadi 35%. Gap yang besar antara training dan validation mengindikasikan overfitting, sementara training accuracy yang rendah menunjukkan underfitting [18], yang terlihat dari hasil akhir dengan akurasi 34.88%, presisi 35.00%, recall 31.82%, dan F1-Score 33.33%. Performa rendah ini kemungkinan disebabkan oleh kompleksitas pengucapan huruf alif dengan harakat dommah dan variasi dalam dataset. Model mengalami masalah fitting ganda dimana terjadi underfitting karena model kesulitan mempelajari pola dasar, sekaligus overfitting karena model mulai menghafal noise dari data training.

3) Analisis Performa Model :

Analisis performa model menunjukkan bahwa karakteristik fonetik dari huruf-huruf hijaiyah memiliki pengaruh signifikan terhadap akurasi klasifikasi. Huruf-huruf yang memiliki karakteristik fonetik yang distingtif dan mudah dibedakan, seperti "shin", "Ha", dan "dzal", secara konsisten mencapai tingkat akurasi yang sangat tinggi dalam rentang 97-100%. Hal ini dapat dijelaskan karena huruf-huruf tersebut memiliki ciri bunyi yang unik dan pola spektral yang jelas dalam representasi mel-spectrogram, memungkinkan model untuk lebih mudah mengidentifikasi dan membedakannya dari huruf-huruf lain. Di sisi lain, model mengalami kesulitan dalam mengklasifikasikan huruf-huruf yang memiliki kemiripan fonetik, terutama terlihat pada kasus "alif" dan "ain" dengan harakat dommah yang hanya mencapai akurasi rendah antara 34-58%. Rendahnya performa pada huruf-huruf ini dapat dikaitkan dengan karakteristik suara yang mirip, yang menghasilkan pola spektral yang serupa dalam mel-spectrogram, sehingga menyulitkan model dalam membedakan nuansa halus antara pengucapan huruf-huruf tersebut.

Berdasarkan jenis harakat menunjukkan perbedaan yang signifikan dalam tingkat akurasi klasifikasi. Pelafalan huruf hijaiyah dengan harakat fathah secara konsisten menunjukkan performa yang lebih baik dengan rata-rata akurasi mencapai 90.12%. Performa ini secara signifikan lebih tinggi dibandingkan dengan pelafalan menggunakan harakat kasroh yang mencapai 79.96% dan harakat dommah dengan 77.82%. Superioritas akurasi pada harakat fathah ini dapat dikaitkan dengan karakteristik pengucapannya yang memiliki pola bunyi yang lebih jelas dan konsisten. Harakat fathah cenderung menghasilkan suara yang lebih terbuka dan tegas, menghasilkan pola spektral yang lebih distingtif dalam representasi mel-spectrogram, sehingga memudahkan model dalam proses klasifikasi.

TABEL 3
 PERBANDINGAN AKURASI BERDASARKAN HARAkat

Harokat	Rata – rata Akurasi
Fathah	90.12%
Kasroh	79.96%
Dommah	77.82%

Berdasarkan tempat keluarnya huruf (makhraj) mengungkapkan pola yang menarik dalam akurasi klasifikasi. Huruf-huruf syafawiyah, yang diucapkan dengan melibatkan bibir seperti "ba", "mim", dan "wau", menunjukkan variasi performa dengan rata-rata akurasi "ba" 77.51%, "mim" 79.84%, dan "wau" 67.44%. Performa ini dapat dikaitkan dengan kejelasan tempat keluarnya suara melalui bibir yang menghasilkan pola spektral terstruktur dalam representasi mel-spectrogram. Di sisi lain, huruf-huruf halqiyah yang diucapkan dari area tenggorokan, seperti "ain" dan "ghaiin", menampilkan variasi performa yang lebih signifikan dalam klasifikasi. Variabilitas ini dapat dijelaskan oleh kompleksitas intrinsik dalam pengucapan huruf-huruf tenggorokan, ditambah dengan adanya

variasi individual dalam cara pengucapan yang dapat mempengaruhi karakteristik spektral suara. Faktor-faktor ini menghadirkan tantangan lebih besar bagi model dalam mengidentifikasi dan mengklasifikasikan huruf-huruf hijaiyah secara akurat.

