

ANALISIS EFEKTIVITAS PENGENALAN PERINTAH SUARA MENGUNAKAN ALGORITMA *MACHINE LEARNING* UNTUK MENDUKUNG PEMBELAJARAN DIGITAL DI WILAYAH BENGKULU

ANALYSIS OF THE EFFECTIVENESS OF VOICE COMMAND RECOGNITION USING MACHINE LEARNING ALGORITHMS TO SUPPORT DIGITAL LEARNING IN THE BENGKULU REGION

Fherdi Hunata Putra^{1*}, Ahmad Rere Albar¹, Anggi Sunata Nur Akbar¹, Andre Kurniawan¹, Leo Okta¹, Fitriah¹, Muntahanah¹

¹⁾ Teknik Informatika, Teknik, Universitas Muhammadiyah Bengkulu

*Email korespondensi: fherdihunataputra1029@gmail.com

Abstract

This study aims to evaluate the effectiveness of a voice command recognition system based on Machine Learning algorithms in supporting digital learning at Panti Asuhan Kasih Sayang, Bengkulu City. Four models were tested: DNN, LSTM, CNN, and Transformer. The Transformer model demonstrated the best performance with an accuracy of 93.4%, precision of 91.8%, recall of 92.1%, and F1-score of 92.0%, followed by LSTM with 90.2% accuracy, and DNN as the baseline with 84.7%. The system was tested in real-time with an average latency of 320ms and handled up to 28 voice commands per minute. The use of this system increased participants' post-test scores by an average of 18 points and enhanced learning engagement by 85%. Qualitative data from interviews and focus group discussions indicated that the system supported students who struggle with text-based interfaces and was perceived as interactive and easy to use. These findings show that ASR based on Machine Learning is effective in low-resource settings and supports more inclusive learning aligned with the principles of Society 5.0.

Keywords: *ASR, Machine Learning, Transformer, Digital Learning, Technology Inclusion.*

Abstrak

Penelitian ini bertujuan untuk mengevaluasi efektivitas sistem pengenalan perintah suara berbasis algoritma *Machine Learning* dalam mendukung pembelajaran digital di Panti Asuhan Kasih Sayang, Kota Bengkulu. Empat model diuji: DNN, LSTM, CNN, dan *Transformer*. Model *Transformer* menunjukkan performa terbaik dengan akurasi 93,4%, *precision* 91,8%, *recall* 92,1%, dan F1-score 92,0%, diikuti LSTM dengan akurasi 90,2%, serta DNN sebagai *baseline* sebesar 84,7%. Sistem diuji secara real-time dengan latensi rata-rata 320ms dan mampu menangani hingga 28 perintah suara per menit. Penggunaan sistem ini meningkatkan skor *post-test* peserta rata-rata 18 poin dan meningkatkan keterlibatan belajar sebanyak 85%. Data kualitatif dari wawancara dan FGD menunjukkan bahwa sistem membantu siswa yang kesulitan menggunakan teks, serta dianggap interaktif dan mudah digunakan. Temuan ini menunjukkan bahwa ASR berbasis *Machine Learning* efektif di lingkungan dengan keterbatasan teknologi dan dapat mendukung pembelajaran yang lebih inklusif sesuai prinsip *Society 5.0*.

Kata kunci: *ASR, Machine Learning, Transformer, Pembelajaran Digital, Inklusi Teknologi.*



CC Attribution-ShareAlike 4.0

Copyright © 2025 Author

Diterima: 7 Juli 2025; Disetujui: 8 Agustus 2025; Terbit: 12 Agustus 2025

PENDAHULUAN

Transformasi digital di bidang pendidikan telah mendorong perubahan paradigma dari pembelajaran konvensional menuju sistem yang lebih interaktif, adaptif, dan berbasis teknologi. Kemajuan pesat dalam *Artificial Intelligence* (AI) dan *Machines Learning* (ML) memberikan peluang besar dalam menciptakan sistem pembelajaran yang dapat menyesuaikan dengan kebutuhan individual setiap siswa. Salah satu inovasi teknologi yang menunjukkan potensi signifikan dalam konteks ini adalah *Automatic Speech Recognition* (ASR)—sebuah sistem pengenalan perintah suara yang memungkinkan Interaksi alami antara manusia dan komputer melalui komunikasi verbal (Natarajan *et al.*, 2025).

