PENGEMBANGAN PENGENDALI KARAKTER PERMAINAN MENGGUNAKAN SUARA REAL TIME BERBASIS LSTM

Alvin Arya Pangestu[1*], Oddy Virgantara Putra[1]
[1]*Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Darussalam Gontor, Jl. Raya Siman, Dusun I, Demangan, Kec. Siman, Kabupaten Ponorogo, Jawa Timur, Indonesia*

alvinaryapangestu80@student.cs.unida.gontor.ac.id
oddy@unida.gontor.ac.id

ABSTRAK

Penggunaan perintah suara sebagai pengendali karakter permainan menawarkan interaksi yang lebih natural. Namun, performa sistem ini sering dipengaruhi oleh gangguan akustik di lingkungan nyata. Penelitian ini mengembangkan dan mengevaluasi sistem pengendali karakter permainan berbasis suara *real-time* menggunakan model *Long Short Term Memory* (LSTM). Empat perintah utama *up, down, left,* dan *right* diproses menggunakan ekstraksi *Mel Frequency Cepstral Coefficients* (MFCC). Evaluasi dilakukan menggunakan pendekatan *multi-seed training* untuk memastikan hasil yang stabil dan tidak bergantung pada satu inisialisasi pelatihan. Sistem diuji pada tiga kondisi, yaitu tanpa *noise*, dengan *white noise*, dan dengan *pink noise*. Hasil menunjukkan bahwa LSTM mencapai akurasi rata-rata 96,50% pada *pink noise* dan 96,24% pada *white noise* dengan standar deviasi yang rendah, menandakan performa yang konsisten. Temuan ini menunjukkan bahwa LSTM cukup robust sebagai pengendali karakter permainan berbasis suara *real-time*, meskipun peningkatan ketahanan terhadap *noise* tetap diperlukan untuk penggunaan pada lingkungan akustik yang lebih kompleks.

Kata kunci: Pengenalan Suara, LSTM, noise akustik, MFCC.

1 PENDAHULUAN

Perkembangan teknologi interaksi manusia dan komputer mendorong penggunaan antarmuka yang lebih natural, salah satunya melalui perintah suara. Pada konteks permainan digital, penggunaan suara sebagai pengendali karakter menawarkan kemudahan dan pengalaman interaksi yang lebih intuitif dibandingkan perangkat input konvensional. Namun, implementasi pengendali berbasis suara di lingkungan nyata masih menghadapi tantangan utama berupa gangguan akustik, seperti noise latar belakang, yang dapat menurunkan akurasi sistem pengenalan suara(Wijaya, 2024). Performa *speech recognition* sangat menurun pada lingkungan ber-noise, sehingga menjadi tantangan penting untuk pengembangan sistem yang robust terhadap gangguan akustik (Li et al., 2023).

Berbagai penelitian telah mengkaji penerapan pembelajaran mendalam untuk pengenalan perintah suara. Long Short-Term Memory (LSTM) merupakan salah satu arsitektur jaringan saraf berulang yang banyak digunakan karena kemampuannya dalam memodelkan ketergantungan temporal pada data sekuensial, termasuk sinyal suara (Aini et al., 2023). Kombinasi fitur Mel-Frequency Cepstral Coefficients (MFCC) dan LSTM telah terbukti efektif dalam tugas klasifikasi perintah suara dengan kosakata terbatas (Putri et al., 2022). Pendekatan berbasis LSTM banyak diterapkan pada sistem real-time karena kemampuannya dalam memprediksi kondisi masa depan

dari data sekuensial, sehingga meningkatkan keandalan sistem pada lingkungan yang dinamis dan tidak ideal (Lashin et al., 2025). Meskipun demikian, performa model pengenalan suara sangat dipengaruhi oleh kondisi lingkungan akustik, terutama keberadaan noise yang bersifat acak maupun terstruktur.