Hasil pada penelitian sebelumnya menunjukkan perbedaan yang signifikan dalam performa klasifikasi huruf hijaiyah. Model HMM yang diimplementasikan oleh Qothrun Nada [19], mencapai akurasi 100% untuk pengujian huruf yang sama dan 54.6% untuk pengujian huruf yang berbeda, menggunakan metode FFT untuk ekstraksi fitur dan Euclidean Distance untuk klasifikasi akhir. Sementara itu, model CNN yang dikembangkan dalam penelitian ini menunjukkan performa yang lebih seimbang dengan rata-rata akurasi keseluruhan 83.80%, dengan beberapa kelas mencapai akurasi 100% dan kelas terendah di sekitar 34-40%. Perbedaan utama terletak pada pendekatan arsitektur model, dimana HMM bergantung pada probabilitas transisi dan observasi, sedangkan CNN mampu mempelajari fitur spasial dari spektrogram mel secara langsung. Meski model HMM menunjukkan performa sempurna untuk huruf yang sama, model CNN memiliki kemampuan generalisasi yang lebih baik untuk membedakan huruf-huruf yang berbeda, tercermin dari rata-rata akurasi yang lebih tinggi pada kasus tersebut. Perbandingan ini menunjukkan bahwa masing-masing pendekatan memiliki kelebihan tersendiri - HMM unggul dalam pengenalan pola yang konsisten, sementara CNN lebih robust dalam menangani variasi dan perbedaan antar kelas.

Bisa kita lihat juga pada penelitian Arcapada [20], dengan penelitian ini menunjukkan perbedaan dalam karakteristik dataset dan hasil klasifikasi huruf hijaiyah. Arcapada menggunakan CNN dengan proses ekstraksi fitur MFCC pada dataset 3360 sampel suara yang terbagi menjadi 28 kelas huruf hijaiyah, menghasilkan akurasi 49.1%, presisi 50%, recall 53%, dan F-measure 0.514. Sementara penelitian ini menggunakan arsitektur CNN untuk mengklasifikasikan 8.904 sampel suara yang mencakup 84 kelas berbeda (kombinasi huruf dengan tiga harakat), mencapai rata-rata akurasi 83.80%. Perbedaan signifikan terletak pada kompleksitas klasifikasi, dimana penelitian ini menangani variasi yang lebih kompleks dengan mempertimbangkan kombinasi huruf-harakat, dibandingkan klasifikasi huruf tunggal pada penelitian Arcapada. Meskipun kedua penelitian menggunakan metode ekstraksi fitur berbasis skala mel, arsitektur CNN yang dikembangkan dalam penelitian ini mampu menangani dataset yang lebih besar dan kompleksitas kelas yang lebih tinggi dengan performa yang lebih baik. Hal ini menunjukkan bahwa pemilihan arsitektur CNN yang tepat dan penggunaan dataset yang lebih besar dapat meningkatkan akurasi dalam klasifikasi huruf hijaiyah, bahkan untuk kasus dengan variasi kelas yang lebih kompleks.

Berdasarkan analisis yang telah dilakukan, performa model CNN dalam klasifikasi pengucapan huruf hijaiyah dipengaruhi oleh beberapa faktor kritis. Kualitas dan konsistensi dataset recording menjadi fondasi utama, dimana variasi dalam pengucapan antar partisipan dapat mempengaruhi akurasi klasifikasi. Kompleksitas intrinsik dari huruf dan harakat tertentu, terutama pada kombinasi huruf-harakat yang memiliki kemiripan fonetik, memberikan tantangan tersendiri bagi model dalam membedakan karakteristik suara. Adanya potensi noise dan interferensi selama proses perekaman juga dapat mempengaruhi kualitas input yang diterima model. Hal ini mengimplikasikan perlunya pengembangan lebih lanjut, termasuk penambahan data khususnya untuk kelas-kelas dengan performa rendah, implementasi teknik augmentasi data untuk meningkatkan variasi dataset, serta pertimbangan penggunaan arsitektur model yang lebih kompleks untuk menangani kelas-kelas yang memiliki tingkat kesulitan lebih tinggi dalam klasifikasi.