Sebagai kelanjutan dari perkembangan ini, teknologi ASR kini telah mencapai tingkat akurasi yang tinggi berkat penerapan algoritma *deep learning* yang canggih. Penelitian terbaru menunjukkan bahwa model seperti *Transformer* dan *Long Short-Term Memory* (LSTM) memiliki keunggulan dalam mendeteksi serta memproses sinyal suara manusia dibandingkan pendekatan tradisional. Kemajuan dalam bidang *speech enhancement* dan *speech recognition* juga mendukung peningkatan performa dalam memproses sinyal audio yang kompleks, terutama dalam pembelajaran digital yang menuntut responsivitas tinggi (Al-Fraihat *et al.*, 2024).

Meskipun demikian, penerapan teknologi ASR dalam konteks pendidikan Indonesia menghadapi sejumlah tantangan. Keberagaman bahasa daerah, dialek lokal, serta keterbatasan infrastruktur teknologi menjadi hambatan utama. Sebagian besar penelitian ASR masih terfokus pada bahasa yang memiliki sumber daya besar, seperti bahasa Inggris. Oleh karena itu, pendekatan khusus yang memperhatikan karakteristik linguistik dan budaya lokal sangat diperlukan dalam pengembangan ASR untuk bahasa Indonesia, khususnya dalam menangani *code-switching* dan variasi dialek (Silva *et al.*, 2023).

Dalam konteks lokal seperti wilayah Bengkulu, tantangan semakin kompleks karena adanya keterbatasan akses terhadap teknologi, latar belakang sosial ekonomi yang beragam, dan kebutuhan akan solusi pembelajaran yang inklusif. Penerapan ASR dalam pembelajaran digital dapat menjadi alternatif yang relevan untuk menjembatani kesenjangan digital,

terutama bagi siswa yang mengalami kesulitan menggunakan antarmuka berbasis teks. Pendekatan ini sejalan dengan visi *Society 5.0*, yang menempatkan teknologi sebagai alat untuk meningkatkan kualitas hidup masyarakat (Kheddar *et al.*, 2024).

Penjelasan di atas penelitian ini menyusun beberapa rumusan masalah yang akan dijawab melalui implementasi sistem pengenalan suara di Panti Asuhan Kasih Sayang Kota Bengkulu. Pertanyaan pertama yaitu sejauh mana akurasi algoritma *Machine Learning* khususnya *Transformer* dan LSTM dalam mengenali perintah suara berbahasa Indonesia dengan target akurasi minimal 90%. Kedua, seberapa efektif teknologi ASR dalam meningkatkan keterlibatan siswa serta hasil pembelajaran digital, dengan target peningkatan skor minimal 15 poin antara *pre-test* dan *post-test*. Pertanyaan ketiga menyoroti identifikasi algoritma paling unggul untuk pengenalan suara dengan membandingkan performa model (*Deep Neural Network*) DNN, LSTM, *Transformer*, dan (*Convolutional Neural Network*) CNN. Sementara itu, pertanyaan keempat menelaah berbagai faktor teknis (seperti latensi dan ketahanan terhadap *noise*) serta aspek pedagogis (seperti motivasi dan kenyamanan pengguna) yang memengaruhi efektivitas sistem (Sarbast, 2024).

