JURNAL ELTEK
Vol. 23, No. 1, April 2025, hal. 1 ~ 8
p-ISSN: 1693 – 4024 | e-ISSN: 2355-0740
DOI: 10.33795/eltek.v23i1.6951

Optimasi penerjemahan bahasa asing dengan teknologi IoT pada kelas internasional Politeknik Negeri Malang

Rizky Ardiansyah[1], Berliana Bastiar[2], Adzikirani[3], Dianthy Marya[4], Atik Novianti[5]
e-mail: [1]rizkyardiansyah@polinema.ac.id, [2]berlianabastiar1903@gmail.com, [3]adzikirani@polinema.ac.id, [4]dianthy@polinema.ac.id, [5]atiknovianti@polinema.ac.id
[1,2,3,4,5]Jurusan Teknik Elektro, Politeknik Negeri Malang, Indonesia

Informasi Artikel

Riwayat Artikel

Diterima 3 Maret 2025
Direvisi 16 April 2025
Diterbitkan 24 April 2025

Kata kunci:

CNN
Deep Learning
Kelas Internasional
Speech to Text

ABSTRAK

Perbedaan bahasa mencerminkan keberagaman budaya dan sejarah suatu negara. Memahami perbedaan ini memperkaya wawasan dan meningkatkan komunikasi global. Dalam rapat internasional, peserta yang tidak memahami bahasa asing sering mengalami kesulitan dalam berpartisipasi aktif, terutama di ruang tertutup, sehingga menghambat pencapaian tujuan diskusi. Penelitian ini mengembangkan sistem Speech to Text berbasis deep learning, menggunakan Mel-Frequency Cepstral Coefficients (MFCC) untuk menganalisis sinyal suara dan Convolutional Neural Network (CNN) untuk meningkatkan akurasi pengenalan serta koreksi ejaan. Raspberry Pi digunakan sebagai penghubung antara headset dan sistem, memungkinkan penerjemahan real-time. Sistem ini juga mengintegrasikan DeepL API sebagai mesin penerjemah dan Text to Speech untuk menghasilkan keluaran suara. Sebagai studi kasus, pengujian dilakukan di Politeknik Negeri Malang kelas internasional, dengan peserta yang terdiri dari mahasiswa dan pengajar yang memiliki latar belakang bahasa yang berbeda. Hasil pengujian menunjukkan metode CNN mencapai rata-rata akurasi 98%, membuktikan efektivitasnya dalam meningkatkan transkripsi dan penerjemahan otomatis. Sistem ini mampu mendukung komunikasi lintas bahasa secara efisien dan real-time, memfasilitasi pemahaman dalam diskusi multibahasa, dan meningkatkan interaksi di lingkungan akademis internasional.

Keywords:

CNN
Deep Learning
International class
Speech to Text

ABSTRACT

Language differences reflect the cultural and historical diversity of a country. Understanding these differences enriches knowledge and enhances global communication. In international meetings, participants who do not understand foreign languages often face difficulties in actively participating, especially in closed settings, thus hindering the achievement of discussion goals. This research develops a Speech to Text system based on deep learning, using Mel-Frequency Cepstral Coefficients (MFCC) to analyze sound signals and Convolutional Neural Networks (CNN) to improve accuracy in recognition and spelling correction. A Raspberry Pi is used as a connector between the headset and the system, enabling real-time translation. The system also integrates the DeepL API as the translation engine and Text to Speech for generating audio output. As a case study, testing was conducted at Politeknik Negeri Malang's international class, with participants consisting of students and instructors from diverse language backgrounds. The test results showed that the CNN method achieved an average accuracy of 98%, demonstrating its effectiveness in improving transcription and automatic translation. This system effectively supports real-time, cross-language communication, facilitates understanding in multilingual discussions, and enhances interaction in the international academic environment.