4) Endpoint API untuk Model :

Implementasi endpoint API untuk model klasifikasi pengucapan huruf hijaiyah menggunakan Flask framework dengan endpoint utama `/predict_biner`. API menerima input berupa file audio dalam format WAV dengan ukuran maksimum 5MB, beserta parameter tambahan seperti huruf yang diuji, kondisi (fathah, kasroh, dhommah), dan informasi waktu pembelajaran. File audio yang diterima kemudian diproses menggunakan `librosa` untuk mengekstrak fitur `mel-spectrogram` dengan parameter yang telah dioptimasi (`n_mels=128`, `n_fft=2048`, `hop_length=512`), yang selanjutnya diumpankan ke model CNN yang sesuai untuk klasifikasi. Output dari API dikembalikan dalam format JSON yang berisi hasil klasifikasi ("benar" atau "kurang") dan disimpan ke dalam database sebagai history pembelajaran pengguna.

```
Request:
POST /predict_biner
Headers:
Authorization: Bearer <jwt_token>
Content-Type: multipart/form-data

Body:
file: audio_sample.wav
huruf: "alif"
kondisi: "fathah"
hasil_prediksi_diinginkan: "01. alif_fathah"
tanggal: "2024-03-23"
waktu: "14:30:00"

Response Success:
{
  "result": "benar"
}

Response Error:
{
  "msg": "No file selected"
}
```

Gambar 6. Contoh (a) Request, (b) Response

Pada Gambar 6, contoh request di atas, API membutuhkan header Authorization yang berisi JWT token untuk autentikasi pengguna. Body request menggunakan format multipart/form-data yang terdiri dari file audio (.wav) dan beberapa parameter penting: huruf yang dilafalkan (huruf), jenis harakat (kondisi), kelas yang akan diprediksi (hasil_prediksi_diinginkan), serta informasi waktu pembelajaran (tanggal dan waktu). Response sukses akan mengembalikan objek JSON dengan properti "result" yang bernilai "benar" jika pengucapan sesuai atau "kurang" jika pengucapan tidak sesuai. Jika terjadi error, response akan mengembalikan pesan error spesifik dalam properti "msg".

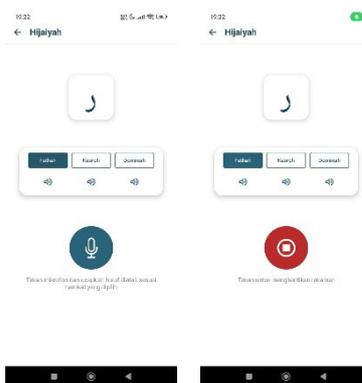
B. Aplikasi Mobile

Pengembangan aplikasi mobile dilakukan menggunakan Jetpack Compose, sebuah toolkit modern untuk membangun UI Android dengan pendekatan deklaratif, yang dipadukan dengan arsitektur MVVM (Model-View-ViewModel) untuk memastikan pemisahan yang jelas antara logika bisnis dan tampilan aplikasi. Implementasi MVVM memungkinkan pengelolaan state aplikasi yang lebih terstruktur, di mana ViewModel bertindak sebagai perantara antara data dari repository dan tampilan UI yang dibangun dengan Compose. Antarmuka pengguna dirancang mengikuti prinsip Material Design 3, terdiri dari beberapa komponen utama seperti layar perekaman dengan visualisasi audio real-time dan layar hasil klasifikasi. Komunikasi dengan backend API diimplementasikan menggunakan Retrofit dan Coroutines untuk operasi asynchronous, sementara Hilt digunakan untuk dependency injection yang memudahkan pengelolaan dependensi antar komponen. Flow dan StateFlow dimanfaatkan untuk mengelola state aplikasi secara reaktif, menciptakan aliran data yang efisien antara ViewModel dan komponen UI. Pendekatan arsitektur ini menghasilkan aplikasi yang tidak hanya responsif dan mudah digunakan, tetapi juga mudah dimaintain dan dikembangkan lebih lanjut.



