

Sistem Bantu Verifikasi Irama Hafalan *Nadzam Aqidatul Awam*

^{1*}Muhamad Nur Sabani, ²Danang Wahyu Widodo, ³Patmi Kasih

¹²³ Teknik Informatika, Universitas Nusantara PGRI Kediri

Email: ¹muhamadnursabani2@gmail.com, ²danangwahyuwidodo@unpkediri.ac.id,
³patmikasih@unpkdr.ac.id

Penulis Korespondens : Muhamad Nur Sabani

Abstrak—Kegiatan menghafal *Nadzam Aqidatul Awam* di pondok pesantren umumnya masih dinilai secara manual. Penelitian ini mengusulkan sistem verifikasi irama hafalan otomatis menggunakan metode *Siamese Long Short-Term Memory* (LSTM) berbasis ekstraksi ciri hibrida (MFCC, *Delta*, *Delta-Delta*, dan kontur *pitch* YIN). Dataset terdiri dari 400 audio rekaman 20 santri Pondok Pesantren Al Ma'ruf Kedunglo dengan acuan irama Ustadz Muhammad Shabri. Pengujian objektif menggunakan data uji terisolasi di Jupyter Notebook menghasilkan akurasi 92%, presisi 88%, *recall* 96%, dan *F1-score* 92%. Namun, pengujian *real-time* pada antarmuka Streamlit menunjukkan variasi performa akibat *environmental noise* dan perbedaan kualitas mikrofon (*hardware gain*). Riset ini berhasil memberikan kontribusi ilmiah dalam memperluas kapabilitas sistem verifikasi suara otomatis ke ranah penilaian estetika keselarasan ketukan irama tradisional keagamaan secara presisi.

Kata Kunci— Pemrosesan Sinyal Audio, Matriks Kebingungan, MFCC, *Nadzam Aqidatul Awam*, LSTM Siamese.

Abstract— The memorization of *Nadzam Aqidatul Awam* in Islamic boarding schools is generally still assessed manually. This study proposes an automatic recitation rhythm verification system using the *Siamese Long Short-Term Memory* (LSTM) method based on hybrid feature extraction (MFCC, *Delta*, *Delta-Delta*, and YIN pitch contour). The dataset consists of 400 audio recordings from 20 students of Al Ma'ruf Kedunglo Islamic Boarding School, using Ustadz Muhammad Shabri's rhythm as the reference standard. Objective testing using isolated data in Jupyter Notebook achieved an accuracy of 92%, precision of 88%, recall of 96%, and an *F1-score* of 92%. However, real-time functional testing via the Streamlit interface indicated performance variations caused by environmental noise and differences in microphone hardware gain. This research successfully contributes to expanding the capability of automatic voice verification systems into the aesthetic assessment of traditional religious rhythmic alignment.

Keywords— Audio Signal Processing, Confusion Matrix, MFCC, *Nadzam Aqidatul Awam*, Siamese LSTM.

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Dalam proses pembelajaran di lingkungan pesantren, kegiatan menghafal merupakan salah satu metode utama dalam memahami dan menguasai ilmu keagamaan[1]. Salah satu materi yang sering dihafalkan oleh para santri adalah *Nadzam Aqidatul Awam*, yaitu bait-bait *nadzam* yang berisi ajaran akidah Islam dalam bentuk syair berbahasa Arab[2]. Dalam praktiknya, seorang santri biasanya melafalkan bait *nadzam* secara berulang tanpa melihat teks sebagai

bentuk latihan dan penguatan daya ingat[3]. Latihan hafalan tersebut tidak hanya membutuhkan ketekunan dalam mengingat, tetapi juga kesesuaian irama nadzam dengan referensi. Oleh karena itu, hasil hafalan santri perlu jika dilakukan secara mandiri membutuhkan alat verifikasi atau diperiksa kebenarannya agar sesuai dengan irama yang benar.

