

**IMPLEMENTASI METODE *FINE-TUNING*
AUTOMATIC SPEECH RECOGNITION MODEL WHISPER
UNTUK TRANSKRIPSI BACAAN AL-QUR'AN**



SKRIPSI

diajukan untuk memenuhi sebagai syarat untuk memperoleh gelar Sarjana Komputer pada
Program Studi Rekayasa Perangkat Lunak

Oleh
Devalino Choiriliusman Syambara
2102325

**PROGRAM STUDI REKAYASA PERANGKAT LUNAK
KAMPUS UPI DI CIBIRU
UNIVERSITAS PENDIDIKAN INDONESIA
2025**

IMPLEMENTASI METODE FINE-TUNING
AUTOMATIC SPEECH RECOGNITION MODEL WHISPER
UNTUK TRANSKRIPSI BACAAN AL-QUR'AN

Oleh

Devalino Choiriliusman Syambara

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer pada Kampus UPI di Cibiru

© Devalino Choiriliusman Syambara

Universitas Pendidikan Indonesia

Agustus 2025

Hak Cipta dilindungi undang-undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
dengan dicetak ulang, difoto kopi, atau cara lainnya tanpa ijin dari penulis.

HALAMAN PENGESAHAN

Devalino Choiriliusman Syambara

IMPLEMENTASI METODE FINE-TUNING AUTOMATIC SPEECH RECOGNITION MODEL WHISPER UNTUK TRANSKRIPSI BACAAN AL-QUR'AN

Disetujui dan disahkan oleh pembimbing:

Pembimbing I



Indira Syawanodya, S.Kom., M.Kom.

NIP 920190219920423201

Pembimbing II



Dian Anggraini, S.ST., M.T

NIP 920190219930526201

Mengetahui,

Ketua Program Studi Rekayasa Perangkat Lunak



Mochamad Iqbal Ardimansyah, S.T., M.Kom.

NIP 920190219910328101

PERNYATAAN BEBAS PLAGIARISME

Saya yang bertanda tangan di bawah ini:

Nama : Devalino Choiriliusman Syambara

NIM : 2102325

Program Studi : Rekayasa Perangkat Lunak

Judul Karya : Implementasi Metode *Fine-tuning Automatic Speech Recognition* Model Whisper untuk Transkripsi Bacaan Al-Qur'an

Dengan ini menyatakan bahwa karya tulis ini merupakan hasil kerja saya sendiri. Saya menjamin bahwa seluruh isi karya ini, baik sebagian maupun keseluruhan, bukan merupakan plagiarisme dari karya orang lain, kecuali pada bagian yang telah dinyatakan dan disebutkan sumbernya dengan jelas.

Jika di kemudian hari ditemukan pelanggaran terhadap etika akademik atau unsur plagiarisme, saya bersedia menerima sanksi sesuai peraturan yang berlaku di Universitas Pendidikan Indonesia.

Bandung, 22 Agustus 2025



Devalino Choiriliusman Syambara

NIM: 2102325

UCAPAN TERIMA KASIH

Alhamdulillah, segala puji syukur penulis panjatkan ke hadirat Allah SWT atas rahmat, hidayah, dan pertolongan-Nya sehingga penulis dapat melalui seluruh rangkaian proses percobaan, pembelajaran, serta perjuangan hingga akhirnya menyelesaikan penelitian ini. Penulis menyadari sepenuhnya bahwa keberhasilan ini tidak terlepas dari dukungan, doa, bimbingan, dan kontribusi banyak pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, dengan penuh rasa hormat dan kerendahan hati, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada pihak-pihak berikut.