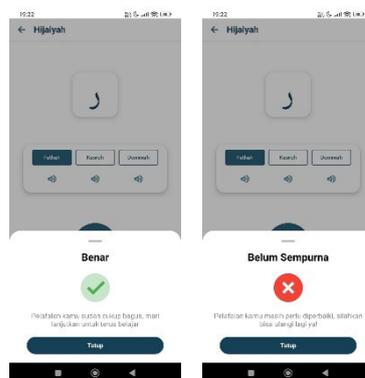
Gambar 7. Tampilan Home Screen aplikasi

Pada tampilan home screen aplikasi Gambar 7, pengguna dapat memilih huruf hijaiyah yang ingin dipelajari dari daftar 28 huruf hijaiyah. Setiap huruf ditampilkan dalam bentuk card yang interaktif, memungkinkan pengguna untuk memilih dan fokus pada huruf-huruf tertentu yang ingin dilatih sebelum melanjutkan ke proses perekaman suara. Fitur ini memberikan fleksibilitas bagi pengguna untuk menyesuaikan pembelajaran sesuai dengan kebutuhan dan tingkat kemampuan mereka.



Gambar 8. Tampilan Detail Huruf Screen (a) Kondisi memilih harokat, (b) Kondisi melakukan record

Setelah memilih huruf hijaiyah yang ingin dipelajari, pengguna akan diarahkan ke halaman detail huruf sesuai Gambar 8, yang menampilkan huruf terpilih beserta tiga pilihan harokat (fathah, kasrah, dan dhammah). Pada setiap harokat terdapat tombol suara yang memungkinkan pengguna untuk mendengarkan contoh pengucapan yang benar dari huruf tersebut. Di bagian bawah layar, tersedia tombol mikrofon yang berfungsi untuk memulai dan menghentikan perekaman suara pengguna. Pengguna dapat merekam pengucapan huruf sesuai dengan harokat yang telah dipilih untuk kemudian dievaluasi oleh sistem. Hasil dari rekaman pengguna akan diteruskan ke endpoint API model.



Gambar 9. Tampilan hasil prediksi (a) Ketika klasifikasi benar, (b) Ketika klasifikasi salah

Setelah pengguna selesai merekam suara pengucapan huruf hijaiyah, sistem akan melakukan klasifikasi menggunakan model CNN yang telah dilatih. Hasil klasifikasi akan ditampilkan dalam bentuk feedback visual kepada pengguna - jika tingkat keyakinan (confidence) model melebihi 50%, maka akan muncul notifikasi "Benar" dengan ikon centang hijau yang menandakan pengucapan sudah cukup bagus, disertai pesan motivasi untuk terus belajar. Sebaliknya, jika tingkat keyakinan di bawah ambang batas tersebut, sistem akan menampilkan notifikasi "Belum Sempurna" dengan ikon silang merah, serta pesan yang mendorong pengguna untuk mencoba kembali memperbaiki pengucapannya.