Permasalahan muncul ketika santri melakukan hafalan secara mandiri tanpa pendamping, sehingga mengalami kesulitan dalam mengetahui apakah bacaan yang diucapkan sudah sesuai dengan teks referensi. Proses penilaian kesesuaian irama hafalan Nadzam Aqidatul Awam di Pondok pesantren Al Ma'ruf umumnya masih dilakukan secara manual dan sangat bergantung pada kehadiran pendamping sebagai verifikator. Perkembangan teknologi pengenalan suara (*speech recognition*) membuka peluang baru dalam mengatasi permasalahan tersebut. Teknologi ini memungkinkan sistem komputer untuk mengenali pola suara manusia[4] dan membandingkannya dengan data suara acuan yang telah disimpan sebelumnya[5].

Beberapa penelitian terdahulu telah menerapkan arsitektur Siamese LSTM untuk verifikasi sekuens vokal, namun analisisnya sering kali masih terbatas pada ekstraksi ciri konvensional sehingga kurang sensitif terhadap dinamika perubahan intonasi dan cengkok vokal[6]. Oleh karena itu, diperlukan integrasi fitur dinamis untuk menangkap laju perubahan irama secara temporal.

Berdasarkan hal tersebut, penelitian ini mengusulkan Sistem Bantu Verifikasi Irama Hafalan Nadzam Aqidatul Awam. Sistem ini dirancang menggunakan bahasa pemrograman Python 3.11.6 dengan library Librosa untuk ekstraksi ciri audio, TensorFlow/Keras untuk pemodelan Siamese LSTM, serta antarmuka berbasis web menggunakan Streamlit. Melalui penelitian ini, diharapkan dapat menyediakan solusi alternatif bagi santri dalam melakukan verifikasi hafalan secara mandiri dan objektif.

II. METODE

Metodologi penelitian pada artikel ini disusun dengan mengacu pada metode pengembangan sistem *Waterfall* dalam membangun sistem bantu verifikasi irama hafalan *Nadzam Aqidatul Awam*. Model *Waterfall* digunakan karena sifatnya yang sistematis dan sekuensial, sehingga memastikan setiap tahapan pengembangan berjalan secara terstruktur[7]. Secara rinci, setiap tahapan pengembangan tersebut dijabarkan pada subbab-subbab sekuensial berikut.

2.1 Skema Korpus Dataset dan Pelatihan

Dataset yang digunakan dalam penelitian ini berupa rekaman suara santri dan audio referensi Nadzam Aqidatul Awam yang dikumpulkan secara mandiri dengan format .wav (PCM), satu kanal (mono), dengan sampling rate 16 kHz dan resolusi 16-bit untuk menjaga konsistensi kualitas audio menggunakan mikrofon ponsel pintar (smartphone). Dataset terdiri dari 400 data audio yang diperoleh dari rekaman suara 20 santri, di mana setiap santri merekam 20 bait Nadzam Aqidatul Awam (bait 1–20). Tempat pengambilan sampel ini berlokasi di Pondok Pesantren Al Ma'ruf Kedunglo. Nada atau irama yang digunakan sebagai acuan referensi adalah irama Ustadz Muhammad Shabri, yang mencakup 2 variasi karakteristik intonasi yaitu nada rendah dan nada tinggi. Karakteristik detail mengenai bentuk spesifikasi elemen kontrol korpus data dijabarkan pada Tabel 1

Tabel 1. Karakteristik Spesifikasi Korpus Dataset

Parameter Elemen Data	Spesifikasi Teknis Operasional
Format Berkas	.wav (Waveform Audio File Format)
Audio Channel	1(Mono)
Sampling Rate	16.000 Hz (16 kHz)
Resolusi Bit	16 bit
Durasi per Berkas	6-10 Detik
Total Sampel Audio	400 file audio

Seluruh berkas audio digital tersebut disimpan secara lokal pada direktori komputer dengan struktur folder yang disusun secara sejajar guna memudahkan proses pemanggilan sequensial oleh algoritma. Visualisasi pemetaan direktori penyimpanan dataset diilustrasikan pada Gambar 1.

```
dataset/
├── referensi/
│   ├── tinggi/
│   │   ├── bait1.wav
│   │   ├── bait2.wav
│   │   └── ... sampai ... bait20.wav
│   └── rendah/
│       ├── bait1.wav
│       ├── bait2.wav
│       └── ... sampai ... bait20.wav
├── santri1/
│   ├── bait1.wav
│   └── ... sampai ... bait20.wav
├── santri2/
│   ├── bait1.wav
│   └── ... sampai ... bait20.wav
├── ... sampai ... santri20/
│   ├── bait1.wav
│   └── ... sampai ... bait20.wav
```

Gambar 1 Bagan Struktur Direktori Penyimpanan Korpus Dataset

Data yang telah terstruktur tersebut kemudian dibagi menggunakan rasio pemisahan data sebesar 80% sebagai data pelatihan (*training data*) untuk melatih model Siamese LSTM, dan 20% dialokasikan sebagai data pengujian (*testing data*) untuk proses evaluasi performa sistem.