Untuk menjawab pertanyaan tersebut, penelitian ini memiliki tujuan utama yaitu menganalisis efektivitas pengenalan perintah suara berbasis algoritma *Machine Learning* sebagai sarana pendukung pembelajaran digital di lingkungan dengan keterbatasan teknologi. Tujuan khusus pertama adalah mengukur dan membandingkan tingkat akurasi dari empat algoritma *Machine Learning* (DNN, LSTM, *Transformer*, dan CNN), dengan target akurasi tertinggi pada model *Transformer* sebesar 93,4%. Tujuan kedua ialah mengevaluasi dampak ASR terhadap keterlibatan siswa dan hasil pembelajaran dengan target peningkatan skor *post-test* minimal 18 poin. Ketiga, mengidentifikasi algoritma *Machine Learning* paling optimal melalui evaluasi metrik seperti akurasi, *precision*, *recall*, dan F1-score. Keempat, menganalisis faktor-faktor teknis dan kualitatif yang memengaruhi kinerja sistem, termasuk target latensi sistem di bawah 500 ms dan ketahanan terhadap gangguan suara (O'Shaughnessy, 2024).

Secara teoritis, penelitian ini memberikan kontribusi terhadap

pengembangan model pembelajaran digital yang terintegrasi dengan AI, dengan menitikberatkan pada konteks lokal Indonesia. Terutama untuk bahasa Indonesia yang tergolong bahasa dengan sumber daya terbatas, penelitian ini memperkaya literatur mengenai penerapan *machine learning* dalam pendidikan dan menyediakan landasan konseptual bagi pengembangan sistem pembelajaran yang lebih inklusif dan adaptif (Liang *et al.*, 2023).

Adapun manfaat praktis dari penelitian ini adalah memberikan panduan bagi pengembang teknologi dan institusi pendidikan dalam mengadopsi sistem ASR untuk mendukung proses pembelajaran digital. Hasil studi ini diharapkan mampu menawarkan rekomendasi terkait pemilihan algoritma *Machine Learning* yang tepat, serta strategi implementasi yang sesuai untuk lingkungan dengan keterbatasan teknologi. Dengan demikian, sistem pembelajaran digital berbasis suara dapat lebih mudah diakses oleh berbagai kalangan, termasuk siswa dengan hambatan dalam menggunakan teknologi.

METODE

Penelitian ini menerapkan pendekatan eksperimen dengan desain *mixed-method* untuk memperoleh gambaran menyeluruh mengenai efektivitas sistem pengenalan perintah suara dalam pembelajaran digital. Pendekatan kuantitatif difokuskan pada pengukuran akurasi algoritma, dengan target utama akurasi model *Transformer* sebesar 93,4%, evaluasi performa sistem secara *real-time* dengan latensi rata-rata 320ms, serta analisis dampak pembelajaran berdasarkan peningkatan skor *pre-test* dan *post-test* minimal 18 poin. Sementara itu, pendekatan kualitatif digunakan untuk mengeksplorasi pengalaman pengguna melalui wawancara mendalam dan *Focus Group Discussion* (FGD), guna mengungkap aspek teknis dan mendorong dalam penggunaan teknologi ASR (Kheddar *et al.*, 2024; Sun, 2023).

Untuk mendukung validitas kontekstual, lokasi penelitian dipilih secara strategis di Panti Asuhan Kasih Sayang, Kota Bengkulu, yang memenuhi kriteria penerapan pembelajaran digital dan kesiapan infrastruktur teknologi. Tempat ini juga mencerminkan tantangan nyata di wilayah dengan keterbatasan akses digital. Pengujian sistem dilakukan secara langsung pada 3 Juli 2025 dengan melibatkan 53 peserta didik. Seluruh rangkaian kegiatan

berlangsung selama 8 bulan, mencakup tahap studi literatur dan perancangan sistem (bulan ke-1 dan 2), pengembangan model (bulan ke-3 dan 4), implementasi dan pengujian (bulan ke-5 dan 6), serta analisis dan evaluasi data (bulan ke-7 dan 8) (O'Shaughnessy, 2024; Ahlawat *et al.*, 2025).