1. Keluarga tercinta, Ayah, Ibu, dan Kakak-kakak yang senantiasa memberikan doa, dukungan secara materiel maupun imateriel, dan kasih sayang yang tak terhingga, serta kesabaran dan pengorbanan yang menjadi cahaya dan kekuatan terbesar dalam setiap langkah hidup penulis. Terima kasih untuk segalanya.
2. Kepada Ibu Indira Syawanodya, S.Kom., M.Kom., selaku dosen pembimbing pertama, atas bimbingan, arahan, ilmu, serta motivasi yang telah diberikan dengan penuh kesabaran dan ketulusan dalam mendampingi penulis selama proses penelitian hingga penyusunan skripsi ini.
3. Kepada Ibu Dian Anggraini, S.ST., M.T., selaku dosen pembimbing kedua, yang telah banyak memberikan masukan berharga, kritik yang membangun, serta arahan yang sangat membantu penulis dalam menyelesaikan penelitian ini.
4. Bapak Mochamad Iqbal Ardimansyah, S.T., M.Kom., selaku Ketua Program Studi Rekayasa Perangkat Lunak, atas segala dukungan, fasilitas, serta kesempatan yang diberikan kepada penulis dalam menempuh pendidikan di program studi ini.
5. Bapak Raditya Muhammad, S.T., M.T., selaku dosen pembimbing akademik yang telah memberikan bimbingan, nasihat, serta arahan selama masa perkuliahan sehingga penulis dapat menyelesaikan studi dan penelitian ini dengan baik.
6. Seluruh Dosen dan Staf Program Studi Rekayasa Perangkat Lunak UPI Kampus Cibiru, yang telah mendidik, membimbing, serta membagikan ilmu pengetahuan dan pengalaman berharga selama masa perkuliahan. Semangat dan keteladanan yang diberikan menjadi inspirasi besar bagi penulis.
7. Kepada teman-teman seperjuangan yang berada di titik kumpul Gusgas, atas kebersamaan, semangat, serta dukungan moral yang selalu menguatkan penulis, baik dalam proses perkuliahan maupun penyusunan skripsi ini. Terima kasih atas setiap diskusi yang membangun, ide-ide kreatif, kerja sama tim yang solid, serta canda tawa yang meringankan beban selama perjalanan akademik. Kehadiran Kalian tidak hanya menjadi rekan belajar, tetapi juga keluarga kedua yang selalu memberi motivasi dan inspirasi untuk terus maju hingga penelitian ini dapat terselesaikan dengan baik.

8. Kepada teman-teman Program Studi Rekayasa Perangkat Lunak UPI Kampus Cibiru Angkatan 2021, terutama yang selalu hadir memberi semangat, berbagi tawa dan air mata, serta menjadi teman diskusi dalam proses penyusunan skripsi ini.
9. Kepada diri sendiri, atas kesabaran, keteguhan hati, serta semangat untuk terus berusaha meskipun dihadapkan dengan berbagai kesulitan dan keterbatasan.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Namun, besar harapan penulis semoga karya ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan dan menjadi pijakan bagi penelitian selanjutnya.

Terima kasih untuk semua pihak yang terlibat dalam kesuksesan penelitian ini, semoga jerih payah serta keringat yang telah dikorbankan, dibalaskan yang terbaik oleh Allah SWT. Aamiin.

Bandung, 22 Agustus 2025

Devalino Choiriliusman Syambara

IMPLEMENTASI METODE *FINE-TUNING*
***AUTOMATIC SPEECH RECOGNITION* MODEL WHISPER**
UNTUK TRANSKRIPSI BACAAN AL-QUR'AN

DEVALINO CHOIRILIUSMAN SYAMBARA
NIM 21012325

ABSTRAK

Automatic Speech Recognition (ASR) atau pengenalan ucapan otomatis merupakan teknologi penting dalam pemrosesan bahasa alami yang berfungsi mengubah sinyal suara menjadi teks, serta memiliki potensi luas dalam berbagai aplikasi, termasuk dalam konteks keislaman untuk membantu proses pembelajaran, pengkajian, dan verifikasi bacaan Al-Qur'an. Bacaan Al-Qur'an menuntut ketelitian tinggi karena setiap huruf, kata, dan ayat memiliki aturan tajwid, makhraj huruf, serta panjang-pendek vokal yang memengaruhi makna secara signifikan. Kompleksitas fonetik bahasa Arab dan kekhasan aturan tajwid menjadikan sistem ASR konvensional yang dilatih pada bahasa percakapan sehari-hari kurang optimal ketika diterapkan pada bacaan Al-Qur'an. Oleh karena itu, penelitian ini dilakukan dengan tujuan mengevaluasi efektivitas metode fine-tuning pada model Whisper-small dalam menghasilkan transkripsi bacaan Al-Qur'an, sekaligus membandingkan performa model hasil fine-tuning dengan model dasar (base model) untuk mengetahui sejauh mana proses adaptasi meningkatkan akurasi sistem. Penelitian ini menggunakan pendekatan Desain dalam Research Methodology (DRM) yang meliputi tahap pengumpulan data berupa bacaan Surah 95–114, pra-pemrosesan data melalui normalisasi teks dan konversi audio, pelatihan model dengan fine-tuning, serta evaluasi kinerja menggunakan metrik Word Error Rate (WER) dan Character Error Rate (CER). Hasil penelitian menunjukkan bahwa model hasil fine-tuning secara signifikan lebih unggul dengan nilai rata-rata WER sebesar 0,1236 dan CER sebesar 0,2071, dibandingkan base model yang mencatatkan WER sebesar 1,0939 dan CER sebesar 0,5244. Perbedaan tersebut membuktikan bahwa fine-tuning mampu meningkatkan akurasi transkripsi baik pada level kata maupun karakter, sehingga dapat menjadi solusi potensial dalam pengembangan sistem ASR untuk bacaan Al-Qur'an.

