

Implementasi Whisper Fine-Tuning sebagai Model Closed Caption pada Browser Extension untuk Aksesibilitas Tunarungu

Isaac Yeremia Nugroho^{*1}, Ryan Putranda Kristianto², Yulia Wahyuningsih³

^{1,2,3*} Program Studi Ilmu Informatika, Universitas Katolik Darma Cendika, Surabaya

*Co-author: isaac.nugroho@student.ukdc.ac.id

Informasi Artikel	Abstrak
Tanggal: Dikirim : 15-Ags-2026 Direvisi : 16-Sep-2026 Diterima : 15-Sep-2026 Kata Kunci: Whisper, fine-tuning, closed caption, pengenalan ucapan otomatis, aksesibilitas tunarungu Keyword: <i>Whisper, fine-tuning, closed caption, automatic speech recognition, deaf accessibility.</i> 	Pengguna tunarungu menghadapi hambatan signifikan dalam mengakses konten audio digital seperti video conference, podcast, dan siaran langsung, karena minimnya closed caption otomatis berbahasa Indonesia yang dapat diandalkan. Kondisi ini diperparah oleh jumlah penyandang disabilitas di Indonesia yang mencapai lebih dari 22 juta jiwa, dengan kelompok tunarungu sebagai salah satu bagian terbesar yang membutuhkan solusi aksesibilitas digital yang konkret. Penelitian ini mengimplementasikan model Whisper hasil fine-tuning sebagai model utama untuk closed caption Bahasa Indonesia pada browser extension bernama CaptionAI, dengan model Whisper baseline (zero-shot) dan Fast Conformer sebagai model pembanding untuk mengevaluasi sejauh mana fine-tuning meningkatkan performa model ASR berskala kecil yang dapat di-hosting mandiri. Model Whisper varian small di-fine-tune penuh menggunakan dataset gabungan FLEURS, Common Voice 17, dan Common Voice 19 Bahasa Indonesia (26.400 sampel), dengan konfigurasi learning rate 5e-6, batch size efektif 32, dan max_steps 1.000 langkah menggunakan HuggingFace Seq2SeqTrainer. Checkpoint terbaik dipilih otomatis melalui EarlyStoppingCallback berdasarkan Word Error Rate (WER) validasi terendah, lalu diekspor ke format ONNX untuk deployment. Evaluasi pada 800 sampel test yang belum pernah dilihat model selama pelatihan menunjukkan model Whisper hasil fine-tuning mencapai WER 15,84% dan Character Error Rate (CER) 5,35%, unggul signifikan dibanding Whisper baseline zero-shot (WER 30,83%, CER 9,29%) maupun Fast Conformer (WER 34,72%, CER 15,40%), dengan latensi rata-rata 0,474 detik yang kompetitif untuk live captioning meski belum memenuhi standar real-time ketat di bawah 400 milidetik. Sistem berhasil diimplementasikan sebagai browser extension dengan server backend FastAPI, menerapkan pola inferensi paralel antara Whisper ONNX lokal dan Groq API sebagai jalur cloud. Evaluasi usability melalui System Usability Scale (SUS) terhadap 25 responden menghasilkan skor rata-rata 67,1 dari skala 100 (kategori "ok"), sementara pengujian black-box terhadap 12 skenario fungsional menunjukkan sistem berjalan sesuai

harapan. Hasil ini menegaskan bahwa fine-tuning memberikan peningkatan performa signifikan terhadap kemampuan dasar model ASR pre-trained untuk konteks Bahasa Indonesia, serta layak diterapkan sebagai solusi aksesibilitas real-world bagi pengguna tunarungu.