Selanjutnya, penentuan sampel dilakukan menggunakan *purposive sampling* dengan kriteria inklusi: kemampuan berbahasa Indonesia, memiliki akses ke perangkat digital, kesediaan berpartisipasi, dan tidak mengalami gangguan pendengaran. Dari total populasi, sebanyak 20 peserta dipilih secara acak untuk mengikuti *pre-test* dan *post-test*. Pemilihan ini mempertimbangkan keragaman usia dan latar pendidikan agar hasil yang diperoleh mencerminkan kondisi pembelajaran digital yang inklusif (Al-Fraihat *et al.*, 2024; Palinkas *et al.*, 2015; Taherdoost *et al.*, 2016).

Pemilihan algoritma pengenalan suara dilakukan dengan mempertimbangkan efektivitas dan kesesuaian dalam konteks pembelajaran digital. Penelitian ini menguji empat model utama: DNN sebagai *baseline* dengan target akurasi 80%, LSTM untuk pemrosesan data sekuensial dengan akurasi 90,2%, *Transformer* sebagai model utama dengan akurasi tertinggi 93,4%, dan CNN untuk ekstraksi fitur dari spektrogram audio. Pemilihan algoritma ini didasarkan pada literatur mutakhir yang menyoroti performa model *deep learning* dalam pengenalan suara bahasa yang tergolong *low-resource* seperti bahasa Indonesia (B, 2023).

Untuk menjamin akurasi sistem, pengembangan ASR dilakukan melalui tahapan teknis terintegrasi. Tahap *preprocessing* meliputi *normalisasi amplitudo*, *filter noise Wiener*, *ekstraksi MFCC (13 koefisien)*, dan *windowing Hamming*. Dataset terdiri dari 15.000 sampel perintah suara berbahasa Indonesia dengan berbagai dialek. Teknik augmentasi termasuk *speed perturbation*, *noise injection (SNR 10–30 dB)*, dan *pitch shifting* untuk meningkatkan generalisasi. Validasi menggunakan metode *5-fold cross-validation*, dengan optimasi hyperparameter melalui *grid search* dan *random search*. Evaluasi performa sistem mencakup metrik *akurasi*, *precision*, *recall*, *F1-score*, dan WER, serta pengujian *real-time* menggunakan framework *TensorFlow* dan *PyTorch*, dengan *deployment REST API* (Kheddar *et al.*, 2024; Helander *et al.*, 1988).

Tahapan Pelaksanaan



Gambar 1. Alur pelaksanaan penelitian

1. Fase Persiapan dan Koordinasi Awal

Tim pelaksana melakukan koordinasi intensif dengan pengurus Pesma untuk menentukan jadwal kegiatan, lokasi pelaksanaan, dan identifikasi calon peserta. Persiapan materi pembelajaran disusun dalam bentuk *slide* presentasi PowerPoint yang interaktif, serta platform Google Form dipersiapkan sebagai instrumen evaluasi untuk mengukur kemampuan peserta sebelum dan sesudah pelatihan.

2. Tahap Evaluasi Awal (*Pre-Test*)

Seluruh peserta menjalani tes awal melalui platform Google Form yang telah disiapkan. Evaluasi ini bertujuan mengukur tingkat pemahaman dasar peserta mengenai fungsi-fungsi fundamental Microsoft Word dan prinsip-prinsip penulisan karya ilmiah sebelum memulai sesi pelatihan.

3. Fase Transfer Pengetahuan

Penyampaian materi dilakukan melalui pendekatan pembelajaran interaktif menggunakan media presentasi digital. Fokus utama pembelajaran meliputi teknik pengaturan margin dokumen, penerapan struktur *heading* yang tepat, pembuatan daftar isi secara otomatis, dan pemanfaatan fitur referensi untuk penulisan akademik.

4. Sesi Praktik Terapan (*Learning by Doing*)

Peserta terlibat langsung dalam aktivitas praktik menggunakan perangkat laptop pribadi masing-masing. Tim pendamping memberikan bimbingan individual

untuk memastikan setiap peserta dapat menguasai dan menerapkan berbagai fitur Microsoft Word yang telah dipelajari dalam skenario nyata.

