

Penerapan *Siamese Long Short-Term Memory* pada Sistem Verifikasi Bacaan Berbasis Suara

Muhamad Nur Sabani¹

¹Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: *¹muhamadnursabani1@gmail.com

Abstrak – Verifikasi bacaan berbasis suara merupakan pendekatan yang dapat digunakan untuk membantu proses evaluasi bacaan secara otomatis, terutama pada aplikasi yang memerlukan perbandingan bacaan tanpa bergantung pada transkripsi teks. Pendekatan ini penting karena mampu memberikan penilaian yang lebih objektif dan efisien dibandingkan evaluasi manual. Penelitian ini membahas penerapan metode *Siamese Long Short-Term Memory (LSTM)* pada sistem verifikasi bacaan berbasis suara sebagai implementasi awal. Sistem dirancang melalui beberapa tahapan, yaitu praproses sinyal suara, ekstraksi fitur menggunakan *Mel-Frequency Cepstral Coefficients (MFCC)* dan *pitch*, pemodelan dengan arsitektur *Siamese LSTM*, serta pengukuran kemiripan menggunakan *Euclidean Distance*. Dua masukan suara, yaitu suara referensi dan suara uji, diproses secara paralel untuk menghasilkan nilai kemiripan sebagai dasar pengambilan keputusan. Hasil penerapan menunjukkan bahwa metode *Siamese LSTM* dapat diimplementasikan pada sistem verifikasi bacaan berbasis suara dan mampu membedakan pasangan bacaan dengan tingkat kemiripan yang berbeda. Hasil ini menunjukkan bahwa pendekatan yang digunakan berpotensi menjadi dasar pengembangan sistem verifikasi bacaan berbasis suara yang lebih komprehensif di masa mendatang.

Kata Kunci : *LSTM, MFCC, Siamese Network, Verifikasi Bacaan, Kemiripan Suara*

1. PENDAHULUAN

Verifikasi bacaan berbasis suara merupakan salah satu aplikasi pengolahan sinyal suara yang bertujuan untuk menentukan tingkat kesesuaian antara dua bacaan audio [1],[2]. Pendekatan ini berbeda dengan pengenalan suara berbasis transkripsi teks karena fokus pada perbandingan pola suara secara langsung tanpa melakukan proses pengenalan teks [1]. Verifikasi bacaan berbasis suara dapat dimanfaatkan pada berbagai aplikasi yang membutuhkan proses evaluasi bacaan secara otomatis [3].

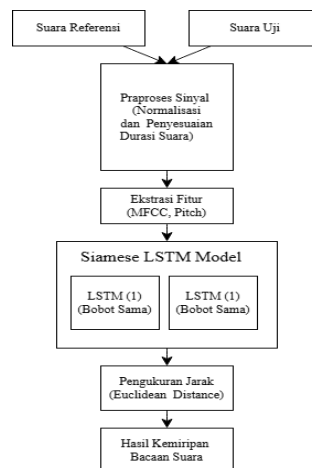
Perkembangan metode *deep learning* memberikan peluang dalam meningkatkan performa sistem verifikasi suara. *Long Short-Term Memory (LSTM)* merupakan salah satu arsitektur jaringan saraf yang efektif dalam memodelkan data sequential, termasuk sinyal suara [4],[5]. Selain itu, arsitektur *Siamese Network* memungkinkan pembelajaran kemiripan dua data melalui dua jaringan identik dengan bobot yang sama [6].

Berdasarkan hal tersebut, penelitian ini membahas penerapan metode *Siamese LSTM* pada sistem verifikasi bacaan berbasis suara sebagai implementasi awal. Fokus penelitian ini adalah pada bagaimana metode *Siamese LSTM* dapat digunakan untuk membandingkan dua sinyal suara berdasarkan tingkat kemiripannya.

2. METODE PENELITIAN

Metodologi penelitian pada artikel ini disusun untuk menerapkan metode *Siamese Long Short-Term Memory (LSTM)* dalam sistem verifikasi bacaan berbasis suara. Metodologi yang digunakan mencakup tahapan perancangan alur sistem, praproses sinyal suara, ekstraksi fitur, pemodelan menggunakan arsitektur *Siamese LSTM*, pengukuran kemiripan, skema pelatihan dan pengujian serta kriteria penentuan hasil verifikasi [4],[6],[7]. Metodologi ini difokuskan pada implementasi awal metode tanpa membahas konteks bacaan tertentu secara spesifik.

