

ABSTRAK

Rendahnya kemampuan membaca Al-Qur'an di Indonesia 72,25% dari 3.111 responden berada pada level cukup dan kurang menunjukkan kebutuhan akan sistem evaluasi pelafalan yang mampu memberikan umpan balik objektif terhadap pelafalan huruf hijaiyah. Penelitian ini mengimplementasikan arsitektur **MobileNetV2** dengan ekstraksi fitur Mel-Frequency Cepstral Coefficients (MFCC) untuk mengklasifikasikan pelafalan huruf hijaiyah, serta membandingkannya dengan Log-Mel Spectrogram sebagai representasi fitur alternatif. Penelitian disusun mengikuti kerangka kerja CRISP-DM dan menggunakan dataset publik berisi 8.904 rekaman audio WAV dari santri Pondok Tahfidz Yanbu'ul Qur'an Menawan, Kudus, yang mencakup 84 kelas (28 huruf hijaiyah $\times$ 3 harakat) dengan distribusi seimbang sempurna (106 sampel per kelas). Tahap pra-pemrosesan meliputi *noise reduction* berbasis spectral gating, trim silence, dan normalisasi durasi menjadi 2 detik. Fitur MFCC ($n_{mfcc} = 40$) dibentuk menjadi tiga kanal (statis, delta, delta-delta), sedangkan Log-Mel Spectrogram ($n_{mels} = 128$) diduplikasi menjadi pseudo-RGB; keduanya di-resize ke $224 \times 224 \times 3$ agar kompatibel dengan input **MobileNetV2** pra-terlatih ImageNet. Pelatihan dilakukan dengan augmentasi *Specaugment* dan strategi *transfer learning* dua fase pada dua skenario pembagian data (80:10:10 dan 70:15:15). Hasil pengujian menunjukkan model *baseline* **MobileNetV2** + MFCC mencapai test *accuracy* 77,54%, sedangkan **MobileNetV2** + Log-Mel Spectrogram mencapai 86,53% dengan *macro precision* 87,25%, *macro recall* 86,47%, dan *macro F1-score* 86,37%, sekaligus melampaui penelitian acuan (83,80%). Model terbaik diintegrasikan ke dalam aplikasi *mobile* berbasis FastAPI dan Flutter untuk pengujian pelafalan secara *real-time*.

Kata kunci: **MobileNetV2**, MFCC, Log-Mel Spectrogram, klasifikasi pelafalan huruf hijaiyah, *transfer learning*

ABSTRACT

The low level of Qur'anic reading proficiency in Indonesia where 72.25% of 3,111 respondents scored at the fair or poor level highlights the need for learning media capable of providing objective feedback on hijaiyah letter pronunciation. This study implements the **MobileNetV2** architecture with Mel-Frequency Cepstral Coefficients (MFCC) feature extraction to classify hijaiyah letter pronunciation, and compares it against the Log-Mel Spectrogram as an alternative feature representation. The research follows the CRISP-DM framework and uses a public dataset of 8,904 WAV audio recordings collected from students of Pondok Tahfidz Yanbu'ul Qur'an Menawan, Kudus, covering 84 classes (28 hijaiyah letters $\times$ 3 diacritics) with a perfectly balanced distribution of 106 samples per class. Preprocessing consists of spectral-gating *noise reduction*, silence trimming, and duration normalization to 2 seconds. MFCC features ($n_mfcc = 40$) are arranged into three channels (static, delta, and delta-delta), while the Log-Mel Spectrogram ($n_mels = 128$) is stacked into a pseudo-RGB representation; both are resized to $224 \times 224 \times 3$ to match the input requirements of ImageNet-pretrained **MobileNetV2**. Training employs *SpecAugment augmentation* and a two-phase *transfer learning* strategy across two data-splitting scenarios (80:10:10 and 70:15:15). Experimental results show that the **MobileNetV2** + MFCC *baseline* achieves a test *accuracy* of 77.54%, whereas **MobileNetV2** + Log-Mel Spectrogram reaches 86.53% with a *macro precision* of 87.25%, *macro recall* of 86.47%, and *macro F1-score* of 86.37%, outperforming the reference study (83.80%). The best model is deployed into a FastAPI- and Flutter-based *mobile* application for *real-time* pronunciation testing.

Kata kunci: **MobileNetV2**, MFCC, Log-Mel Spectrogram, hijaiyah pronunciation classification, *transfer learning*