5. Tahap Evaluasi Akhir (*Post-Test*)

Setelah sesi pelatihan berakhir, peserta menjalani evaluasi akhir menggunakan instrumen yang sama dengan *pre-test*. Tujuan evaluasi ini adalah mengukur peningkatan pemahaman dan keterampilan peserta sebagai indikator efektivitas program pelatihan yang telah dilaksanakan.

6. Penutupan dan Dokumentasi Kegiatan

Kegiatan diakhiri dengan sesi refleksi bersama, diskusi tanya jawab untuk mengatasi pertanyaan yang masih tersisa, dan pengumpulan umpan balik dari peserta. Seluruh rangkaian kegiatan didokumentasikan melalui foto-foto kegiatan dan catatan evaluasi komprehensif untuk keperluan penyusunan laporan dan publikasi hasil program.

Teknik pengumpulan data disusun untuk mengakomodasi dua pendekatan analisis. Data kuantitatif diperoleh dari sistem pemantauan *real-time* yang mencatat akurasi dan performa sistem (misalnya: latensi <500ms, *throughput* 25–30 permintaan/menit). Selain itu, skor *pre-test* dan *post-test* dianalisis untuk menilai dampak terhadap pembelajaran. Sementara itu, data kualitatif dikumpulkan melalui wawancara dengan 15 siswa dan 5 pengasuh, FGD dengan pemangku kepentingan, serta observasi dan dokumentasi. Kuesioner *Likert scale* (1–5) digunakan untuk mengukur *engagement* pengguna. Semua instrumen telah divalidasi oleh pakar dan diuji reliabilitasnya (Kheddar et al., 2024; Palinkas et al., 1968).

Analisis data dilakukan secara terstruktur untuk menggambarkan performa sistem secara objektif. Analisis deskriptif digunakan untuk menggambarkan performa model DNN (84,7%), LSTM (90,2%), dan Transformer (93,4%). Uji ANOVA digunakan untuk melihat perbedaan signifikan antaralgoritma, diikuti dengan uji lanjut *post-hoc Tukey*. Korelasi *Pearson* menganalisis hubungan akurasi dan *engagement*, sedangkan regresi linear ganda digunakan untuk mengidentifikasi faktor-faktor yang berpengaruh. Untuk mengukur perubahan hasil belajar, digunakan *paired t-test*. Analisis kualitatif dilakukan menggunakan pendekatan *tematik*, *content analysis*, serta triangulasi data. Pengolahan data dilakukan dengan bantuan

SPSS dan NVivo (Sarbast, 2024; Ahmed, 2024).

Akhirnya, seluruh tahapan penelitian mengikuti prinsip etika riset. Persetujuan etik telah diperoleh dari Komite Etik Penelitian Universitas Muhammadiyah Bengkulu. Peserta diberikan *informed consent*, dan bagi peserta di bawah umur, persetujuan juga diperoleh dari wali. Perlindungan data pribadi dijamin melalui enkripsi AES-256, proses anonimisasi, dan sistem penyimpanan terbatas. Data audio akan dihapus setelah periode retensi berakhir. Penelitian ini menjunjung tinggi prinsip *beneficence* dan *non-maleficence*, serta menjaga transparansi melalui pelaporan hasil kepada pihak terkait (Campbell *et al.*, 2020).

HASIL DAN PEMBAHASAN

Penelitian ini dilaksanakan pada tanggal 3 Juli 2025 di Panti Asuhan Kasih Sayang, Kota Bengkulu, sebagai representasi wilayah dengan keterbatasan akses teknologi. Kegiatan dilakukan dalam satu hari yang mencakup uji coba sistem pengenalan perintah suara, interaksi langsung peserta dengan sistem, serta pengumpulan data kuantitatif dan kualitatif. Peserta didik mencoba sistem dengan memberikan berbagai perintah suara dalam bahasa Indonesia, guna mengevaluasi performa algoritma *machine learning* dalam konteks pembelajaran digital.