Abstract

Deaf and hard-of-hearing users face significant barriers accessing digital audio content such as video conferences, podcasts, and live broadcasts, due to the lack of reliable automatic Indonesian closed captioning. This is compounded by Indonesia's disability population exceeding 22 million, with the deaf and hard-of-hearing community representing one of the largest groups requiring concrete digital accessibility solutions. This study implements a fine-tuned Whisper model as the primary model for Indonesian closed captioning in a browser extension called CaptionAI, with a zero-shot Whisper baseline and Fast Conformer as comparison models to evaluate how much fine-tuning improves a small-scale, self-hostable ASR model. The Whisper-small model was fully fine-tuned on a combined Indonesian dataset from FLEURS, Common Voice 17, and Common Voice 19 (26,400 samples), using a learning rate of 5e-6, an effective batch size of 32, and 1,000 max training steps with the HuggingFace Seq2SeqTrainer. The best checkpoint was automatically selected via EarlyStoppingCallback based on the lowest validation Word Error Rate (WER), then exported to ONNX format for deployment. Evaluation on 800 test samples never seen during training shows the fine-tuned Whisper model achieves a WER of 15.84% and Character Error Rate (CER) of 5.35%, significantly outperforming the zero-shot Whisper baseline (WER 30.83%, CER 9.29%) and Fast Conformer (WER 34.72%, CER 15.40%), with an average latency of 0.474 seconds, competitive for live captioning despite not fully meeting the strict real-time standard under 400 milliseconds. The system was successfully implemented as a browser extension with a FastAPI backend, applying a parallel inference pattern between local Whisper ONNX and the Groq API as a cloud pathway. Usability evaluation via the System Usability Scale (SUS) with 25 respondents yielded an average score of 67.1 out of 100 ("ok" category), while black-box testing across 12 functional scenarios showed the system performing as expected. These results confirm that fine-tuning provides significant performance gains over pre-trained ASR models for the Indonesian language context, and that this approach is viable as a real-world accessibility solution for deaf and hard-of-hearing users.

1. PENDAHULUAN

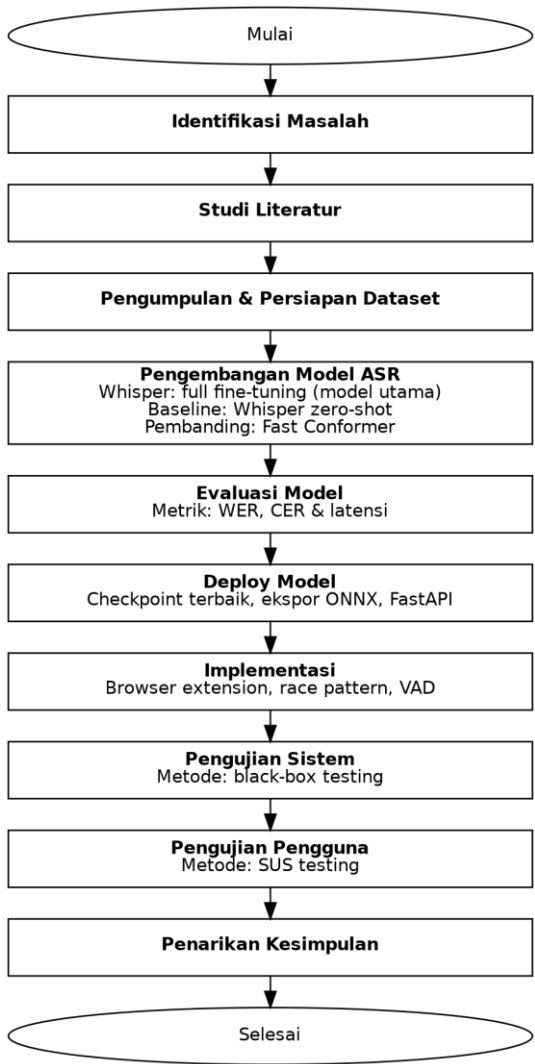
Tunarungu merupakan salah satu kelompok disabilitas sensorik yang menghadapi hambatan signifikan dalam mengakses informasi berbasis audio di era digital. Gangguan pendengaran saat ini memengaruhi lebih dari 1,5 miliar orang di dunia dan diprediksi mencapai 2,5 miliar pada tahun 2050 [1]. Di Indonesia, jumlah penyandang disabilitas mencapai 22,97 juta jiwa atau sekitar 8,5% dari total penduduk, dengan kelompok tunarungu sebagai salah satu bagian terbesar yang membutuhkan solusi aksesibilitas digital yang konkret [2]. Teknologi berbasis kecerdasan buatan seperti large language model untuk penyempurnaan caption otomatis [3] dan sistem penerjemah bahasa isyarat [4] menjadi beberapa pendekatan yang terus dikembangkan untuk menjawab kebutuhan aksesibilitas tersebut. Sebagian besar konten audio di web—video conference, podcast, siaran langsung—tidak dilengkapi teks otomatis, sehingga pengguna tunarungu mengalami hambatan akses nyata.