2.2 Pra-pemrosesan Sinyal Suara

Sebelum dilakukan ekstraksi fitur akustik, setiap berkas audio mentah (.wav) yang ada di dalam korpus dataset harus melalui serangkaian tahapan pra-pemrosesan sinyal digital secara sekuensial. Proses ini bertujuan untuk mengondisikan sinyal vokal agar berada pada kondisi optimal dan seragam saat diumpankan ke dalam model jaringan. Tahapan pra-pemrosesan sinyal suara pada penelitian ini dijabarkan sebagai berikut:

- a. Normalisasi Amplitudo: Proses ini dilakukan untuk menyamakan skala sinyal audio dari seluruh sampel rekaman santri. Melalui teknik ini, amplitudo dari setiap berkas audio disesuaikan ke dalam rentang nilai yang seragam (misalnya antara -1.0 hingga 1.0) guna menghindari ketimpangan volume atau intensitas suara akibat perbedaan jarak mikrofon saat perekaman mandiri.
- b. Pengurangan Noise (Derau): Tahapan ini bertujuan untuk meningkatkan kualitas suara dengan mereduksi interferensi derau latar belakang (*environmental noise*) yang ikut terekam oleh ponsel pintar. Proses pembersihan sinyal ini memastikan bahwa komponen frekuensi yang dominan dalam audio benar-benar merepresentasikan karakteristik irama vokal santri, bukan derau luar.
- c. Pemotongan dan Penyesuaian Panjang Sinyal: Tahapan akhir ini dilakukan untuk menyesuaikan durasi atau panjang sinyal audio agar sesuai dengan dimensi masukan yang dibutuhkan oleh sistem. Bagian keheningan (*silence*) pada awal dan akhir rekaman dipotong, kemudian panjang sekuens temporal audio diselaraskan agar model *Siamese LSTM* dapat memproses pasangan input secara simultan dengan ukuran matriks yang konsisten.

2.3 Ekstraksi Fitur Suara

Setelah melalui tahapan pra-pemrosesan, sinyal audio diekstraksi ke dalam bentuk representasi numerik untuk menangkap karakteristik unik dari irama vokal santri dan audio referensi. Proses ekstraksi fitur pada penelitian ini menggabungkan parameter spektral dan parameter pitch (nada dasar) sebagai berikut:

- a. Mel-Frequency Cepstral Coefficients (MFCC): Digunakan untuk mengekstrak representasi amplop spektral dari sinyal suara yang meniru cara indra pendengaran manusia merespons frekuensi audio[8]. Fitur MFCC statis ini berperan penting dalam menangkap karakteristik fonetik dari bait Nadzam yang dilafalkan.
- b. Karakteristik Dinamis (Delta dan Delta-Delta MFCC): Untuk menangkap perubahan transisi temporal dan kontinuitas irama dari waktu ke waktu[9], fitur MFCC statis dilengkapi dengan koefisien turunan pertama (Delta) dan turunan kedua (Delta-Delta atau akselerasi). Penambahan koefisien dinamis ini sangat krusial bagi model guna mengenali fluktuasi cepat pada penekanan tajwid dan cengkok irama hafalan.
- c. Pitch Detection (F0): Fitur nilai frekuensi fundamental (F_0) diekstrak untuk mendeteksi tinggi rendahnya nada yang dihasilkan oleh vokal santri secara objektif[10]. Mengingat audio referensi memiliki variasi karakteristik intonasi nada tinggi dan nada rendah, ekstraksi parameter pitch ini berfungsi sebagai indikator utama penentu keselarasan melodi hafalan.