Beberapa studi menunjukkan bahwa noise seperti white noise dan pink noise dapat memengaruhi kualitas fitur akustik yang diekstraksi dari sinyal suara dan berdampak pada performa model klasifikasi(Buono & Uliniansyah, 2025). Oleh karena itu, diperlukan evaluasi lebih lanjut mengenai ketahanan model LSTM terhadap berbagai jenis gangguan akustik, khususnya dalam konteks aplikasi real-time yang menuntut stabilitas dan konsistensi performa.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem pengendali karakter permainan berbasis suara real-time menggunakan model LSTM. Fokus penelitian diarahkan pada pengujian performa dan stabilitas model dalam mengenali empat perintah suara utama, yaitu *up*, *down*, *left*, dan *right*, yang dipilih karena mewakili perintah navigasi dasar pada pengendalian karakter permainan serta umum digunakan pada penelitian pengenalan perintah suara dengan kosakata terbatas.serta menganalisis ketahanan model terhadap gangguan akustik berupa *white noi*se dan *pink noise*. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pengendali permainan berbasis suara yang lebih andal dan robust untuk penggunaan di lingkungan nyata.

2 METODE

2.1 Dataset dan Lingkungan Penelitian

Dataset dalam penelitian ini terdiri dari empat perintah suara utama, yaitu *up, down, left,* dan *right*. Dataset perintah suara yang digunakan berasal dari *Simple Commands Dataset* yang tersedia secara publik melalui platform DagsHub dan disimpan dalam format WAV. Setiap sampel audio memiliki durasi pendek dengan frekuensi sampling sebesar 16 kHz dan direkam dalam satu kanal (mono), sehingga sesuai untuk tugas klasifikasi perintah suara dengan kosakata terbatas. Dataset ini berisi perintah suara diskret yang diucapkan oleh beberapa penutur dengan variasi intonasi, yang memungkinkan model mempelajari karakteristik temporal ucapan secara lebih representatif. Penggunaan dataset audio berdurasi pendek untuk pengenalan perintah suara telah umum diterapkan dalam penelitian *command recognition* karena mampu menyeimbangkan antara performa klasifikasi dan efisiensi komputasi (Sharif et al., 2023). Selain itu, penelitian sebelumnya menunjukkan bahwa sistem pengendali berbasis suara dengan jumlah kelas perintah terbatas dapat mencapai performa tinggi ketika dikombinasikan dengan ekstraksi fitur yang tepat dan model sekuensial seperti LSTM (Bakouri, 2022).Seluruh proses pemrosesan dan pelatihan dilakukan dengan menerapkan teknik pemrosesan sinyal digital untuk ekstraksi fitur suara, diikuti oleh pelatihan model *deep learning* berbasis sekuens guna mempelajari pola temporal dari sinyal audio. Pelatihan dilakukan dalam lingkungan komputasi GPU untuk mempercepat proses training.

2.2 Pra Pemrosesan Audio Menggunakan MFCC

Tahap pra pemrosesan dilakukan untuk mengubah sinyal suara mentah menjadi fitur numerik yang dapat dipelajari oleh model. Audio dinormalisasi dan dipotong secara seragam menggunakan *padding* dengan panjang maksimum 16 frame**.** Ekstraksi dilakukan menggunakan 40 *koefisien Mel-Frequency Cepstral Coefficients* (MFCC) karena jumlah koefisien tersebut mampu merepresentasikan karakteristik spektral sinyal suara secara lebih detail tanpa

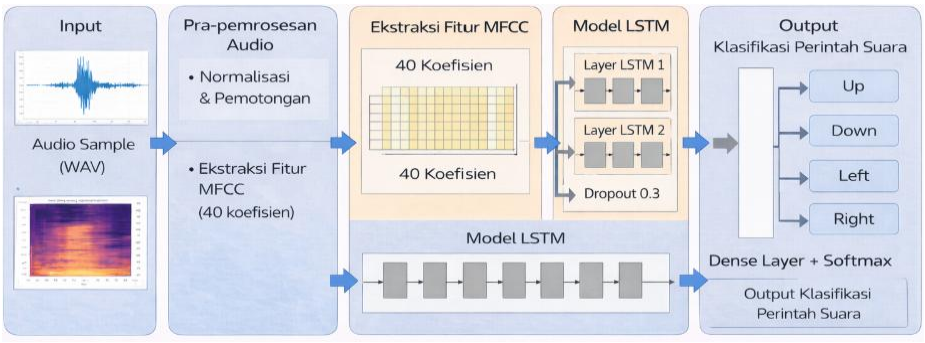
meningkatkan kompleksitas komputasi secara signifikan, sehingga sesuai untuk tugas klasifikasi perintah suara berbasis *deep learning*. Pendekatan ini banyak digunakan pada sistem pengenalan suara berbasis *deep learning* karena memberikan keseimbangan antara kualitas representasi fitur dan efisiensi komputasi(Rahman et al., 2025). Teknik MFCC dipilih karena merupakan standar dalam pengenalan suara dan terbukti efektif dalam menangkap karakteristik frekuensi sinyal ucapan(Sharif et al., 2023).Penelitian terbaru juga menunjukkan bahwa MFCC tetap menjadi fitur yang stabil dan kompetitif dalam sistem pengenalan suara berbasis *deep learning* (Hermanto & Sen, 2024). Hasil ekstraksi MFCC disusun dalam bentuk matriks dua dimensi yang digunakan sebagai input bagi model LSTM.