Model Whisper dari OpenAI terbukti mampu melakukan transkripsi multibahasa secara andal, mencapai Word Error Rate (WER) 2,7% pada LibriSpeech test-clean secara zero-shot [5]. Untuk Bahasa Indonesia, Whisper unggul dibanding XLS-R dan XLSR-53 hasil fine-tuning, dengan WER 19,77% berbanding 56,35% dan 71,91% [6]. Pada bahasa daerah beresource rendah seperti Mandar, fine-tuning Whisper-small menurunkan WER dari 73,7% menjadi 37,4% [7]. Namun tanpa fine-tuning, performa Whisper turun drastis pada bahasa low-resource seperti Bahasa Jawa, dengan WER mencapai 89,40% [8]. Pendekatan deep learning berbasis Recurrent Neural Network juga telah dieksplorasi untuk membangun sistem speech-to-text Bahasa Indonesia [9]. Penelitian-penelitian ini membuktikan efektivitas fine-tuning Whisper, namun masih bersifat evaluasi offline dan belum diintegrasikan ke sistem hampir real-time yang dapat langsung digunakan, apalagi dalam bentuk browser extension untuk aksesibilitas tunarungu.

Berdasarkan kesenjangan tersebut, penelitian ini mengimplementasikan model Whisper hasil fine-tuning sebagai model utama untuk closed caption Bahasa Indonesia pada browser extension. Model Whisper varian small di-fine-tune secara penuh menggunakan dataset gabungan FLEURS, Common Voice 17, dan Common Voice 19 Bahasa Indonesia, kemudian diekspor ke format ONNX agar dapat dijalankan efisien. Fine-tuning dipilih alih-alih hanya mengandalkan layanan cloud karena memberi nilai praktis berupa redundansi saat layanan cloud terganggu serta independensi dari biaya API berkelanjutan. Model Fast Conformer disertakan sebagai pembandingan evaluasi performa. Penelitian ini bertujuan mengevaluasi sejauh mana fine-tuning meningkatkan performa model ASR berskala kecil yang dapat di-hosting mandiri untuk Bahasa Indonesia, diukur melalui metrik WER, Character Error Rate (CER), dan latensi.

2. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif, di mana data dari dataset publik Bahasa Indonesia (FLEURS, Common Voice 17, dan Common Voice 19) digunakan untuk melakukan full fine-tuning terhadap model Whisper varian small sebagai model utama. Sebagai pembandingan, model Whisper tanpa fine-tuning (zero-shot) dan model Fast Conformer [10], varian efisien dari arsitektur Conformer [11], dievaluasi secara paralel menggunakan dataset dan metrik yang sama, tanpa disertakan ke dalam sistem yang di-deploy. Tahapan penelitian terdiri dari sepuluh tahap berurutan, mulai dari identifikasi masalah hingga penarikan kesimpulan, sebagaimana ditunjukkan pada Gambar 1.



Gambar.1 Tahapan Penelitian

Rincian tiap tahapan ditunjukkan pada Tabel 1 berikut ini.

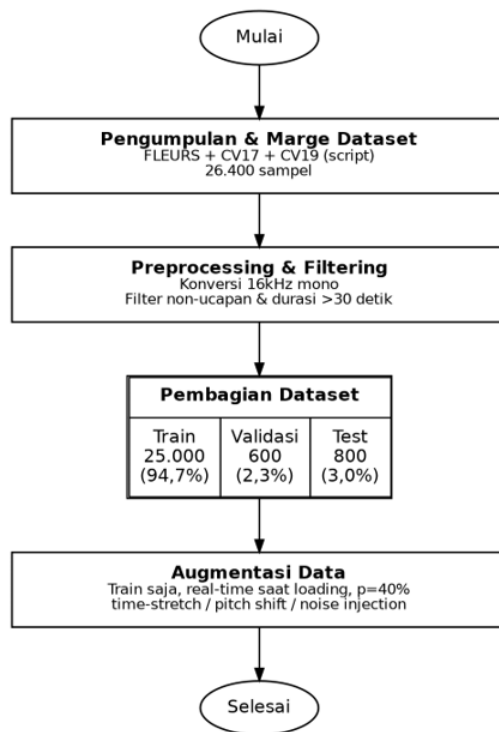
Tabel.1 Aktivitas dalam Metode Waterfall