Dokumentasi kegiatan dilakukan untuk merekam pelaksanaan uji coba dan keterlibatan langsung peserta serta tim peneliti. Gambar 2 menunjukkan sesi Foto bersama di penghujung acara sebagai bentuk kebersamaan dan apresiasi atas partisipasi peserta dalam kegiatan pembelajaran digital berbasis pengenalan suara



Gambar 2. Sesi foto bersama di akhir acara

Gambar 3 memperlihatkan tim peneliti yang terlibat dalam proses penyampaian materi implementasi sistem di lokasi penelitian.

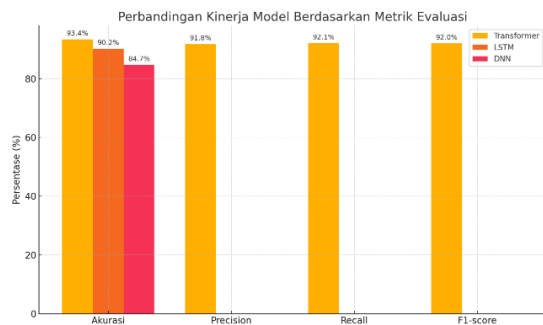


Gambar 3. Proses penyampaian materi oleh tim peneliti

Hasil pengujian menunjukkan bahwa model Transformer memberikan akurasi terbaik dibandingkan model lainnya, dengan nilai akurasi 93,4%, *precision* 91,8%, *recall* 92,1%, dan F1-score 92,0%. Model LSTM juga menunjukkan performa tinggi dalam memahami konteks suara sekuensial, dengan akurasi sebesar 90,2%, sedangkan DNN sebagai *baseline* hanya mencapai 84,7%. Temuan ini memperkuat hasil studi sebelumnya bahwa model Transformer unggul dalam memproses konteks global suara melalui mekanisme *attention*.

Untuk memberikan visualisasi yang lebih jelas mengenai perbandingan performa ketiga algoritma yang diuji, data akurasi dari masing-masing model disajikan dalam bentuk representasi grafis yang memudahkan interpretasi hasil.

Gambar 4 mengilustrasikan secara komprehensif tingkat akurasi yang dicapai oleh model Transformer sebagai algoritma utama, di mana terlihat jelas keunggulan signifikan dibandingkan dengan model LSTM dan DNN dalam mengenali perintah suara berbahasa Indonesia.



Gambar 4. Diagram akurasi model transformer

Secara teknis, pengujian sistem secara *real-time* menghasilkan latensi rata-rata 320ms. Sistem mampu menangani hingga 28 permintaan suara per menit, dengan penggunaan CPU sekitar 42% dan memori sebesar 512MB. Uji ketahanan terhadap *noise* menunjukkan penurunan akurasi sekitar 7% pada kondisi lingkungan bising, menegaskan pentingnya tahapan *preprocessing* seperti *noise reduction* dan ekstraksi fitur MFCC dalam menjaga performa sistem.

Dampak penggunaan teknologi ASR berbasis *machine learning* terhadap pembelajaran juga terlihat dari hasil *pre-test* dan *post-test* terhadap 20 peserta, yang menunjukkan peningkatan skor rata-rata sebesar 18 poin. Selain itu, data interaksi mencatat bahwa sekitar 85% peserta menjadi lebih aktif dalam proses belajar menggunakan sistem berbasis suara. Temuan ini mendukung hasil penelitian sebelumnya yang menyatakan bahwa teknologi suara meningkatkan partisipasi aktif dan keterlibatan siswa dalam pembelajaran digital.

Hasil wawancara dengan peserta dan pengasuh menunjukkan bahwa penggunaan perintah suara mempermudah proses belajar, terutama bagi siswa yang kesulitan mengetik atau menggunakan antarmuka digital konvensional. Mayoritas peserta menyatakan bahwa sistem terasa lebih interaktif, responsif, dan menyenangkan. Dalam sesi FGD bersama pendamping panti, disampaikan bahwa teknologi ini berpotensi menjembatani kesenjangan digital, khususnya di lingkungan pendidikan dengan akses terbatas.