2.3 Penambahan Noise

Untuk mensimulasikan kondisi akustik nyata, penelitian ini menambahkan dua jenis noise pada data uji, yaitu *white noise* dan *pink noise*. *White noise* memiliki distribusi frekuensi yang merata sehingga sering digunakan untuk menguji sistem pengenalan suara pada kondisi gangguan acak (Xia et al., 2020). Sebaliknya, *pink noise* memiliki energi dominan pada frekuensi rendah dan lebih menyerupai suasana akustik lingkungan manusia (Christian & Tan, 2023).Kedua jenis noise ini digunakan untuk mengevaluasi ketahanan model LSTM terhadap gangguan audio yang umum terjadi pada penggunaan sistem suara secara *real-time*.

2.4 Arsitektur Model LSTM

Model utama yang digunakan untuk mengenali perintah suara *adalah* ***Long Short-Term Memory*** **(LSTM)**, yaitu jaringan saraf berulang yang mampu mempelajari pola temporal pada data sekuensial. LSTM banyak digunakan dalam penelitian *speech recognition* modern karena kemampuannya mempertahankan konteks jangka panjang (Sharif et al., 2023). Arsitektur model pada penelitian ini terdiri dari dua lapisan LSTM bertumpuk dengan 128 *hidden units*, *dropout* sebesar 0.3, dan keluaran berupa empat kelas perintah suara. Model dilatih menggunakan *optimizer* Adam, *batch size* 64, dan fungsi loss *CrossEntropyLoss*, yang umum digunakan pada klasifikasi audio berbasis *deep learning*. Total parameter model sekitar 147 ribu dengan ukuran model akhir sekitar 2.3 MB. Arsitektur model klasifikasi perintah suara menggunakan LSTM yang digunakan dalam penelitian ini ditunjukkan pada **Gambar 1**.



Gambar 1. Arsitektur Model Klasifikasi Perintah Suara Menggunakan LSTM

2.5 Evaluasi dan Multi Seed Training

Evaluasi performa model dilakukan menggunakan ***multi-seed training***, yaitu melatih model menggunakan beberapa nilai inisialisasi acak (seed 42, 123, dan 456) untuk memperoleh hasil yang lebih reliabel dan tidak bergantung pada satu kondisi awal. Pendekatan ini umum

digunakan dalam penelitian deep learning untuk mengukur kestabilan model serta variasi performa antar pelatihan (Sharif et al., 2023). Seluruh eksperimen dilakukan pada lingkungan komputasi berbasis Kaggle Notebook yang menyediakan akselerasi GPU. Eksperimen dijalankan menggunakan GPU NVIDIA Tesla T4 dengan kapasitas VRAM 16 GB**.** Proses pelatihan model menggunakan *learning rate* sebesar 0,001**,** *batch size* 64, dan dilakukan selama 50 *epoch* pada setiap percobaan untuk mencapai konvergensi yang stabil.

Pada setiap *seed*, sebagian data digunakan untuk pelatihan dan sebagian lainnya untuk pengujian. Metrik yang dihitung meliputi *akurasi, precision, recall, dan F1-score*, disertai perhitungan statistik seperti *mean* dan *standar deviasi* untuk menilai konsistensi performa. Evaluasi dilakukan pada tiga kondisi, yaitu tanpa *noise*, *white noise*, dan *pink noise*, guna menilai ketahanan LSTM terhadap gangguan akustik.