No	Tahapan	Deskripsi
1	Identifikasi Masalah	Menentukan kebutuhan closed caption Bahasa Indonesia hampir real-time untuk aksesibilitas tunarungu berbasis browser extension.
2	Studi Literatur	Kajian arsitektur Whisper, Fast Conformer, strategi fine-tuning, ONNX Runtime, dan Manifest V3.
3	Persiapan Dataset	Pengumpulan dan pembersihan dataset FLEURS, Common Voice 17 dan 19 Bahasa Indonesia.
4	Pemodelan	Full fine-tuning Whisper-small sebagai model utama; Fast Conformer dilatih paralel sebagai pembanding.
5	Evaluasi Model	Perbandingan WER, CER, dan latensi antar model.
6	Deploy Model	Ekspor checkpoint terbaik ke ONNX, penyiapan server inferensi FastAPI.
7	Implementasi	Integrasi model ke browser extension dengan Groq API sebagai jalur cloud paralel.
8	Pengujian Sistem	Pengujian fungsional menggunakan metode black-box testing.

9	Pengujian Pengguna	Pengukuran usability menggunakan System Usability Scale (SUS).
10	Kesimpulan	Menjawab rumusan masalah penelitian.

A. Dataset

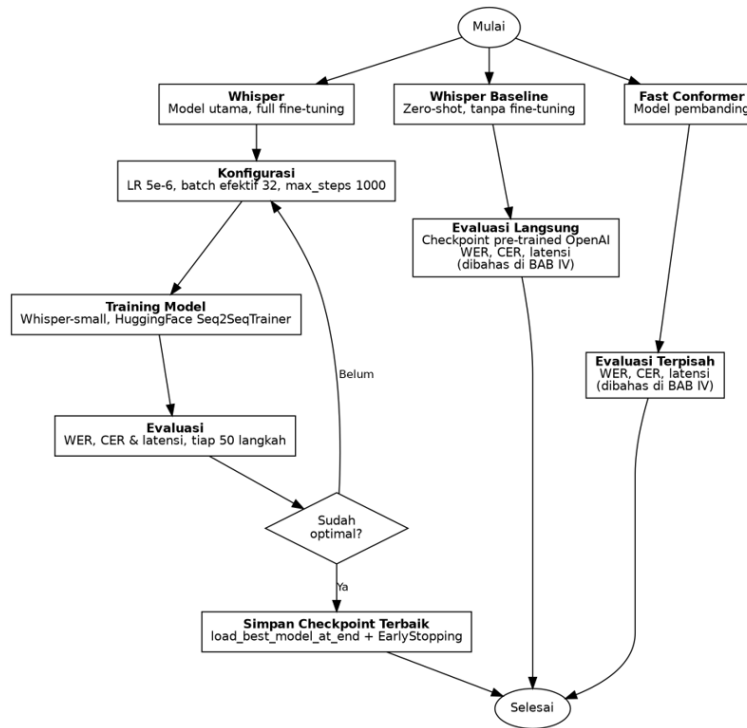
Dataset gabungan berjumlah 26.400 sampel audio tervalidasi (47,7 jam), dikonversi ke 16 kHz mono dan disaring dari sampel non-ucapan ([music], [noise], dll.) serta durasi di atas 30 detik, sebagaimana ditunjukkan pada Gambar 2. Dataset dibagi menjadi 25.000 sampel (94,7%) train, 600 sampel (2,3%) validasi, dan 800 sampel (3,0%) test. Augmentasi (time-stretch, pitch shift, noise injection) diterapkan pada 40% data train secara acak-tunggal.



Gambar.2 Diagram Alur Pengolahan Data

B. Fine-Tuning Model Whisper

Model Whisper varian small menjalani full fine-tuning dengan learning rate 5e-6, batch size efektif 32, dan max_steps 1.000, menggunakan HuggingFace Seq2SeqTrainer pada GPU Kaggle, sebagaimana ditunjukkan pada Gambar 3. Model dievaluasi setiap 50 langkah dengan load_best_model_at_end dan EarlyStoppingCallback (patience 5). Sebagai pembandingan, Whisper zero-shot dievaluasi langsung dari checkpoint pre-trained OpenAI tanpa pelatihan, dan Fast Conformer dilatih terpisah menggunakan pipeline NeMo. Checkpoint terbaik (checkpoint-900, WER validasi 17,78%) diekspor ke format ONNX (encoder, decoder, decoder-with-past, decoder-merged) menggunakan HuggingFace Optimum tanpa kuantisasi, kemudian diunggah ke HuggingFace Hub (Isaacyn/whisper-small-id).