2.4 Arsitektur Siamese LSTM

Proses verifikasi keselarasan irama hafalan pada penelitian ini bertumpu pada pemodelan *deep learning* menggunakan arsitektur *Siamese LSTM*. Berbeda dengan jaringan saraf standar, arsitektur *Siamese* menerapkan konsep jaringan kembar (*twin networks*) yang memiliki dua sub-jaringan LSTM (*Long Short-Term Memory*) dengan struktur, bobot parameter (*shared weights*), dan konfigurasi bias yang identik secara paralel[11]. Pemilihan lapisan LSTM didasarkan pada kemampuannya yang sangat optimal dalam memproses data sekuensial temporal dan mengenali dependensi jangka panjang (*long-term dependencies*) dari matriks fitur akustik (MFCC, *Delta*, *Delta-Delta*, dan *Pitch*) yang memiliki panjang durasi bervariasi.

Secara operasional, sub-jaringan pertama bertugas memproses vektor fitur dari audio referensi ustadz (X_{ref}), sedangkan sub-jaringan kedua secara simultan memproses vektor fitur dari audio rekaman vokal santri (X_{san}). Kedua jalur tersebut menghasilkan output berupa representasi vektor *embedding* berdimensi rendah yang merangkum esensi karakteristik temporal dan melodi dari masing-masing input audio.

2.5 Pengukuran Jarak Selisih Absolut

Setelah kedua sub-jaringan kembar Siamese LSTM menghasilkan *embedding vector* berdimensi 16 dari masing-masing jalur input, langkah selanjutnya adalah menghitung matriks nilai selisih jarak spasial antar-vektor tersebut. Pengukuran jarak dalam penelitian ini dilakukan menggunakan operasi selisih absolut (*Absolute Difference*) memanfaatkan lapisan Lambda pada Keras melalui Persamaan (7):

$$distance = |f(X_{ref}) - f(X_{san})| \quad (1)$$

Keterangan:

- X_{ref} = vektor fitur dari suara referensi Ustadz Muhammad Shabri.
- X_{san} = vektor fitur dari suara uji santri.
- $f()$ = fungsi ekstraksi fitur pemetaan sub-jaringan kembar (*embedding vector*).

Hasil perhitungan selisih absolut ini berupa nilai vektor jarak tunggal yang merepresentasikan tingkat kesenjangan intonasi, lalu diumpankan menuju satu neuron akhir (*Dense layer*) beraktivasi *sigmoid*. Fungsi *sigmoid* memetakan luaran spasial tersebut ke dalam nilai probabilitas kemiripan kontinu dengan rentang nilai antara 0 hingga 1[12].

2.6 Skema Evaluasi Kuantitatif

Pengujian non-fungsional dilakukan menggunakan porsi data uji (*testing data*) sebesar 20% dari total data yang belum pernah dikenali oleh model selama fase pelatihan. Evaluasi kinerja arsitektur Siamese LSTM diukur secara komparatif dengan memetakan sebaran hasil prediksi biner menggunakan diagram *Confusion Matrix* melalui bantuan pustaka *Scikit-Learn*.

Berdasarkan sebaran matriks kesalahan tersebut, tingkat keandalan dan kepekaan model dalam mengenali keselarasan irama hafalan Nadzam Aqidatul Awam santri dianalisis secara objektif berdasarkan empat metrik evaluasi utama, yaitu *Accuracy* (Akurasi), *Precision* (Presisi), *Recall* (Sensitivitas), dan *F1-Score*. Hasil kalkulasi performa ini kemudian disajikan dan dibahas secara mendalam pada bab hasil dan pembahasan untuk membuktikan validitas sistem yang dibangun.