3 HASIL DAN PEMBAHASAN

3.1 Akurasi Model LSTM

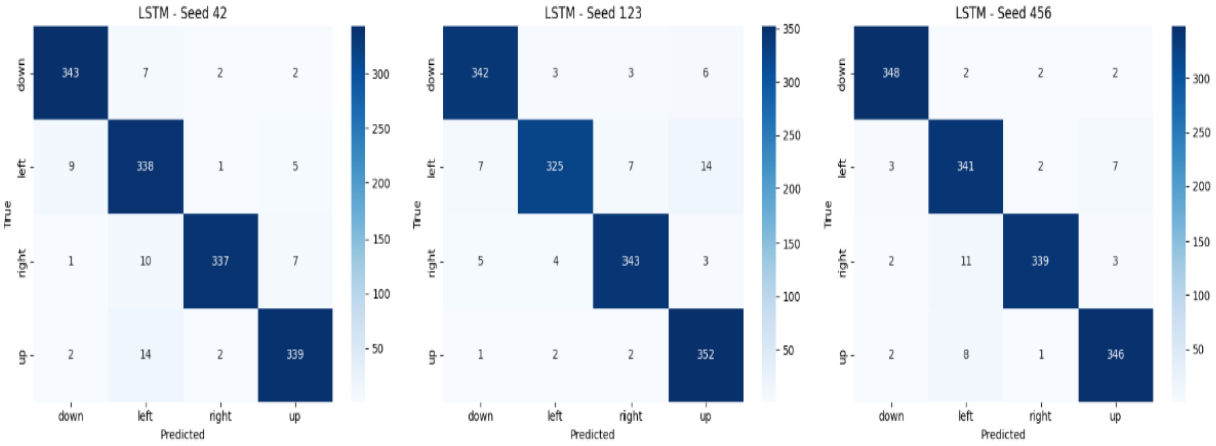
Evaluasi performa model Long Short-Term Memory (LSTM) dilakukan untuk mengukur kemampuan sistem dalam mengenali empat perintah suara, yaitu *up*, *down*, *left*, dan *right*. Pengujian dilakukan menggunakan tiga nilai inisialisasi acak yang berbeda (seed 42, 123, dan 456) untuk menilai konsistensi performa model terhadap variasi kondisi awal pelatihan.

Tabel 1. Ringkasan Statistik Performa Model LSTM

Metrik	Rata – Rata	Standar Deviasi	Minimum	Maximum
Accuracy	0,9615	0,0050	0,9563	0,9683
Precision	0,9620	0,0049	0,9570	0,9686
Recall	0,9615	0,0050	0,9563	0,9683
F1-score	0,9615	0,0050	0,9565	0,9683

Hasil evaluasi menunjukkan bahwa model LSTM memiliki performa yang tinggi dan stabil. Ringkasan statistik performa model ditampilkan pada **Tabel 1**. yang menunjukkan bahwa model mencapai akurasi rata-rata sebesar **0,9615** dengan standar deviasi **0,0050**. Nilai akurasi minimum tercatat sebesar **0,9563**, sedangkan nilai maksimum mencapai **0,9683**, yang mengindikasikan variasi performa yang relatif kecil antar seed. Selain akurasi, nilai **precision, recall, dan F1-score** juga menunjukkan hasil yang seimbang dengan rata-rata masing-masing berada di atas **0,96**.

Rendahnya nilai standar deviasi pada seluruh metrik menunjukkan bahwa model LSTM tidak sensitif terhadap perubahan inisialisasi parameter dan memiliki stabilitas pelatihan yang baik. Hal ini menandakan bahwa performa model tidak bergantung pada satu kondisi pelatihan tertentu, sehingga memiliki kemampuan generalisasi yang lebih baik.



Gambar 2. Confusion matrix hasil klasifikasi perintah suara menggunakan model LSTM

Analisis confusion matrix pada masing-masing seed menunjukkan bahwa sebagian besar prediksi berada pada diagonal utama, yang menandakan tingkat klasifikasi yang tinggi. Kesalahan prediksi yang terjadi relatif sedikit dan tidak menunjukkan pola kesalahan sistematis. Beberapa kesalahan klasifikasi ditemukan pada perintah *left* yang sesekali tertukar dengan *down* atau *up*, yang diduga disebabkan oleh kemiripan karakteristik fonetik dan variasi artikulasi antar penutur. Fenomena serupa juga dilaporkan pada penelitian pengenalan perintah suara berbasis ucapan pendek (Warden, 2018).