Gambar.3 Alur Pengembangan Model ASR

C. Evaluasi Model

Tiga metrik digunakan untuk membandingkan Whisper hasil fine-tuning, Whisper baseline zero-shot, dan Fast Conformer:

$$\text{WER} = (S + D + I) / N \times 100\% \quad (1)$$

$$\text{CER} = (Sc + Dc + Ic) / Nc \times 100\% \quad (2)$$

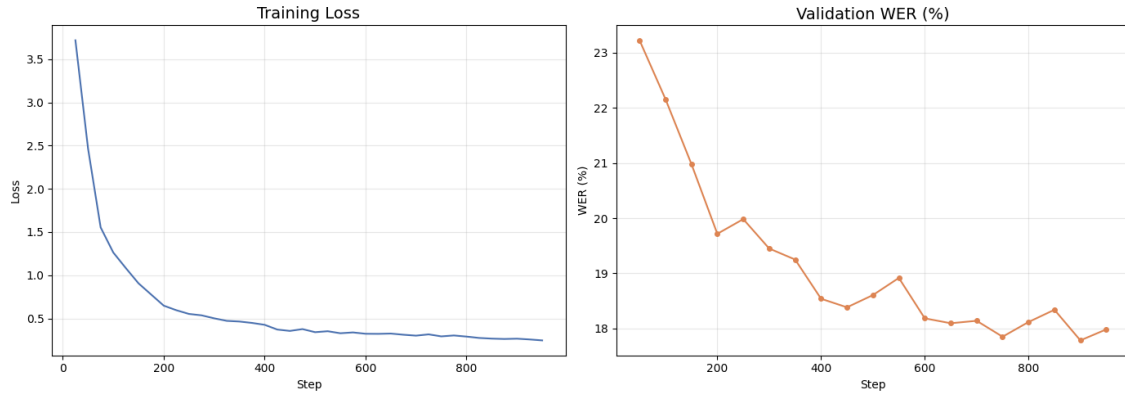
$$L = t_{\text{output}} - t_{\text{input}} \quad (3)$$

di mana S, D, I berturut-turut adalah jumlah substitusi, penghapusan, dan penyisipan kata (huruf untuk CER), N adalah total kata/karakter referensi, dan L adalah latensi end-to-end. Evaluasi dilakukan pada 800 sampel test yang tidak pernah dilihat model selama pelatihan.

3. HASIL DAN PEMBAHASAN

A. Hasil Fine-Tuning Whisper

Model Whisper varian small menjalani proses full fine-tuning menggunakan HuggingFace Seq2SeqTrainer dengan konfigurasi sebagaimana dijelaskan pada bagian Metode Penelitian. Perkembangan proses pelatihan dipantau melalui kurva training loss dan validation WER sepanjang 1.000 langkah, sebagaimana ditunjukkan pada Gambar 4.



Gambar.4 Alur Pengembangan Model ASR

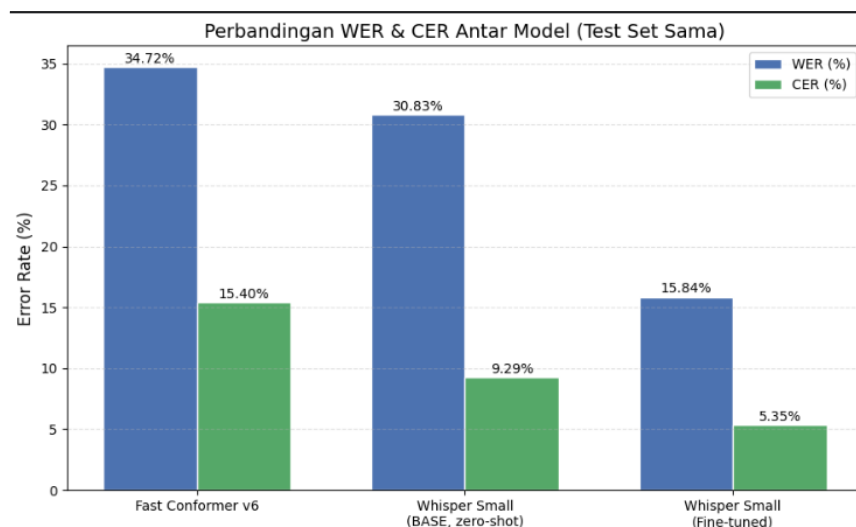
Kurva training loss menunjukkan penurunan tajam pada 200 langkah pertama, kemudian melandai secara bertahap hingga mendekati konvergen pada langkah ke-1.000, mengindikasikan proses pembelajaran yang stabil tanpa tanda overfitting yang ekstrem. Kurva validation WER menunjukkan tren penurunan dari 23,26% pada langkah ke-50 menjadi nilai terendah 17,78% pada langkah ke-900. Checkpoint-900 kemudian dipilih secara otomatis oleh EarlyStoppingCallback sebagai model terbaik, untuk selanjutnya dievaluasi pada 800 sampel data test yang belum pernah dilihat model sebelumnya.