2.1 Alur Sistem Verifikasi Bacaan



Gambar 1 Alur Sistem Verifikasi Bacaan

Alur sistem verifikasi bacaan berbasis suara yang digunakan pada penelitian ini ditunjukkan pada Gambar 1. Sistem menerima dua masukan berupa sinyal suara, yaitu suara referensi dan suara uji. Kedua sinyal suara tersebut diproses melalui tahapan yang sama secara paralel untuk menjaga konsistensi proses pengolahan data.

Tahapan utama dalam sistem ini meliputi praproses sinyal suara, ekstraksi fitur, pemodelan menggunakan arsitektur *Siamese LSTM*, serta pengukuran tingkat kemiripan. Hasil akhir dari sistem berupa nilai kemiripan yang menunjukkan tingkat kesesuaian antara dua bacaan suara yang dibandingkan.

Untuk memperjelas tahapan proses pada sistem verifikasi bacaan berbasis suara, ringkasan setiap tahapan sistem disajikan pada Tabel 1.

Tabel 1. Tahapan Sistem Verifikasi Bacaan Berbasis Suara

Tahapan Sistem	Deskripsi
Input Suara	Sistem menerima dua masukan suara berupa suara referensi dan suara uji
Praproses Sinyal	Normalisasi amplitudo dan penyesuaian durasi sinyal suara
Ekstraksi Fitur	Ekstraksi fitur MFCC dan pitch dari sinyal suara
Pemodelan	Pemodelan fitur menggunakan arsitektur Siamese LSTM
Pengukuran Kemiripan	Perhitungan jarak menggunakan Euclidean Distance
Ouput Sistem	Nilai kemiripan sebagai dasar verifikasi bacaan

Tabel 1 menyajikan ringkasan tahapan utama pada sistem verifikasi bacaan berbasis suara yang diusulkan. Setiap tahapan menggambarkan proses yang dilalui sinyal suara mulai dari penerimaan masukan suara hingga dihasilkan nilai kemiripan sebagai dasar verifikasi bacaan. Penyajian tahapan dalam bentuk tabel bertujuan untuk memperjelas alur sistem secara terstruktur dan melengkapi visualisasi alur sistem yang ditunjukkan pada Tabel 1

2.2 Praproses Sinyal Suara

Praproses sinyal suara dilakukan untuk meningkatkan kualitas data audio sebelum dilakukan ekstraksi fitur. Tahapan praproses meliputi normalisasi amplitudo untuk menyamakan skala intensitas suara serta penyesuaian durasi sinyal agar setiap masukan memiliki panjang yang seragam.

Proses praproses ini bertujuan untuk mengurangi pengaruh perbedaan volume, durasi bacaan, dan kondisi perekaman terhadap proses pemodelan dan hasil verifikasi bacaan berbasis suara.

2.3 Ekstraksi Fitur Suara

Ekstraksi fitur merupakan tahapan penting dalam sistem verifikasi suara karena bertujuan untuk merepresentasikan karakteristik sinyal audio dalam bentuk numerik. Pada penelitian ini digunakan fitur *Mel-Frequency Cepstral Coefficients* (MFCC) untuk merepresentasikan karakteristik spektral suara yang sesuai dengan persepsi pendengaran manusia [8],[9].

Selain itu, fitur *pitch* digunakan untuk menangkap informasi nada dasar suara yang berperan dalam membedakan pola bacaan dan variasi suara [3],[10]. Fitur-fitur yang dihasilkan kemudian disusun dalam bentuk sekuens sebagai masukan ke jaringan LSTM.

2.4 Arsitektur *Siamese* LSTM

Arsitektur yang digunakan pada penelitian ini adalah *Siamese* LSTM, yang terdiri dari dua jaringan LSTM identik dengan struktur dan bobot yang sama (*shared weights*) [4],[6],[7]. Masing-masing jaringan bertugas memetakan sinyal suara masukan ke dalam representasi vektor fitur.

Dengan penggunaan bobot yang sama, perbedaan vektor fitur yang dihasilkan sepenuhnya disebabkan oleh perbedaan sinyal suara yang dibandingkan. Pendekatan ini memungkinkan sistem mempelajari pola kemiripan bacaan suara secara langsung dan konsisten.