Secara keseluruhan, hasil ini menunjukkan bahwa model LSTM mampu memberikan performa klasifikasi yang akurat dan konsisten dalam mengenali perintah suara. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa LSTM efektif dalam memodelkan dependensi temporal pada sinyal suara dan cocok digunakan untuk tugas pengenalan perintah suara berbasis MFCC (Fadhilah & Prasetio, 2025)

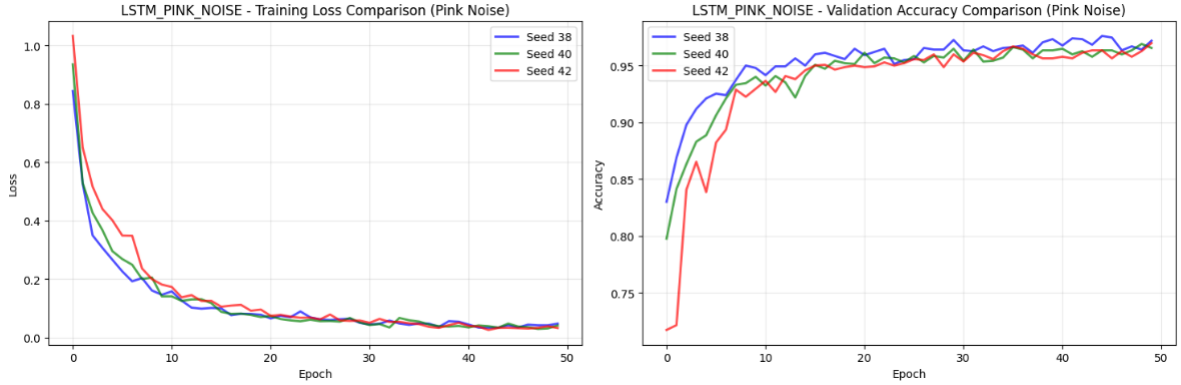
Ketahanan model LSTM terhadap gangguan akustik diuji dengan menambahkan **pink noise** dan **white noise** pada data pengujian. Evaluasi dilakukan menggunakan beberapa seed untuk memastikan konsistensi performa model. Metrik yang digunakan meliputi **akurasi, precision, recall, dan F1-score**.

Berdasarkan **Tabel 2**, model LSTM menunjukkan performa yang stabil pada kedua kondisi noise. Pada pink noise, akurasi rata-rata mencapai **96,50%** dengan standar deviasi **0,37%**, sedangkan pada white noise akurasi rata-rata sebesar **96,24%** dengan standar deviasi **0,26%**. Nilai standar deviasi yang rendah menunjukkan bahwa performa model konsisten dan tidak sensitif terhadap variasi pelatihan.

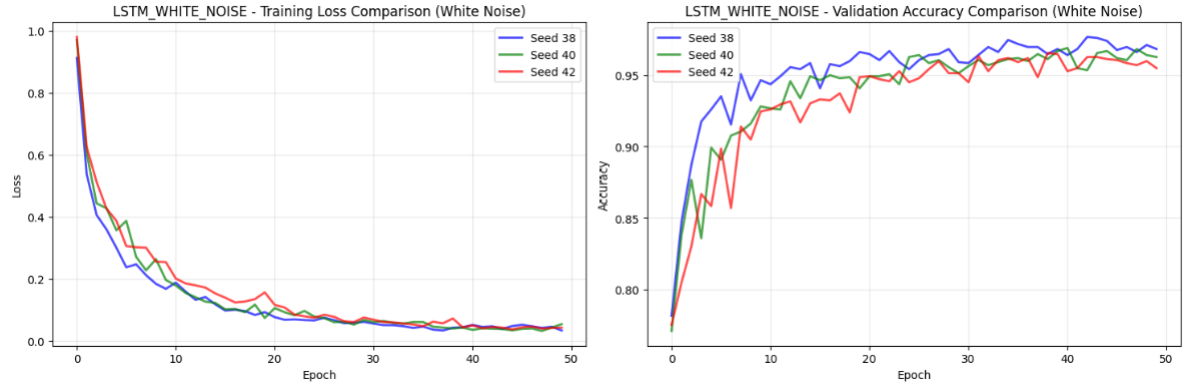
Tabel 2. Ringkasan Statistik Performa LSTM pada Kondisi Noise