III. HASIL DAN PEMBAHASAN

3.1 Pengujian Fungsional Aplikasi

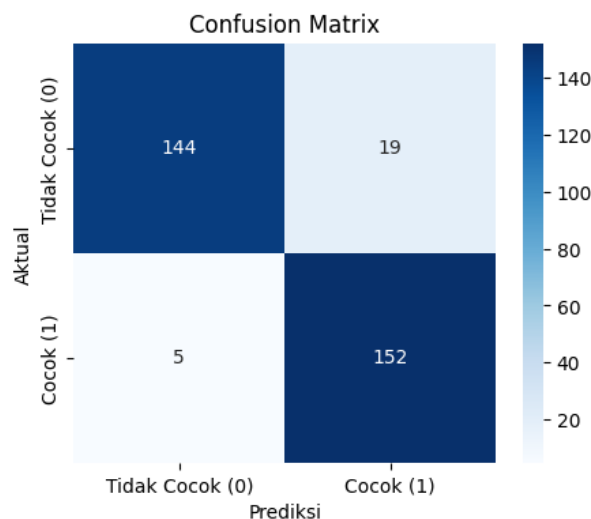


Gambar 3 Antarmuka Pengguna

Pengujian fungsional sistem dilakukan menggunakan metode Black-Box Testing untuk memvalidasi seluruh fungsi antarmuka pada framework Streamlit, mulai dari proses perekaman audio secara real-time hingga penentuan keputusan hasil verifikasi irama hafalan. Berdasarkan skenario uji yang dieksekusi, komponen pemilihan bait nadzam berhasil memuat berkas audio acuan ustadz, widget mikrofon sukses menangkap sinyal vokal santri, dan modul back-end secara instan mampu melakukan ekstraksi ciri hingga mencetak status klasifikasi biner pada layar secara fungsional.

3.2 Hasil Evaluasi Kuantitatif Performa Model

Pengujian non-fungsional untuk mengukur kinerja arsitektur Siamese LSTM dilakukan menggunakan data uji sebesar 20% (setara dengan 80 sampel audio santri) yang diisolasi di lingkungan Jupyter Notebook menggunakan fungsi `model.predict()`. Hasil sebaran prediksi dari pengujian biner terhadap data murni tersebut dipetakan ke dalam diagram *Confusion Matrix* berbantuan pustaka *Scikit-Learn* seperti yang disajikan pada Gambar 3.1.



Gambar 3. 1 Confuision Matrix

Berdasarkan hasil plot pada Gambar 3.1, diperoleh rincian sebaran prediksi berupa parameter *True Positive* (TP) sebanyak 152 sampel, *True Negative* (TN) sebanyak 144 sampel, *False Positive* (FP) sebanyak 19 sampel, dan *False Negative* (FN) sebanyak 5 sampel. Nilai sebaran kesalahan tersebut kemudian dikalkulasikan untuk mengukur tingkat keandalan model berdasarkan empat metrik evaluasi utama. Hasil pengujian performa non-fungsional secara mendetail disajikan pada Tabel 3.

Tabel 3. 1 Hasil Pengujian Non-Fungsional Performa Model Siamese LSTM

Metrik Evaluasi Model	Parameter Evaluasi			
	Akurasi	Presisi	Recall	F1-Score
Siamese LSTM	92%	88%	96%	92%

3.3 Pengujian Data

Untuk memvalidasi kinerja sistem secara nyata, dilakukan pengujian langsung menggunakan kombinasi pelafalan bait *Nadzam Aqidatul Awam* dengan variasi model cengkok irama (Irama Ganjil dan Irama Genap). Pengujian ini bertujuan untuk melihat sensitivitas algoritma *Siamese LSTM* dalam mengenali keselarasan irama hafalan berdasarkan nilai ambang batas (*threshold*) probabilitas kemiripan yang dihasilkan.

Tabel 3. 2 Hasil Pengujian Data

Nadzam Yang Dipilih	Hafalan Santri	Hasil Pengujian	Tingkat Kemiripan
1	Nadzam 1 Irama Ganji	Sesuai	83%
1	Nadzam 1 Irama Genap	Tidak Sesuia	0%
2	Nadzam 2 Irama Ganjil	Tidak Sesuia	0%
2	Nadzam 2 Irama Genap	Sesuia	95%
3	Nadzam 3 Irama Ganjil	Tidak Sesuia	4%
3	Nadzam 3 Irama Genap	Sesuia	92%

3.4 Pembahasan Analitis dan Kebaruan Riset

Perolehan nilai akurasi yang menyentuh angka 92% serta *F1-score* sebesar 92% membuktikan bahwa model *Siamese LSTM* bekerja dengan sangat optimal dalam mengenali karakteristik prosodi suara secara sekuensial. Integrasi ciri spektral dari 13 koefisien MFCC yang dikombinasikan dengan kontur *pitch* (algoritma YIN) terbukti kokoh dalam menangkap fluktuasi nada pelantunan nadzam secara hibrida. Nilai *recall* yang mencapai 96% mengindikasikan bahwa model memiliki sensitivitas yang tinggi dalam mengenali sampel hafalan santri yang benar-benar sesuai dengan irama referensi Ustadz Muhammad Shabri, dengan tingkat kegagalan deteksi (*false negative*) yang sangat minim yaitu hanya sebesar 5 sampel.