Secara matematis, Long Short-Term Memory (LSTM) terdiri dari beberapa gerbang (gate) yang berfungsi untuk mengatur aliran informasi pada data sekuens. Gerbang-gerbang tersebut meliputi forget gate, input gate, dan output gate, yang bekerja bersama dalam memperbarui nilai cell state dan hidden state. Mekanisme ini memungkinkan LSTM mempertahankan informasi penting dalam jangka waktu yang lebih panjang.

- *Forget Gate*

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

Keterangan:

f_t : nilai forget gate pada waktu ke- t

W_f : matriks bobot *forget gate*

h_{t-1} : hidden state pada waktu ke- $t-1$

x_t : vektor input pada waktu ke- t

b_f : bias *forget gate*

σ : fungsi aktivasi sigmoid (menghasilkan nilai antara 0 dan 1)

- *Input Gate*

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

Keterangan:

i_t : nilai input gate pada waktu ke- t

W_i : matriks bobot *input gate*

b_i : bias *input gate*

- Kandidat *Cell State*

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

Keterangan:

$\tilde{C}_t$: kandidat nilai cell state baru

W_c : matriks bobot kandidat memori

b_c : bias kandidat memori

- Pembaruan *Cell State*

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (4)$$

Keterangan:

C_t : *cell state* pada waktu ke- t

C_{t-1} : *cell state* pada waktu sebelumnya

$\odot$: operasi perkalian elemen (*Hadamard product*)

- *Output Gate*

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

Keterangan:

o_t : nilai *output gate* pada waktu ke- t

W_o : matriks bobot *output gate*

b_o : bias *output gate*

- Nilai *hidden state*

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

Keterangan:

h_t : *hidden state* pada waktu ke- t

C_t : *cell state* pada waktu ke- t

2.5 Pengukuran Kemiripan

Pengukuran kemiripan dilakukan dengan membandingkan vektor fitur keluaran dari kedua jaringan LSTM menggunakan fungsi jarak. Pada penelitian ini digunakan *Euclidean Distance* untuk menghitung jarak antara dua vektor fitur.

Nilai jarak yang lebih kecil menunjukkan tingkat kemiripan bacaan yang lebih tinggi, sedangkan nilai jarak yang lebih besar menunjukkan perbedaan bacaan yang lebih signifikan. Nilai jarak ini digunakan sebagai dasar dalam menentukan hasil verifikasi bacaan berbasis suara. Secara matematis, *Euclidean Distance* dirumuskan pada Persamaan (1).

$$d(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1,i} - x_{2,i})^2} \quad (7)$$

Keterangan :

x_1 : vektor fitur dari suara referensi

x_2 : vektor fitur dari suara uji

$x_{1,i} - x_{2,i}$: elemen ke- i dari masing-masing vektor fitur

n : jumlah dimensi vektor fitur

Nilai jarak yang lebih kecil menunjukkan tingkat kemiripan bacaan yang lebih tinggi.

2.6 Skema Pelatihan dan Pengujian

Skema pelatihan dilakukan dengan membentuk pasangan data suara yang diproses secara bersamaan oleh arsitektur *Siamese LSTM*. Pada tahap ini, model dilatih untuk mempelajari representasi fitur yang mampu membedakan pasangan bacaan dengan tingkat kemiripan tinggi dan rendah.

Pengujian dilakukan dengan memberikan pasangan suara yang belum pernah digunakan pada tahap pelatihan untuk mengamati kemampuan sistem dalam menghasilkan nilai kemiripan. Skema ini digunakan untuk memastikan bahwa sistem verifikasi bacaan berbasis suara dapat bekerja sesuai dengan tujuan sebagai implementasi awal metode.

2.7 Kriteria Penentuan Hasil Verifikasi

Nilai kemiripan yang dihasilkan oleh sistem digunakan sebagai dasar dalam menentukan hasil verifikasi bacaan. Secara umum, semakin kecil nilai jarak yang dihasilkan maka semakin tinggi tingkat kemiripan antara bacaan suara uji dan bacaan suara referensi.

Pada penelitian ini, penentuan hasil verifikasi masih bersifat deskriptif tanpa menetapkan nilai ambang batas tertentu. Oleh karena itu, sistem difokuskan pada analisis tingkat kemiripan suara sebagai implementasi awal penerapan metode *Siamese LSTM*.