Noise Type	Akurasi	Precision	Recall	F1-Score	Std Dev
Pink Noise	96,50	96,52	96,50	96,50	0,37
White Noise	96,24	96,26	96,24	96,24	0,26

Selain itu, **Gambar 3** dan **Gambar 4** memperlihatkan bahwa nilai loss menurun dan akurasi validasi meningkat secara stabil hingga epoch akhir pada kedua jenis noise. Hal ini menunjukkan bahwa model mampu belajar dengan baik tanpa indikasi overfitting.



Gambar 3. Kurva Pelatihan Model LSTM pada Kondisi Pink Noise



Gambar 4. Kurva Pelatihan Model LSTM pada Kondisi White Noise

Selain itu Secara keseluruhan, hasil ini menunjukkan bahwa model LSTM cukup robust terhadap gangguan noise akustik, baik yang bersifat spektrum merata (white noise) maupun dominan frekuensi rendah (pink noise). Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa kombinasi MFCC dan LSTM efektif untuk pengenalan suara pada kondisi bising (Le & Feng, 2024).

Analisis stabilitas dilakukan untuk menilai konsistensi performa model LSTM terhadap variasi inisialisasi pelatihan. Evaluasi dilakukan menggunakan tiga seed berbeda, yaitu 42, 123, dan 456. Hasil pengujian menunjukkan bahwa model LSTM memiliki performa yang relatif konsisten pada setiap percobaan.

Tabel 3. Performa Model LSTM pada Berbagai Seed

Seed	Accuracy	Precision	Recall	F1-Score
42	95,63	95,70	95,63	95,65
123	95,98	96,03	95,98	95,97
456	96,83	96,86	96,83	96,83
Rata – Rata	96,15	96,20	96,15	96,15
Std. Dev	0,50	0,49	0,50	0,50

Berdasarkan **Tabel 3**, model LSTM memperoleh **akurasi rata-rata sebesar 96,15% dengan standar deviasi 0,50%**, yang menunjukkan variasi performa yang rendah antar-seed. Pola serupa juga terlihat pada metrik precision, recall, dan F1-score. Nilai standar deviasi yang kecil mengindikasikan bahwa model tidak sensitif terhadap perubahan inisialisasi parameter dan memiliki stabilitas pelatihan yang baik.

Hasil ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa penggunaan beberapa seed merupakan pendekatan yang efektif untuk mengevaluasi konsistensi dan keandalan model deep learning pada tugas pengenalan suara (Bethard, 2022).

4 KESIMPULAN

Penelitian ini berhasil mengembangkan sistem pengendali karakter permainan berbasis suara real-time menggunakan model Long Short-Term Memory (LSTM) dengan empat perintah utama, yaitu *up*, *down*, *left*, dan *right*. Berdasarkan hasil evaluasi, model LSTM menunjukkan kinerja yang tinggi dan konsisten, dengan akurasi rata-rata sebesar **96,15%** serta variasi performa yang rendah pada beberapa inisialisasi pelatihan. Hal ini menunjukkan bahwa model mampu melakukan generalisasi dengan baik dan tidak bergantung pada satu kondisi pelatihan tertentu.

Pengujian pada kondisi gangguan akustik menunjukkan bahwa model LSTM tetap mempertahankan performa yang baik pada *pink noise* dan *white noise*, sehingga membuktikan bahwa pendekatan yang digunakan relevan untuk lingkungan penggunaan nyata yang tidak sepenuhnya bebas dari gangguan suara. Dengan demikian, tujuan penelitian untuk mengevaluasi kelayakan LSTM sebagai pengendali karakter permainan berbasis suara real-time telah tercapai.