Penulis Korespondensi:
Rizky Ardiansyah,
Jurusan Teknik Elektro,
Politeknik Negeri Malang,
Jl. Sukarno Hatta No. 9, Malang, Jawa Timur, Indonesia.
Email: rizkyardiansyah@polinema.ac.id

1. PENDAHULUAN

Perbedaan bahasa dapat memengaruhi secara signifikan perkembangan komunikasi. Tantangan komunikasi muncul karena hambatan bahasa, terutama ketika bertemu orang baru atau membentuk persahabatan. Bahasa berfungsi sebagai alat utama untuk komunikasi yang efektif, yang menjadikannya penting untuk memiliki alat yang dapat membantu dalam penerjemahan bahasa selama percakapan. Masalah ini telah memunculkan inovasi dalam bentuk alat penerjemah, seperti alat berbasis headset Internet of Things (IoT) yang dirancang untuk memfasilitasi komunikasi antara individu yang berbicara dalam bahasa yang berbeda [1][2][3].

Penelitian sebelumnya yang berjudul "Pengembangan Aplikasi Text-to-Speech untuk Bahasa Indonesia" menggunakan metode Automata Finite State berbasis Android. Ditulis oleh Rieke Adriati W, Herman Tolle, dan Onny Setyawati, penelitian ini mengembangkan aplikasi Text-to-Speech (TTS) berbasis Android untuk Bahasa Indonesia, memanfaatkan Google API untuk TTS dan data suara dalam bahasa Inggris, yang memungkinkan aplikasi ini digunakan secara mobile di mana saja [4].

Penelitian lainnya berjudul "Mobile Gramabot: Pengembangan Aplikasi Chatbot untuk Pembelajaran Interaktif Tata Bahasa Jerman" yang ditulis oleh M. Kharis, berfokus pada pembuatan aplikasi chatbot untuk mahasiswa Indonesia yang belajar bahasa Jerman. Aplikasi ini, yang dapat diakses melalui smartphone dan PC desktop, menggunakan metode String Matching, meningkatkan pengalaman pembelajaran dengan sistem Gramabot [5]. Penelitian ini menghadirkan beberapa inovasi signifikan dibandingkan dengan penelitian sebelumnya. Salah satu inovasi utama adalah penggunaan Convolutional Neural Network (CNN) dalam sistem Speech-to-Text untuk meningkatkan akurasi pengenalan suara dan penerjemahan bahasa. Berbeda dengan pendekatan sebelumnya yang menggunakan metode Automata Finite State berbasis Android untuk aplikasi Text-to-Speech (TTS), penelitian ini memanfaatkan teknologi deep learning yang lebih canggih dan efektif dalam mengenali suara dengan akurasi yang lebih tinggi. Selain itu, penelitian ini juga mengintegrasikan DeepL API sebagai mesin penerjemah otomatis, yang dikenal memiliki akurasi lebih baik dibandingkan dengan penerjemah otomatis lain yang digunakan dalam penelitian sebelumnya. Inovasi lainnya adalah penerjemahan real-time, yang memungkinkan komunikasi antarbahasa berlangsung langsung, tanpa adanya keterlambatan yang sering terjadi pada sistem sebelumnya. Dengan menggabungkan teknologi Speech-to-Text, DeepL Translator, dan Text-to-Speech, sistem ini tidak hanya mentranskripsikan suara menjadi teks, tetapi juga langsung menerjemahkannya dan mengeluarkan hasil terjemahan dalam bentuk suara, yang sangat berguna dalam situasi komunikasi internasional. Keberadaan Raspberry Pi sebagai platform perangkat keras juga menjadi inovasi tersendiri, karena menawarkan solusi dengan biaya rendah namun tetap efektif, memudahkan penerapan sistem ini di berbagai lingkungan pendidikan dengan anggaran terbatas. Dengan demikian, penelitian ini tidak hanya memberikan solusi penerjemahan yang lebih cepat dan akurat, tetapi juga membuka peluang untuk penggunaan teknologi yang lebih terjangkau dan dapat diterapkan secara luas.

Sebagai bagian dari penelitian ini, dilakukan studi kasus di Politeknik Negeri Malang, Kelas Internasional di bawah Program Studi Jaringan Telekomunikasi Digital di Jurusan Teknik Elektro. Penelitian ini bertujuan untuk meningkatkan komunikasi dan pemahaman bahasa di antara mahasiswa internasional dan dosen dengan memanfaatkan sistem penerjemahan suara secara real-time. Sistem ini berhasil memfasilitasi komunikasi multibahasa dan meningkatkan lingkungan pembelajaran dengan memungkinkan interaksi yang lebih mudah antar bahasa yang berbeda dalam setting kelas.