Munculnya *error rate* berupa 19 sampel *false positive* (kondisi di mana irama santri tidak cocok namun diprediksi cocok oleh model) disebabkan oleh adanya kemiripan temporal pada jeda ketukan antartransisi bait nadzam. Namun secara keseluruhan, penggunaan komponen dinamis *Delta* dan *Delta-Delta* terbukti memberikan pengaruh signifikan dalam merekam laju perubahan artikulasi vokal terhadap waktu, sehingga sel memori LSTM dapat mempelajari ketergantungan temporal jangka panjang secara kontekstual.

Meskipun pengujian kuantitatif menggunakan data uji terisolasi menghasilkan tingkat akurasi yang tinggi (92%), pengujian secara langsung pada antarmuka *real-time* Streamlit menunjukkan adanya variasi performa di mana sistem terkadang memerlukan beberapa kali pengulangan rekaman untuk menghasilkan keputusan verifikasi yang sepenuhnya valid. Ketidaksesuaian operasional ini diidentifikasi terjadi akibat adanya faktor eksternal lingkungan (*environmental noise*) saat pengambilan sampel vokal secara langsung, serta perbedaan spesifikasi fungsional perangkat keras mikrofon (*hardware gain*) yang digunakan oleh pengguna. Faktor-faktor tersebut menyebabkan adanya pergeseran nilai magnitudo desibel dan fluktuasi kontur *pitch* yang tidak konstan jika dibandingkan dengan karakteristik lingkungan korpus data latih yang cenderung bersih dan terkontrol.

Jika dikomparasikan dengan penelitian sejenis terdahulu, seperti riset yang dilakukan oleh Rajagede & Hastuti (2021) yang berfokus pada verifikasi ketepatan pelafalan teks ayat Al-

Qur'an menggunakan jaringan Siamese LSTM standar, penelitian ini membawa kontribusi kebaruan ilmiah (*novelty*) yang berfokus pada keselarasan dendang irama lagu tradisional pesantren. Penambahan estimasi kontur *pitch* pada penelitian ini terbukti efektif menjaga performa model tetap tinggi. Sistem menjadi jauh lebih peka dalam membedakan antara santri yang melantunkan irama dengan benar dengan santri yang hanya membaca teks secara datar atau menyimpang dari ketukan acuan.

IV. KESIMPULAN

4.1 Kesimpulan

Penelitian ini berhasil menjawab tujuan utama dalam mengotomatisasi proses verifikasi keselarasan irama hafalan *Nadzam Aqidatul Awam* menggunakan pendekatan komputasi hibrida berbasis arsitektur *Siamese Long Short-Term Memory* (LSTM). Melalui integrasi fitur spektral *Mel-Frequency Cepstral Coefficients* (MFCC) bersama estimasi kontur *pitch*, sistem terbukti andal dalam mengenali karakteristik prosodi, fluktuasi nada, dan ketergantungan temporal jangka panjang sekuensial vokal santri terhadap jangkar audio referensi ustadz secara objektif. Implementasi model pada platform Streamlit secara fungsional juga telah sukses memfasilitasi proses perekaman interaktif yang mampu memberikan keputusan klasifikasi biner secara instan. Secara umum, riset ini memberikan kontribusi peningkatan yang signifikan pada bidang sains pengolahan sinyal suara (*audio signal processing*) dan kecerdasan buatan, khususnya dalam memperluas kapabilitas verifikasi suara (*voice verification*) dari yang semula hanya berfokus pada pengenalan teks atau identifikasi pembicara, kini beralih ke ranah penilaian estetika ketukan lagu tradisional keagamaan secara otomatis dan presisi.