3. HASIL DAN PEMBAHASAN

Hasil penerapan sistem verifikasi bacaan berbasis suara menunjukkan bahwa sistem mampu memproses dua masukan suara, yaitu suara referensi dan suara uji, melalui tahapan praproses sinyal, ekstraksi fitur, pemodelan menggunakan arsitektur *Siamese Long Short-Term Memory (LSTM)*, serta pengukuran tingkat kemiripan. Proses praproses yang dilakukan berupa normalisasi amplitudo dan penyesuaian durasi sinyal bertujuan untuk mengurangi perbedaan karakteristik dasar suara sebelum dilakukan ekstraksi fitur. Setelah melalui tahapan tersebut, sistem menghasilkan nilai jarak yang merepresentasikan tingkat kemiripan antara dua bacaan suara yang dibandingkan.

Untuk memperjelas proses pengujian yang dilakukan oleh sistem, data uji yang digunakan disajikan dalam bentuk pasangan bacaan suara referensi dan bacaan suara uji sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Contoh Dataset Berpasangan

No	Bacaan Referensi	Bacaan Uji	Keterangan
1	Ref_01.wav	Uji_01.wav	Bacaan Sama
2	Ref_01.wav	Uji_02.wav	Bacaan Beda
	Ref_01.wav	Uji_03.wav	Bacaan Beda
3	Ref_01.wav	Uji_04.wav	Bacaan Beda

4	Ref_01.wav	Uji_05.wav	Bacaan Beda
	Ref_02.wav	Uji_02.wav	Bacaan Sama
	Ref_02.wav	Uji_03.wav	Bacaan Beda
	Ref_02.wav	Uji_04.wav	Bacaan Beda
	Ref_02.wav	Uji_05.wav	Bacaan Beda
	Ref_03.wav	Uji_03.wav	Bacaan Sama
	Ref_03.wav	Uji_04.wav	Bacaan Beda
	Ref_03.wav	Uji_05.wav	Bacaan Beda
	Ref_04.wav	Uji_04.wav	Bacaan Sama
	Ref_04.wav	Uji_05.wav	Bacaan Beda
	Ref_05.wav	Uji_05.wav	Bacaan Sama

Dataset yang digunakan dalam penelitian ini berupa kumpulan data suara yang terdiri dari pasangan bacaan suara referensi dan bacaan suara uji. Dataset disusun dengan tujuan untuk mengamati kemampuan sistem dalam membandingkan tingkat kemiripan dua bacaan suara. Setiap pasangan data suara diproses melalui tahapan yang sama sehingga perbedaan hasil yang diperoleh sepenuhnya disebabkan oleh perbedaan karakteristik suara, bukan oleh perbedaan proses pemodelan.

Berdasarkan hasil pengujian secara kualitatif, pasangan bacaan suara yang memiliki pola pengucapan dan karakteristik akustik yang serupa cenderung menghasilkan nilai jarak yang lebih kecil dibandingkan pasangan bacaan suara yang memiliki perbedaan pola pengucapan yang signifikan. Nilai jarak hasil pengujian untuk beberapa pasangan suara disajikan pada Tabel 3.

Tabel 3. Nilai Jarak hasil Pengujian

No	Bacaan Referensi	Bacaan Uji	Nilai Jarak
1	Ref_01.wav	Uji_01.wav	0,31
2	Ref_01.wav	Uji_02.wav	0,46
	Ref_01.wav	Uji_03.wav	0,63
3	Ref_01.wav	Uji_04.wav	0,79
4	Ref_01.wav	Uji_05.wav	0,88

Hal ini menunjukkan bahwa representasi fitur yang dihasilkan oleh jaringan *Siamese LSTM* mampu menangkap kesamaan dan perbedaan pola suara antar bacaan. Dengan demikian, nilai jarak yang dihasilkan dapat digunakan sebagai dasar untuk menilai tingkat kecocokan bacaan suara dalam sistem verifikasi berbasis kemiripan.

Penggunaan fitur *Mel-Frequency Cepstral Coefficients* (MFCC) berperan penting dalam merepresentasikan karakteristik spektral suara yang relevan dengan persepsi pendengaran manusia. Sementara itu, fitur pitch membantu sistem dalam menangkap variasi nada dasar dan intonasi bacaan suara. Kombinasi kedua fitur tersebut memberikan representasi suara yang cukup informatif untuk mendukung proses pembandingan bacaan. Pendekatan ini menunjukkan bahwa verifikasi bacaan berbasis kemiripan suara dapat dilakukan tanpa melalui proses pengenalan teks secara eksplisit, sehingga lebih sederhana dibandingkan metode *speech recognition* berbasis transkripsi

Meskipun sistem mampu menghasilkan nilai kemiripan yang membedakan bacaan suara dengan tingkat kecocokan yang berbeda, hasil pengujian juga menunjukkan bahwa kondisi perekaman memiliki pengaruh terhadap nilai jarak yang dihasilkan. Untuk memberikan gambaran proses pengujian, dilakukan simulasi perhitungan nilai jarak menggunakan Euclidean Distance.