B. Perbandingan Performa Model

Ketiga model — Whisper hasil fine-tuning, Whisper baseline (zero-shot), dan Fast Conformer v6 — dievaluasi pada 800 sampel data test yang identik dan tidak pernah dilihat model selama pelatihan. Hasil perbandingan ditunjukkan pada Tabel 2 dan Gambar 5.

Tabel.2 Biaya Komponen

Metrik	Whisper (Fine-Tuned)	Whisper (Baseline)	Fast Conformer v6
WER	15,84%	30,83%	34,72%
CER	5,35%	9,29%	15,40%



Gambar.5 Alur Pengembangan Model ASR

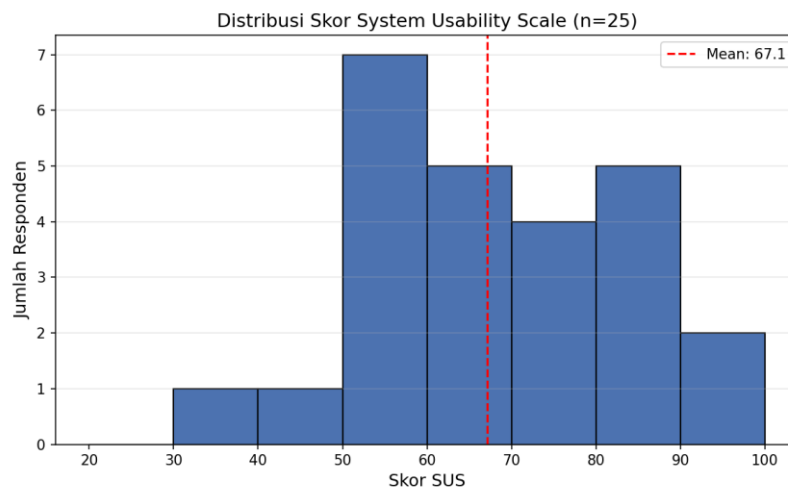
Whisper hasil fine-tuning unggul konsisten pada kedua metrik. Dibanding Fast Conformer v6, Whisper fine-tuned unggul 18,88 poin persentase pada WER (2,19 kali lebih baik) dan 10,05 poin persentase pada CER (2,88 kali lebih baik). Dibanding Whisper baseline, unggul 14,99 poin persentase pada WER (1,95 kali lebih baik) dan 3,94 poin persentase pada CER (1,74 kali lebih baik), mengonfirmasi efek fine-tuning yang signifikan terhadap performa dasar model pre-trained. Berdasarkan kategori WER [12], nilai 15,84% berada pada rentang "Good and Usable" (6–12%) mendekati batas atas. Temuan ini sejalan dengan studi banding arsitektur ASR lain yang turut menunjukkan keunggulan Whisper dibanding model berbasis Conformer pada berbagai skenario evaluasi [13], meski pendekatan continued pretraining pada model berbasis wav2vec2 turut menunjukkan potensi peningkatan performa serupa pada domain spesifik [14]. Selisih performa Whisper terhadap Fast Conformer sejalan dengan karakteristik arsitektur: Whisper diuntungkan skala data pretraining multibahasa yang masif (680.000 jam audio), sementara Fast Conformer dilatih dari bobot dasar Bahasa Inggris tanpa keunggulan pretraining multibahasa yang sebanding.

C. Evaluasi Latensi

Model Whisper hasil fine-tuning mencatatkan latensi rata-rata 0,474 detik dan median 0,438 detik pada 800 sampel test. Berdasarkan standar ITU untuk sistem real-time ketat (<400 ms) [15], nilai ini belum sepenuhnya memenuhi ambang tersebut, namun berada jauh di bawah ambang batas praktis 1,5 detik yang masih dapat diterima untuk live captioning [16]. Nilai ini juga kompetitif dibanding penelitian sejenis, di mana sistem speech-to-text berbasis edge computing melaporkan latensi end-to-end di bawah 1,1 detik [17], dan sistem audio dubbing multibahasa berbasis Chrome extension melaporkan latensi 1,7 hingga 1,9 detik per segmen [18].