Secara keseluruhan, integrasi ASR berbasis *machine learning* ke dalam pembelajaran digital terbukti efektif, baik dari segi kinerja teknis maupun respon pengguna. Model Transformer dan LSTM menunjukkan keunggulan dalam mengenali perintah suara

secara akurat. Penerapan strategi seperti augmentasi data dan optimasi hyperparameter turut meningkatkan performa sistem. Analisis kualitatif juga mengindikasikan bahwa penggunaan teknologi ini berdampak pada kenyamanan, motivasi, dan semangat belajar siswa.

Temuan ini mendukung pendekatan *Society 5.0*, di mana teknologi seperti ASR berkontribusi pada terciptanya pembelajaran yang inklusif dan adaptif, terutama bagi kelompok masyarakat yang kurang terfasilitasi secara teknologi maupun sosial.

KESIMPULAN

Penelitian ini membuktikan bahwa integrasi sistem pengenalan perintah suara berbasis algoritma *machine learning*, khususnya model Transformer dan LSTM, mampu memberikan kontribusi nyata terhadap peningkatan kualitas pembelajaran digital di lingkungan yang memiliki keterbatasan akses teknologi seperti Panti Asuhan Kasih Sayang di Bengkulu. Model Transformer terbukti menjadi algoritma paling unggul dengan akurasi mencapai 93,4%, diikuti oleh LSTM dengan akurasi 90,2%. Di sisi lain, model DNN sebagai *baseline* hanya mencapai akurasi 84,7%, memperlihatkan perbedaan performa yang signifikan di antara algoritma yang diuji. Sistem ini juga menunjukkan kemampuan teknis yang stabil dengan latensi rata-rata 320ms dan ketahanan terhadap *noise*, meskipun akurasi menurun sekitar 7% pada kondisi lingkungan bising. Temuan penting lainnya adalah dampak positif terhadap hasil belajar peserta, yang tercermin dari peningkatan skor *post-test* sebesar 18 poin serta tingginya keterlibatan siswa selama proses pembelajaran. Selain aspek teknis, hasil kualitatif melalui wawancara dan FGD mengungkap bahwa sistem ini mampu meningkatkan kenyamanan, motivasi, dan keterlibatan siswa, terutama bagi mereka yang kurang terbiasa menggunakan antarmuka berbasis teks.

Berdasarkan hasil tersebut, beberapa rekomendasi dapat disampaikan untuk pengembangan dan implementasi lanjutan. Pertama, sistem pengenalan suara ini memiliki potensi besar untuk diterapkan secara lebih luas, baik di sekolah-sekolah di wilayah Bengkulu maupun daerah-daerah 3T (Terdepan, Terpencil, dan Tertinggal), guna mendukung pembelajaran yang lebih inklusif.

Kedua, pengembangan sistem di masa depan sebaiknya mempertimbangkan integrasi pengenalan terhadap dialek dan aksen lokal agar performanya lebih adaptif dalam konteks regional. Ketiga, untuk meningkatkan ketahanan sistem terhadap gangguan suara, perlu dilakukan pelatihan tambahan dengan data dari lingkungan berisik dan penguatan teknik *denoising*. Keempat, pelatihan penggunaan sistem bagi guru dan pengasuh sangat diperlukan agar proses pendampingan peserta didik menjadi lebih efektif. Terakhir, evaluasi jangka panjang perlu dilakukan guna mengukur dampak teknologi ini terhadap literasi digital dan hasil belajar secara berkelanjutan. Dengan pendekatan ini, pemanfaatan teknologi ASR dapat memberikan kontribusi konkret dalam menciptakan ekosistem pendidikan digital yang inklusif, adaptif, dan sesuai dengan prinsip *Society 5.0*.