Contoh, apabila vektor fitur suara referensi adalah $X_1 = [0,25, 0,40, 0,35]$ dan vektor fitur suara uji adalah $X_2 = [0,30, 0,45, 0,80]$, maka nilai jarak yang dihasilkan adalah 0,46. Nilai tersebut menunjukkan bahwa perbedaan yang cukup besar akan menghasilkan nilai kemiripan yang lebih rendah. Seperti pada Tabel 3, Semakin kecil nilai jaraknya maka nilai kemiripannya besar.

Perbedaan kualitas rekaman, keberadaan noise, variasi jarak mikrofon, serta perbedaan tempo dan intonasi bacaan dapat menyebabkan variasi nilai kemiripan. Praproses sinyal yang diterapkan membantu mengurangi sebagian pengaruh tersebut, namun belum sepenuhnya menghilangkan variasi akibat perbedaan kondisi perekaman.

Penelitian ini masih bersifat implementasi awal, sehingga evaluasi performa sistem difokuskan pada analisis kualitatif terhadap kecenderungan nilai kemiripan yang dihasilkan. Penentuan hasil verifikasi belum menggunakan nilai ambang batas tertentu dan belum dilakukan pengukuran performa secara kuantitatif menggunakan metrik evaluasi seperti akurasi, presisi, atau recall. Meskipun demikian, hasil yang diperoleh menunjukkan bahwa

penerapan arsitektur *Siamese* LSTM memiliki potensi yang baik untuk dikembangkan lebih lanjut sebagai sistem verifikasi bacaan berbasis suara.

Secara keseluruhan, hasil dan pembahasan menunjukkan bahwa sistem verifikasi bacaan berbasis suara yang dibangun mampu menjalankan fungsi dasarnya dalam membandingkan dua bacaan suara berdasarkan tingkat kemiripan. Sistem ini dapat dijadikan sebagai dasar pengembangan lanjutan dengan penambahan evaluasi kuantitatif, optimasi parameter model, serta penggunaan dataset yang lebih besar dan beragam agar sistem menjadi lebih andal dan siap diterapkan pada skala yang lebih luas.

4. SIMPULAN

Berdasarkan hasil penerapan metode dan pembahasan yang telah dilakukan pada penelitian ini, dapat diambil beberapa simpulan sebagai berikut:

1. Metode *Siamese Long Short-Term Memory* (LSTM) dapat diterapkan pada sistem verifikasi bacaan berbasis suara untuk membandingkan dua masukan audio dan menghasilkan nilai kemiripan sebagai dasar proses verifikasi bacaan.
2. Sistem verifikasi bacaan yang dibangun mampu memproses pasangan suara referensi dan suara uji melalui tahapan praproses, ekstraksi fitur, pemodelan, dan pengukuran kemiripan secara terstruktur, sehingga menunjukkan bahwa alur sistem dapat berjalan sesuai dengan metodologi yang dirancang.
3. Penggunaan fitur suara *Mel-Frequency Cepstral Coefficients* (MFCC) dan *Pitch* memberikan representasi karakteristik suara yang relevan dalam menggambarkan pola bacaan, sehingga mendukung proses pembandingan bacaan berbasis kemiripan suara.
4. Kelebihan dari sistem yang dikembangkan adalah kemampuannya dalam melakukan verifikasi bacaan tanpa memerlukan proses pengenalan teks secara eksplisit, sehingga pendekatan berbasis kemiripan suara lebih sederhana dan sesuai untuk aplikasi verifikasi bacaan.
5. Keterbatasan penelitian ini terletak pada belum dilakukannya evaluasi performa sistem secara kuantitatif dan menyeluruh, sehingga hasil yang diperoleh masih bersifat implementasi awal dari penerapan metode yang digunakan.
6. Penelitian ini dapat dikembangkan lebih lanjut dengan melakukan evaluasi performa secara kuantitatif, menambahkan variasi fitur suara atau metode pemodelan lain, serta mengimplementasikan sistem pada konteks bacaan yang lebih spesifik dan skala data yang lebih besar.