D. Implementasi dan Pengujian Sistem

Sistem CaptionAI berhasil diimplementasikan sebagai browser extension berbasis WebExtensions API [19] dengan server backend FastAPI di HuggingFace Spaces, menerapkan pola inferensi paralel antara Whisper ONNX lokal dan Groq API sebagai jalur cloud. Pengujian black-box terhadap 12 skenario fungsional menunjukkan mayoritas skenario berjalan valid sesuai harapan. Evaluasi usability melalui System Usability Scale (SUS) terhadap 25 responden menghasilkan skor rata-rata 67,1 dari skala 100, dengan distribusi skor individual ditunjukkan pada Gambar 6.



Gambar.6 Distribusi Skor System Usability Scale

Berdasarkan skala adjective rating [20], skor rata-rata 67,1 berada pada kategori "ok" — mengindikasikan sistem tergolong usable namun masih terdapat ruang perbaikan, khususnya pada aspek kecepatan transkripsi yang paling banyak dikeluhkan responden.

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan model Whisper hasil fine-tuning sebagai model utama untuk closed caption Bahasa Indonesia pada browser extension CaptionAI. Model Whisper hasil fine-tuning mencapai WER 15,84% dan CER 5,35%, unggul signifikan dibanding Whisper baseline zero-shot (WER 30,83%, CER 9,29%) maupun Fast Conformer v6 (WER 34,72%, CER 15,40%), mengonfirmasi bahwa proses fine-tuning memberikan peningkatan performa yang substansial terhadap kemampuan dasar model pre-trained untuk konteks Bahasa Indonesia. Model tersebut mencatatkan latensi rata-rata 0,474 detik, kompetitif untuk skenario live captioning meski belum memenuhi standar real-time ketat. Evaluasi usability menghasilkan skor SUS 67,1 (kategori "ok"), dan pengujian black-box terhadap 12 skenario fungsional menunjukkan sistem berjalan sesuai harapan.

Penelitian selanjutnya dapat mengeksplorasi teknik akselerasi inferensi lebih lanjut untuk menekan latensi model secara lebih agresif [21], serta penambahan modul koreksi pasca-transkripsi berbasis large language model atau pendekatan direct speech translation untuk meningkatkan kualitas caption pada kalimat pendek dengan artikulasi ambigu [22].

DAFTAR PUSTAKA

- [1] WHO, "WORLD REPORT ON HEARING," Mar. 2021. Accessed: Apr. 10, 2026. [Online]. Available: <https://www.who.int/publications/i/item/9789240020481>
- [2] Kemenko PMK, "Pemerintah Penuhi Hak Penyandang Disabilitas di Indonesia | Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan," <https://www.kemenkopmk.go.id/pemerintah-penuhi-hak-penyandang-disabilitas-di-indonesia>. Accessed: Apr. 10, 2026. [Online]. Available: <https://www.kemenkopmk.go.id/pemerintah-penuhi-hak-penyandang-disabilitas-di-indonesia>
- [3] N. Fathallah, M. Bhole, and S. Staab, "Empowering the Deaf and Hard of Hearing Community: Improving Video Captions with Large Language Models," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Jul. 2025, pp. 345–353. doi: 10.1145/3696593.3696617.
- [4] G. Kalaiselvi, N. Badri, and C. Karnan, "SignSpeak – Sign language translation system for hearing impaired," *International Journal of Science and Research Archive*, vol. 15, no. 2, pp. 921–930, May 2025, doi: 10.30574/ijrsra.2025.15.2.1523.
- [5] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision," Dec. 2023, [Online]. Available: <http://arxiv.org/abs/2212.04356>
- [6] E. William and A. Zahra, "SPEECH RECOGNITION DENGAN WHISPER DALAM BAHASA INDONESIA," *Action Research Literate*, vol. 9, no. 2, 2025.
- [7] J. Jafar, M. A. W. Tb, F. Aziz, R. Iriany, and N. Nasir, "Fine-Tuning Whisper Model for Mandar Speech Recognition: Approach and Performance Evaluation," *Journal of Applied Engineering and Technological Science (JAETS)*, vol. 7, no. 1, pp. 261–273, Dec. 2025, doi: 10.37385/jaets.v7i1.7170.
- [8] R. S. A. Pratama and A. Amrullah, "ANALYSIS OF WHISPER AUTOMATIC SPEECH RECOGNITION PERFORMANCE ON LOW RESOURCE LANGUAGE," *Jurnal Pilar Nusa Mandiri*, vol. 20, no. 1, pp. 1–8, Mar. 2024, doi: 10.33480/pilar.v20i1.4633.
- [9] N. Lubis, Mhd. Z. Siambaton, and R. Aulia, "Implementasi Algoritma Deep Learning pada Aplikasi Speech to Text Online dengan Metode Recurrent Neural Network (RNN)," *sudo Jurnal Teknik Informatika*, vol. 3, no. 3, pp. 113–126, Sep. 2024, doi: 10.56211/sudo.v3i3.583.