4.2 Saran

Berdasarkan hasil pengujian di lingkungan dunia nyata, terdapat beberapa keterbatasan operasional yang dapat dijadikan acuan perbaikan untuk penelitian selanjutnya. Pengembangan sistem berikutnya disarankan untuk mengintegrasikan modul peredam bising (*noise cancellation*) atau normalisasi audio sebelum masuk ke tahap ekstraksi ciri guna meminimalkan distorsi akibat gangguan lingkungan sekitar (*environmental noise*). Selain itu, diperlukan penambahan variabilitas data latih yang mencakup berbagai variasi perangkat mikrofon pengguna (*hardware gain*) serta karakteristik vokal santri yang lebih heterogen guna meningkatkan kemampuan generalisasi laten dan stabilitas sistem saat dioperasikan secara *real-time*.

DAFTAR PUSTAKA

- [1] N. Fazwanis, B. Othman, M. F. Bin, and M. Yusof, "Utilising Dialogical Pedagogy as a Mode of Instruction with the Purpose of Improving Future Islamic Education through the Study of Qur'anic Principles," vol. 13, no. 10, pp. 394–400, 2023, doi: 10.6007/IJARBSS/v13-i10/18784.
- [2] P. Pembelajaran, K. Aqidatul, P. N.-Nilai, T. Santri, M. Diniyah, and T. Qulub, "Reslaj : Religion Education Social Laa Roiba Journal Reslaj : Religion Education Social Laa Roiba Journal," vol. 7, pp. 2363–2375, 2025, doi: 10.47476/reslaj.v7i8.9318.

- [3] A.-Q. U. R. An, D. I. Pondok, and A. Mu, “Tradisi sima’an dalam penguatan hafalan al-qur’an di pondok pesantren,” vol. 2, no. 2, pp. 309–320, 2025, doi: 10.63424/amsal.v2i2.375.
- [4] S. Alharbi et al., “Automatic Speech Recognition : Systematic Literature Review,” IEEE Access, vol. PP, p. 1, 2021, doi: 10.1109/ACCESS.2021.3112535.
- [5] S. Mariyanto, A. Sasongko, S. Tsauray, S. Ariessaputra, and S. Ch, “Mel Frequency Cepstral Coefficients (MFCC) Method and Multiple Adaline Neural Network Model for Speaker Identification,” vol. 7, no. December, 2023, doi: 10.62527/joiv.7.4.1376.
- [6] H. Yakura, K. Watanabe, and M. Goto, “Self-Supervised Contrastive Learning for,” IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 30, pp. 1614–1623, 2022, doi: 10.1109/TASLP.2022.3169627.
- [7] A. Zahra, D. Nur, A. Jl, L. Pol, S. No, and K. P. Utara, “Analisa Perbandingan Penggunaan Metodologi Pengembangan Perangkat Lunak (Waterfall , Prototype , Iterative , Spiral , Rapid Application Development (RAD)),” no. 4, 2024, doi: 10.61132/mercurius.v2i4.148.
- [8] Z. K. Abdul and A. K. Al-talabani, “Mel Frequency Cepstral Coefficient and its applications : A Review,” IEEE Access, vol. PP, p. 1, 2022, doi: 10.1109/ACCESS.2022.3223444.
- [9] R. Sebastian and S. O. Keefe, “Bridging Biological Hearing and Neuromorphic Computing : End-to-End Time-Domain Audio Signal Processing with Reservoir Computing”, doi: 10.48550/arXiv.2603.24283.
- [10] F. Islam, N. Parvin, M. Rahman, T. Ahmed, D. Chakraborty, and S. Rahman, “Fundamental Frequency Extraction by Utilizing the Combination of Spectrum in Noisy Speech,” vol. 13, no. 4, pp. 730–734, doi: 10.37391/IJEER.130413.
- [11] N. Sharma et al., “Siamese Convolutional Neural Network-Based Twin Structure Model for Independent Offline Signature Verification,” pp. 1–14, 2022, doi: 10.3390/su141811484.
- [12] Y. Lu, Y. Xu, E. Herrera-viedma, and Y. Han, “Consensus of large-scale group decision making in social network : the minimum cost model based on robust optimization,” Inf. Sci. (Ny), vol. 547, pp. 910–930, 2021, doi: 10.1016/j.ins.2020.08.022.