IV. SIMPULAN

Penelitian ini telah berhasil mengembangkan sistem klasifikasi pengucapan huruf hijaiyah menggunakan metode Convolutional Neural Network (CNN) dengan fitur mel-spectrogram, yang diimplementasikan dalam bentuk aplikasi mobile berbasis Android. Hasil analisis menunjukkan bahwa karakteristik fonetik dari huruf hijaiyah memiliki pengaruh signifikan terhadap akurasi klasifikasi, dengan rata-rata akurasi keseluruhan mencapai 83.80%. Pelafalan huruf dengan harakat fathah secara konsisten menunjukkan performa terbaik (90.12%) dibandingkan dengan harakat kasroh (79.96%) dan dommah (77.82%). Beberapa kelas seperti "shin", "Ha", dan "dzal" mencapai akurasi sempurna 100% karena karakteristik fonetik yang distingtif, sementara kelas seperti "alif_dommah" dan "ghaiin_dommah" mengalami tantangan dengan akurasi di bawah 40% akibat kemiripan pola spektral dalam mel-spectrogram. Sistem yang dikembangkan telah berhasil memberikan platform pembelajaran yang interaktif melalui integrasi aplikasi mobile dengan API Flask, namun masih memerlukan pengembangan lebih lanjut terutama dalam perluasan dataset untuk mengatasi masalah overfitting dan meningkatkan kemampuan generalisasi model agar dapat menangani variasi pengucapan yang lebih beragam.

UCAPAN TERIMAKASIH

Ucapan terima kasih yang sebesar-besarnya kepada Pondok Tahfidz Yanbuul Qur'an Menawan Kudus yang telah memberikan dukungan baik secara materi maupun non-materi selama proses penelitian berlangsung. Kontribusi dan partisipasi aktif dari pihak pondok pesantren telah sangat membantu dalam pengumpulan data dan pengembangan sistem klasifikasi pengucapan huruf hijaiyah ini.