Kata Kunci: *Automatic Speech Recognition, Whisper, Fine-tuning, Word Error Rate, Character Error Rate, Al-Qur'an*

**IMPLEMENTATION OF FINE-TUNING METHOD ON
AUTOMATIC SPEECH RECOGNITION WHISPER MODEL FOR
QUR'ANIC RECITATION TRANSCRIPTION**

DEVALINO CHOIRILIUSMAN SYAMBARA

NIM 21012325

ABSTRACT

Automatic Speech Recognition (ASR) is an essential technology in natural language processing that converts speech signals into textual representations and has been widely applied in various domains, including Islamic contexts such as assisting in learning, verifying, and evaluating the recitation of the Qur'an. Qur'anic recitation requires a high level of precision since every letter, word, and verse must adhere to tajwid rules, articulation points (makhraj), and vowel length, all of which significantly influence meaning. The complexity of Arabic phonetics and the strict rules of tajwid make conventional ASR systems, which are generally trained on everyday speech or Modern Standard Arabic, less effective when applied to Qur'anic recitation. Therefore, this study aims to evaluate the effectiveness of fine-tuning the Whisper-small model for Qur'anic recitation transcription and to compare the performance of the fine-tuned model with its base version to determine the extent of improvement through domain adaptation. This research employs the Design Research Methodology (DRM), covering data collection from short Surahs (95–114), data preprocessing through text normalization and audio conversion, model training via fine-tuning, and performance evaluation using Word Error Rate (WER) and Character Error Rate (CER) metrics. The results indicate that the fine-tuned model significantly outperforms the base model, achieving an average WER of 0.1236 and CER of 0.2071, compared to the base model's WER of 1.0939 and CER of 0.5244. These findings demonstrate that fine-tuning not only enhances word-level recognition but also improves character-level precision, making it a potential solution for developing ASR systems tailored for Qur'anic recitation.

Keywords: *Automatic Speech Recognition, Whisper, Fine-tuning, Word Error Rate, Character Error Rate, Qur'an*

DAFTAR ISI

HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
PERNYATAAN BEBAS PLAGIARISME	Error! Bookmark not defined.
UCAPAN TERIMA KASIH	v
ABSTRAK	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan.....	4
1.4. Manfaat	4
1.5. Batasan	5
1.6. Struktur Organisasi Skripsi	5
BAB II KAJIAN PUSTAKA	7
2.1. Penelitian Terdahulu.....	7
2.2. Al-Qur'an.....	11
2.3. Natural Language Processing.....	11
2.2.1. <i>Automatic Speech Recognition (ASR)</i>	12
2.2.2. Model Whisper.....	13
2.4. <i>Fine-tuning Model</i>	14
2.5. Metrik Evaluasi	15
BAB III METODOLOGI PENELITIAN.....	16
3.1. Desain Penelitian.....	16
3.2. Klarifikasi Penelitian.....	16
3.3. Studi Deskriptif I.....	17
3.4. Studi Preskriptif	17
3.4.1. Pengembangan Model Whisper-small	17
3.4.2. <i>Dataset</i>	19