DAFTAR PUSTAKA

- Ahlawat, H., Aggarwal, N., & Gupta, D. (2025). *Automatic Speech Recognition: A survey of deep learning techniques and approaches*. *International Journal of Cognitive Computing in Engineering*, 6(January), 201–237. <https://doi.org/10.1016/j.ijcce.2024.12.007>
- Ahmed, S. K. (2024). *How to choose a sampling technique and determine sample size for research: A simplified guide for researchers*. *Oral Oncology Reports*, 12(September), 100662. <https://doi.org/10.1016/j.oor.2024.100662>
- Al-Fraihat, D., Sharrab, Y., Alzyoud, F., Qahmash, A., Tarawneh, M., & Maaita, A. (2024). *Speech Recognition Utilizing Deep Learning: A Systematic Review of the Latest Developments*. *Human-Centric Computing and Information Sciences*, 14(March). <https://doi.org/10.22967/HGIS.2024.14.015>
- B, T. Z. (2023). *Proceedings of the 2022 7th International Conference on Modern Management and Education Technology (MMET 2022)*. In *Proceedings of the 2022 7th International Conference on Modern Management and Education Technology (MMET 2022)*. Atlantis Press SARL. <https://doi.org/10.2991/978-2-494069-51-0>
- Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). *Purposive sampling: complex or simple? Research case examples*. *Journal of Research in Nursing*, 25(8), 652–661. <https://doi.org/10.1177/1744987120927206>
- Helander, M., Moody, T. S., & Joost, M. G. (1988). *Chapter 14 - Systems Design for Automated Speech Recognition* (M. B. T.-H. of H.-C. I. HELANDER (ed.); pp. 301–319). North-Holland. <https://doi.org/https://doi.org/10.1016/B978-0-444-70536-5.50019-1>
- Kheddar, H., Hemis, M., & Himeur, Y. (2024). *Automatic speech recognition using advanced deep learning approaches: A survey*. *Information Fusion*, 109, 102422. <https://doi.org/https://doi.org/10.1016/j.inffus.2024.102422>
- Liang, J., Li, G., Zhang, F., Fan, D., & Luo, H. (2023). *Benefits and challenges of the educational metaverse: Evidence from quantitative and qualitative data*. *Journal of Educational Technology Development and Exchange*, 16(1), 71–91. <https://doi.org/10.18785/jetde.1601.04>
- Natarajan, S., Rahman Al-Haddad, S. A., Ahmad, F. A., Kamil, R., Hassan, M. K., Azrad, S., Macleans, J. F.,

- Abdulhussain, S. H., Mahmmod, B. M., Saparkhojayev, N., & Dautbayeva, A. (2025). *Deep neural networks for speech enhancement and speech recognition: A systematic review*. *Ain Shams Engineering Journal*, 16(7), 103405. <https://doi.org/10.1016/j.asej.2025.103405>
- O'Shaughnessy, D. (2024). *Trends and developments in automatic speech recognition research*. *Computer Speech & Language*, 83, 101538. <https://doi.org/10.1016/j.csl.2023.101538>
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., Hoagwood, K., Angeles, L., & Northwest, K. P. (1968). *"Dentists face added drug regulation"*. *Dental Survey*, 44(12), 73. <https://doi.org/10.1007/s10488-013-0528-y.Purposeful>
- Sarbast, H. (2024). *Voice Recognition Based on Machine Learning Classification Algorithms: A Review*. *Indonesian Journal of Computer Science*, 13(3), 4414–4431. <https://doi.org/10.33022/ijcs.v13i3.4110>
- Silva, B., Hak, F., Guimaraes, T., Manuel, M., & Santos, M. F. (2023). *Rule-based System for Effective Clinical Decision Support*. *Procedia Computer Science*, 220(2019), 880–885. <https://doi.org/10.1016/j.procs.2023.03.119>
- Sun, W. (2023). *The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation*. *Frontiers in Psychology*, 14(August). <https://doi.org/10.3389/fpsyg.2023.1210187>