DAFTAR PUSTAKA

- [1] R. Helmalia, L. Suzanti, and Rr. D. Widjayatri, "Pengenalan Huruf Hijaiyah Melalui Metode Tilawati bagi Anak Usia 5-6 Tahun," *Aulad: Journal on Early Childhood*, vol. 7, no. 1, pp. 199–209, Apr. 2024, doi: 10.31004/aulad.v7i1.634.
- [2] F. G. Safinatunnajah, A. Prasetiadi, and M. Wibowo, "CLASSIFICATION OF CAT SOUNDS USING CONVOLUTIONAL NEURAL NETWORK (CNN) AND LONG SHORT-TERM MEMORY (LSTM) METHODS," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 5, pp. 1349–1353, Oct. 2022, doi: 10.20884/1.jutif.2022.3.5.373.
- [3] V. Karenina, M. F. Erinsyah, and D. S. Wibowo, "Klasifikasi Rentang Usia Dan Gender Dengan Deteksi Suara Menggunakan Metode Deep Learning Algoritma Cnn (Convolutional Neural Network)," *Komputika : Jurnal Sistem Komputer*, vol. 12, no. 2, pp. 75–82, Sep. 2023, doi: 10.34010/komputika.v12i2.10516.
- [4] A. Maccagno, A. Mastropietro, U. Mazziotta, M. Scarpiniti, Y.-C. Lee, and A. Uncini, "A CNN Approach for Audio Classification in Construction Sites."
- [5] K. M. Rezaul *et al.*, "Enhancing Audio Classification Through MFCC Feature Extraction and Data Augmentation with CNN and RNN Models," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 7, pp. 37–53, 2024, doi: 10.14569/IJACSA.2024.0150704.
- [6] J. Lee, J. Nam, T. Kim, and J. Park, "Raw Waveform-based Audio Classification Using Sample-level CNN Architectures," 2017, doi: 10.48550/arXiv.1712.00866.
- [7] H. Mahmud Nawawi, A. Baitul Hikmah, A. Mustopa, and G. Wijaya, "Model Klasifikasi Machine Learning untuk Prediksi Ketepatan Penempatan Karir," *Jurnal SAINTEKOM*, vol. 14, no. 1, pp. 13–25, Mar. 2024, doi: 10.33020/saintekom.v14i1.512.
- [8] M. Aminuddin, "English Spoken Digit Recognition using Convolutional Neural Network (CNN)," *Jurnal EEICT (Electric Electronic Instrumentation Control Telecommunication)*, vol. 6, no. 2, Oct. 2023, doi: 10.31602/eeict.v6i2.11877.
- [9] G. Ajinurseto, L. O. Bakrim, and N. Islamuddin, "Penerapan Metode Mel Frequency Cepstral Coefficients pada Sistem Pengenalan Suara Berbasis Desktop," *Infomatek*, vol. 25, no. 1, pp. 11–20, Jun. 2023, doi: 10.23969/infomatek.v25i1.6109.
- [10] Ulfah Nur Oktaviana, Ricky Hendrawan, Alfian Dwi Khoirul Annas, and Galih Wasis Wicaksono, "Klasifikasi Penyakit Padi berdasarkan Citra Daun Menggunakan Model Terlatih Resnet101," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, pp. 1216–1222, Dec. 2021, doi: 10.29207/resti.v5i6.3607.
- [11] M. Agil Izzulhaq, "Indonesian Journal of Mathematics and Natural Sciences Penerapan Algoritma Convolutional Neural Network Arsitektur ResNet50V2 Untuk Mengidentifikasi Penyakit Pneumonia." [Online]. Available: <https://journal.unnes.ac.id/journals/JM/index>
- [12] N. Fadlia and R. Kosasih, "KLASIFIKASI JENIS KENDARAAN MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN)," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 24, no. 3, pp. 207–215, 2019, doi: 10.35760/tr.2019.v24i3.2397.
- [13] H. W. Sulistyono, H. Oktavianto, Z. Arifin, A. E. Wardoyo, and Q. Fitriyah, "PEMANFAATAN API PADA APLIKASI PORTAL BERITA SEDERHANA," *JUTIKOMP*, 2024.
- [14] R. Nandang Pratama and Y. A. Susetyo, "Implementasi Python API dengan Framework Flask sebagai Cloud Run Service Untuk Proses Update di PT. XYZ," 2024.
- [15] N. Puspitasari, E. Budiman, Y. N. Sulaiman, and M. B. Firdaus, "Microservice API Implementation for E-Government Service Interoperability," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Apr. 2021. doi: 10.1088/1742-6596/1807/1/012005.
- [16] F. Dzulqarnain, "Computer Based Information System Journal RANCANG BANGUN APLIKASI BELAJAR ARAB UNTUK ANDROID MENGGUNAKAN JETPACK COMPOSE DAN KOTLIN," *CBIS JOURNAL*, vol. 11, no. 01, 2023, [Online]. Available: <http://ejournal.upbatam.ac.id/index.php/cbis>
<http://ejournal.upbatam.ac.id/index.php/cbis>
- [17] B. Rizki, P. Surya, A. Putra Kharisma, and N. Yudistira, "Perbandingan Kinerja Pola Perancangan MVC, MVP, dan MVVM Pada Aplikasi Berbasis Android (Studi kasus : Aplikasi Laporan Hasil Belajar Siswa SMA BSS)," 2020. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [18] Mahardika Yoshi Putra, "Rancang Bangun Deteksi Bentuk Wajah Untuk Menentukan Gaya Rambut Menggunakan Algoritma CNN," *Repeater : Publikasi Teknik Informatika dan Jaringan*, vol. 2, no. 3, pp. 206–212, Jul. 2024, doi: 10.62951/repeater.v2i3.139.
- [19] Q. Nada, C. Ridhuandi, P. Santoso, and D. Apriyanto, "Speech Recognition dengan Hidden Markov Model untuk Pengenalan dan Pelafalan Huruf Hijaiyah," 2019.
- [20] R. Punggawa Arcapada, W. Setiawan, I. Made Arsa Suyadnya, J. Raya Kampus Unud Jimbaran, and K. Badung, "Desember 2021 Robbani Punggawa Arcapada, Widyadi Setiawan, I Made Arsa Suyadnya 1."