3.4.3.	Pra-Pemrosesan	21
3.4.4.	Proses <i>Fine-tuning</i>	22
3.4.5.	Transkrip Bacaan Al-Qur'an	23
3.5.	Studi Deskriptif II	24
3.5.1.	<i>Word Error Rate</i> (WER)	24
3.5.2.	<i>Character Error Rate</i> (CER).....	25
3.6.	Instrumen Penelitian.....	26
3.7.	Alat dan Bahan Penelitian.....	26
3.7.1.	Library Python	27
BAB IV HASIL DAN PEMBAHASAN.....		28
4.1.	Pengembangan Model.....	28
4.1.1.	Persiapan <i>Dataset</i>	28
4.1.2.	Pra-pemrosesan Data.....	30
4.1.3.	Pelatihan Model Whisper	32
4.1.4.	Impelementasi Model.....	35
4.2.	Evaluasi Kinerja Model.....	37
4.2.1.	Uji Evaluasi dengan WER dan CER.....	37
4.2.2.	Perbandingan Model <i>Fine-tuned</i> dan Model <i>Base</i>	39
BAB V SIMPULAN DAN SARAN		41
5.1.	Simpulan	41
5.2.	Saran.....	42
DAFTAR PUSTAKA		44
LAMPIRAN.....		49

DAFTAR TABEL

Tabel 2.1. Pratinjau Penelitian Terkait	9
Tabel 3.1 Desain Penelitian.....	16
Tabel 3.2 Distribusi Dataset	20
Tabel 3.3 Daftar Library Python	27
Tabel 4.1 Contoh Sampel File transcripts.tsv.....	29
Tabel 4.2 Tabel Pembagian Dataset	31
Tabel 4.3 Kolom Audio Setelah Pra-Pemrosesan	31
Tabel 4.4. Tabel Parameter Pelatihan	33
Tabel 4.5 Contoh Hasil Transkripsi Model	36
Tabel 4.6 Hasil Pengujian Model Menggunakan Mean Nilai WER dan CER.....	38

DAFTAR GAMBAR

Lampiran 1 Hasil Log Pelatihan.....	49
Lampiran 2 Hasil Uji Perbandingan Transkripsi Model Base dan Fine-tuned....	49

DAFTAR PUSTAKA

- Adnan, R., & Hassani, H. (2025, Agustus 13). *Which one Performs Better? Wav2Vec or Whisper? Applying both in Badini Kurdish Speech to Text (BKSTT)*. arXiv.
- Ahmad, M. (2020). Literary Miracle of The Quran. *Ar-Raniry: International Journal of Islamic Studies*, 3, 205–220.
- Aji, M. H., Hilmi, M. Z., & Rahman, M. T. (2021). The Living Qur'an as a Research Object and Research Methodology in the Qur'anic Studies. *Jurnal Iman Dan Spiritualitas*, 1, 78–84.
- Andreyev, A. (2025, Maret 12). *Quantization for OpenAI's Whisper Models: A Comparative Analysis*. arXiv.
- Cui, M., Geng, M., Deng, J., Deng, C., Kang, J., Hu, S., ... Liu, X. (2025, Agustus 14). *Exploring Cross-Utterance Speech Contexts for Conformer-Transducer Speech Recognition Systems*. arXiv.
- Elitawati, E. (2022). METODE TILAWATI QUR'AN SEBAGAI UPAYA MENINGKATKAN SENI BACA QUR'AN. *JURNAL PUSAKA*, 12, 26–33.
- Ferraz, T. P., Zanon Boito, M., Brun, C., & Nikoulina, V. (2024). Multilingual Distilwhisper: Efficient Distillation of Multi-Task Speech Models Via Language-Specific Experts. *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 10716–10720.