- [10] D. Rekesh *et al.*, “Fast Conformer with Linearly Scalable Attention for Efficient Speech Recognition,” Sep. 2023, [Online]. Available: <http://arxiv.org/abs/2305.05084>
- [11] A. Gulati *et al.*, “Conformer: Convolution-augmented Transformer for Speech Recognition,” May 2020, [Online]. Available: <http://arxiv.org/abs/2005.08100>
- [12] N. Giovanni and W. Nurhayat, “The Effectiveness Of Automatic Speech Recognition In The Capcut Application For Developing Inclusive Learning Media,” JKTP: Jurnal Kajian Teknologi Pendidikan, vol. 8, no. 2, pp. 157–169, May 2025, doi: 10.17977/um038v8i22025p157.
- [13] A. Barcovich, R. Jain, and P. Corcoran, “A comparative analysis between Conformer-Transducer, Whisper, and wav2vec2 for improving the child speech recognition,” Nov. 2023, [Online]. Available: <http://arxiv.org/abs/2311.04936>
- [14] A. A. Attia, D. Demszky, T. Ogunremi, J. Liu, and C. Espy-Wilson, “Continued Pretraining for Domain Adaptation of Wav2vec2.0 in Automatic Speech Recognition for Elementary Math Classroom Settings,” May 2024, [Online]. Available: <http://arxiv.org/abs/2405.13018>
- [15] I. Degtyarenko, O. Radyvonenko, N. Tkach, V. Seliukhov, K. Seliuk, and O. Ivanov, “Real-Time Multilingual Video Subtitle Spotting on Resource-Constrained Devices,” IEEE Access, vol. 13, pp. 211701–211714, 2025, doi: 10.1109/ACCESS.2025.3643512.
- [16] P. Romero-Fresco and Y. Van Gauwbergen, “Fit for What Purpose? NER Certification of Automatic Captions in English and Spanish,” Applied Sciences (Switzerland), vol. 15, no. 3, Feb. 2025, doi: 10.3390/app15031387.
- [17] S. Di Leo, L. De Cicco, and S. Mascolo, “Real-Time Speech-to-Text on Edge: A Prototype System for Ultra-Low Latency Communication with AI-Powered NLP,” Information (Switzerland), vol. 16, no. 8, Aug. 2025, doi: 10.3390/info16080685.
- [18] R. Kannoja, A. K. Singh, I. Sharma, and S. Gupta, “Gen AI driven multilingual audio dubbing and synthesis system for cross-language video platforms,” Results in Engineering, vol. 27, Sep. 2025, doi: 10.1016/j.rineng.2025.106241.
- [19] Mozilla Foundation, “Browser extensions,” <https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions>. Accessed: May 01, 2026. [Online]. Available: <https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions>
- [20] R. Anrahvi, N. Pratama, and S. Stevani, “Penerapan Metode System Usability Scale (SUS) dalam Mengukur Kepuasan Mahasiswa terhadap Website Direktori Akademik,” Indonesian Journal of Business Economics and Management, vol. 3, no. 2, pp. 74–80, Jun. 2024, doi: 10.57152/ijbem.v3i2.2020.
- [21] I. J. Ratul, Y. Zhou, and K. Yang, “Accelerating Deep Learning Inference: A Comparative Analysis of Modern Acceleration Frameworks,” Electronics (Switzerland), vol. 14, no. 15, Aug. 2025, doi: 10.3390/electronics14152977.
- [22] S. Papi, M. Gaido, A. Karakanta, M. Cettolo, M. Negri, and M. Turchi, “Direct Speech Translation for Automatic Subtitling,” Jul. 2023, [Online]. Available: <http://arxiv.org/abs/2209.13192>