5. SARAN

Berdasarkan hasil penelitian dan keterbatasan yang telah diidentifikasi, beberapa saran untuk pengembangan penelitian selanjutnya dapat disampaikan sebagai berikut:

1. Penelitian selanjutnya disarankan untuk melakukan evaluasi performa sistem secara kuantitatif menggunakan metrik evaluasi yang lebih komprehensif guna mengukur tingkat keandalan sistem verifikasi bacaan berbasis suara.
2. Pengembangan sistem dapat dilakukan dengan menambahkan atau mengombinasikan fitur suara lain selain *Mel-Frequency Cepstral Coefficients* (MFCC) dan *Pitch*, sehingga representasi karakteristik bacaan suara menjadi lebih optimal.
3. Metode pemodelan dapat dikembangkan lebih lanjut dengan mengeksplorasi arsitektur jaringan saraf lain atau melakukan penyempurnaan pada arsitektur *Siamese* LSTM untuk meningkatkan kemampuan pembelajaran pola kemiripan suara.
4. Sistem verifikasi bacaan berbasis suara ini dapat diuji pada objek bacaan yang lebih spesifik atau pada skala data yang lebih besar guna mengamati tingkat generalisasi metode yang digunakan.
5. Penelitian selanjutnya juga dapat mempertimbangkan implementasi sistem dalam bentuk aplikasi berbasis web atau perangkat bergerak agar sistem dapat digunakan secara lebih praktis dalam proses pembelajaran.

DAFTAR PUSTAKA

- [1] H. Zhao, W. Du, J. Guo, and G. Liu, "A Universal Identity Backdoor Attack against Speaker Verification based on Siamese Network," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2022-Sept, pp. 4770–4774, 2022, doi: 10.21437/Interspeech.2022-446.
- [2] A. Aloradi, W. Mack, M. Elminshaw, and E. A. P. Habets, "Speaker Verification in Multi-Speaker Kediri, 24 Januari 2026

- Environments Using Temporal Feature Fusion,” *Eur. Signal Process. Conf.*, vol. 2022-Augus, pp. 354–358, 2022, doi: 10.23919/eusipco55093.2022.9909705.
- [3] A. Wirdiani, S. N. Machetho, I. K. G. D. Putra, M. Sudarma, R. S. Hartati, and H. A. Ferdian, “Improvement Model for Speaker Recognition using MFCC-CNN and Online Triplet Mining,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 14, no. 2, pp. 420–427, 2024, doi: 10.18517/ijaseit.14.2.19396.
- [4] I. Dewa Agung Adwitya Prawangsa and A. Eka Karyawati, “Penerapan Metode MFCC dan LSTM untuk Speech Emotion Recognition,” *J. Elektron. Ilmu Komput. Udayana*, vol. 12, no. 4, pp. 2654–5101, 2024, [Online]. Available: www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess.
- [5] N. M. Almarshady, A. A. Alashban, and Y. A. Alotaibi, “Analysis and Investigation of Speaker Identification Problems Using Deep Learning Networks and the YOHO English Speech Dataset,” *Appl. Sci.*, vol. 13, no. 17, 2023, doi: 10.3390/app13179567.
- [6] R. De Prisco, C. Fusco, M. Iannucci, D. Malandrino, and R. Zaccagnino, “Text-independent voice recognition based on Siamese networks and fusion embeddings,” *CEUR Workshop Proc.*, vol. 3488, 2023.
- [7] X. Chen *et al.*, “Control of Raman Scattering Quantum Interference Pathways in Graphene,” *ACS Nano*, vol. 17, no. 6, pp. 5956–5962, 2023, doi: 10.1021/acsnano.3c00180.
- [8] S. Lata, “A Comparative Analysis of CNN-LSTM and MFCC-LSTM for Sentiment Recognition from Speech Signals,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 21s, p. 4392, 2024.
- [9] E. Abdul, W. Hachim, M. Talib Gaata, and T. Abbas, “Voice-Authentication Model Based on Deep Learning for Cloud Environment,” vol. 7, no. September, pp. 864–870, 2023, [Online]. Available: www.joiv.org/index.php/joiv
- [10] S. Manna, S. Nandy, and B. Roy, “Higher-order topological phases on fractal lattices,” *Phys. Rev. B*, vol. 105, no. 20, pp. 1–7, 2022, doi: 10.1103/PhysRevB.105.L201301.