- Gastaldi, J. L., Terilla, J., Malagutti, L., DuSell, B., Vieira, T., & Cotterell, R. (2025, April 3). *The Foundations of Tokenization: Statistical and Computational Concerns*. arXiv.
- Ghaffarzadegan, S., Bořil, H., & Hansen, J. H. L. (2016). Generative Modeling of Pseudo-Whisper for Robust Whispered Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24, 1705–1720.
- Hadwan, M., Alsayadi, H., & AL-Hagree, S. (2022). An End-to-End Transformer-Based Automatic Speech Recognition for Qur'an Reciters. *Computers, Materials & Continua*, 74, 3471–3487.
- Haidar, M. A., Xing, C., & Rezagholizadeh, M. (2021, Maret 17). *Transformer-based ASR Incorporating Time-reduction Layer and Fine-tuning with Self-Knowledge Distillation*. arXiv.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., Jaitly, N., ... Kingsbury, B. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine*, 29, 82–97.
- Iosifova, O., Iosifov, I., Rolik, O., & Sokolov, V. (2020, Juni 16). *Techniques Comparison for Natural Language Processing*. Zenodo.
- J, M. R., VM, K., Warriar, H., & Gupta, Y. (2024, Maret 23). *Fine Tuning LLM for Enterprise: Practical Guidelines and Recommendations*. arXiv.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349, 255–260.

- K, T. D., James, J., Gopinath, D. P., & K, M. A. (2024, Oktober 18). *Advocating Character Error Rate for Multilingual ASR Evaluation*. arXiv.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.
- Martinez, A. R. (2012). Part-of-speech tagging. *WIREs Computational Statistics*, 4, 107–113.
- Nadia, N., Zega, S. A., Salim, T. H., Prameswari, N. F., & Yunan, P. D. (2023). MENINGKATKAN KEMAMPUAN BACAAN AL-QUR'AN DENGAN METODE TAHSIN DAN TAJWID. *Prosiding Seminar Nasional LPPM UMJ*.
- Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association*, 18, 544–551.
- Nakache, D., Metais, E., & Timsit, J. F. (2005). Evaluation and NLP. Dalam K. V. Andersen, J. Debenham, & R. Wagner (Ed.), *Database and Expert Systems Applications* (hlm. 626–632). Berlin, Heidelberg: Springer.
- Park, C., Chen, M., & Hain, T. (2024, April 26). *Automatic Speech Recognition System-Independent Word Error Rate Estimation*. arXiv.
- Pillai, L. G., Manohar, K., Raju, B. K., & Sherly, E. (2024, November 7). *Multistage Fine-tuning Strategies for Automatic Speech Recognition in Low-resource Languages*. arXiv.
- Pratama, R. S. A., & Amrullah, A. (2024). ANALYSIS OF WHISPER AUTOMATIC SPEECH RECOGNITION PERFORMANCE ON LOW RESOURCE LANGUAGE. *Jurnal Pilar Nusa Mandiri*, 20, 1–8.

- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022, Desember 6). *Robust Speech Recognition via Large-Scale Weak Supervision*. arXiv.
- Rahali, A., & Akhloufi, M. A. (2023). End-to-End Transformer-Based Models in Textual-Based NLP. *AI*, 4, 54–110.
- Rahman, A., Kabir, Md. M., Mridha, M. F., Alatiyyah, M., Alhasson, H. F., & Alharbi, S. S. (2024). Arabic Speech Recognition: Advancement and Challenges. *IEEE Access*, 12, 39689–39716.
- Shah, P., Chadha, H. S., Gupta, A., Dhuriya, A., Chhimwal, N., Gaur, R., & Raghavan, V. (2022, Juni 15). *Is Word Error Rate a good evaluation metric for Speech Recognition in Indic Languages?* arXiv.
- Tc, S. F. (2024). Holy Quranic Manuscripts: Examining Historical Variants and Transmission Methods. *Al-Afkar, Journal For Islamic Studies*, 7, 1163–1177.
- Timmel, V., Paonessa, C., Kakooee, R., Vogel, M., & Perruchoud, D. (2025, April 22). *Fine-tuning Whisper on Low-Resource Languages for Real-World Applications*. arXiv.
- William, E., & Zahra, A. (2025). Speech Recognition Dengan Whisper Dalam Bahasa Indonesia. *Action Research Literate*, 9, 386–397.
- Zakariya, D. M., & Shafwan, M. H. (2023, Mei 13). *The Miracles of the Qur'an According to Classical and Modern Islamic Scientists and the Conformity of the Qur'an and Science and Technology*. 155–162. Atlantis Press.

Zevario, R. E., Chen, Y.-W., Fu, S.-W., Tsao, Y., Wang, H.-M., & Fuh, C.-S. (2024).

A Study On Incorporating Whisper For Robust Speech Assessment. 2024

IEEE International Conference on Multimedia and Expo (ICME), 1–6.