Sebagai pengembangan selanjutnya, penelitian ini dapat ditingkatkan dengan menerapkan metode reduksi noise, atau pengujian pada jenis gangguan akustik yang lebih kompleks guna meningkatkan ketahanan sistem pada lingkungan yang lebih dinamis.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Darussalam Gontor atas dukungan fasilitas dan lingkungan akademik yang mendukung pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada dosen pembimbing yang telah memberikan arahan, masukan, dan bimbingan selama proses penelitian hingga penulisan naskah ini. Selain itu, penulis mengapresiasi seluruh pihak yang telah membantu secara langsung maupun tidak langsung dalam penyelesaian penelitian ini.

DAFTAR PUSTAKA

Aini, N., Asri, L., Adam, R. I., & Dermawan, B. A. (2023). *SPEECH RECOGNITION UNTUK KLASIFIKASI PENGUCAPAN NAMA HEWAN DALAM BAHASA SUNDA MENGGUNAKAN METODE LONG-SHORT TERM MEMORY*. *7*(2), 1242–1247.

Bakouri, M. (2022). Development of Voice Control Algorithm for Robotic Wheelchair Using MIN and LSTM Models. *Computers, Materials and Continua*, *73*(2), 2441–2456. https://doi.org/10.32604/cmc.2022.025106

Bethard, S. (2022). *We need to talk about random seeds*. http://arxiv.org/abs/2210.13393

Buono, A., & Uliniansyah, M. T. (2025). *Pengembangan model akustik dengan deep neural network untuk sistem pengenalan wicara bahasa Indonesia*. *22*(1), 84–100.

Christian, Y., & Tan, C. (2023). KOMPARASI DAN ANALISIS AI BASE NOISE SUPPRESSION: STUDI KASUS RTX VOICE COMPARISON AND ANALYSIS ON AI BASE NOISE SUPPRESSION: STUDY CASE OF RTX VOICE. *Journal of Information Technology and Computer Science (INTECOMS)*, *6*(2).

Fadhilah, K., & Prasetio, B. H. (2025). *Pengembangan Sistem Smart Home Berbasis Pengenalan Suara Menggunakan Model Long Short-Term Memory*. *9*(2), 1–8.

Lashin, M., El-mashad, S. Y., & Elgammal, A. T. (2025). Real-time path planning in dynamic environments using LSTM-augmented A∗ search. *Results in Engineering*, *27*. https://doi.org/10.1016/j.rineng.2025.106324

Le, N. C. J., & Feng, L. (2024). *Self-Organization Towards $1/f$ Noise in Deep Neural Networks*. http://arxiv.org/abs/2301.08530

Li, D., Gao, Y., Zhu, C., Wang, Q., & Wang, R. (2023). Improving Speech Recognition Performance in Noisy Environments by Enhancing Lip Reading Accuracy. *Sensors*, *23*(4). https://doi.org/10.3390/s23042053

Putri, H. M., Fadlisyah, F., & Fuadi, W. (2022). Pendeteksian Bahasa Isyarat Indonesia Secara Real-Time Menggunakan Long Short-Term Memory (Lstm). *Jurnal Teknologi Terapan and Sains 4.0*, *3*(1), 663. https://doi.org/10.29103/tts.v3i1.6853

Rahman, F. N., Listyorini, T., & Supriyati, E. (2025). *ANALISIS AKURASI CNN PADA DATA OLAH SUARA MANUSIA MENGGUNAKAN PARAMETER KOEFISIEN MFCC DAN MAX LENGTH* (Vol. 15, Issue 1).

Sharif, A., Sitompul, O. S., & Nababan, E. B. (2023). Analysis Of Variation In The Number Of MFCC Features In Contrast To LSTM In The Classification Of English Accent Sounds. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, *6*(2), 587–601. https://doi.org/10.31289/jite.v6i2.8566

Warden, P. (2018). *Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition*. http://arxiv.org/abs/1804.03209

Wijaya, H. (2024). Teknologi Pengenalan Suara tentang Metode, Bahasa dan Tantangan: Systematic Literature Review. *Bit-Tech*, *7*(2), 533–544. https://doi.org/10.32877/bt.v7i2.1888

Xia, L., Chen, G., Xu, X., Cui, J., & Gao, Y. (2020). *Audiovisual speech recognition: A review and forecast*. International Journal of Advanced Robotic Systems, 17(6), 1–17. https://doi.org/10.1177/1729881420976082