2. METODE PENELITIAN

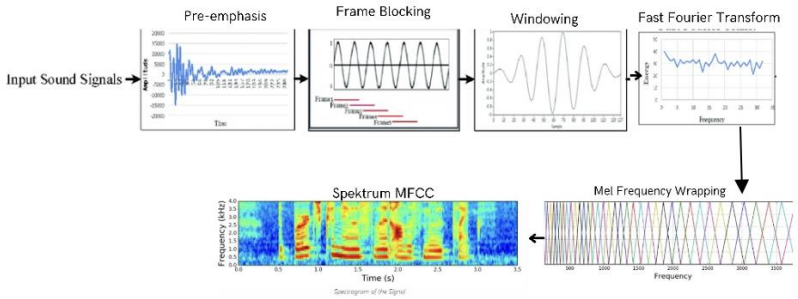
2.1 *Convolutional Neural Network* (CNN)

Convolutional Neural Network (CNN) atau jaringan saraf convolutional adalah arsitektur yang umum digunakan dalam pengolahan gambar dan pengenalan suara. Dalam konteks pengenalan suara, CNN bertugas untuk menganalisis sinyal audio dan mengklasifikasikannya menjadi unit-unit yang lebih kecil, seperti fonem atau kata [6]. Arsitektur CNN untuk pengenalan suara terdiri dari beberapa lapisan. Pertama, Convolutional Layer yang menerima masukan dan menggunakan filter untuk mengekstraksi fitur dari suara, dengan parameter yang dapat disesuaikan seperti ukuran filter, stride, dan padding [7][8][9]. Selanjutnya, Max Pooling Layer digunakan untuk mengurangi dimensi data, menjaga konsistensi ukuran data, dan memudahkan pengendalian.

ReLU Layer meningkatkan representasi model dengan menambahkan nonlinieritas melalui fungsi aktivasi Rectified Linear Units (ReLU). Setelah itu, Fully Connected Layer menggabungkan semua fitur yang diekstraksi untuk melakukan klasifikasi, dengan beberapa lapisan tersembunyi dan fungsi loss untuk optimasi. Terakhir, Softmax Activation digunakan untuk klasifikasi multi-kelas, menghasilkan probabilitas yang belum dinormalisasi untuk setiap kelas [10][11]. Dengan pendekatan ini, CNN dapat meningkatkan akurasi dalam pengenalan suara dan mengelola informasi secara efisien.

2.2 *Mel-Frequency Cepstral Coefficients* (MFCC)

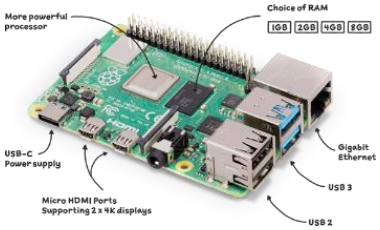
MFCC (*Mel-Frequency Cepstral Coefficients*) adalah transformasi sinyal audio yang mengubahnya ke dalam skala frekuensi non-linier menggunakan spektrum daya logaritma pendek, menyesuaikan dengan cara pendengaran manusia merespons frekuensi. Proses MFCC dimulai dengan pre-emphasis, yaitu filter untuk mengurangi noise dan meningkatkan rasio sinyal terhadap noise. Selanjutnya, frame blocking memecah sinyal menjadi frame-frame kecil, masing-masing berisi N sampel, dengan pemisahan antar frame sejauh M sampel, untuk analisis sinyal yang lebih akurat. Windowing digunakan untuk menghaluskan spektrum dan mengurangi efek diskontinuitas pada ujung frame. Fast Fourier Transform (FFT) mengubah sinyal dari domain waktu ke domain frekuensi. Terakhir, Mel Frequency Wrapping (MFW) menerapkan filter bank untuk mengubah hasil FFT ke dalam mel-scale, menyesuaikan dengan cara pendengaran manusia dan menghasilkan filter bank kritikal. Proses ini memungkinkan representasi suara yang lebih sesuai dengan persepsi manusia terhadap frekuensi [12][13][14].



Gambar 1. Arsitektur MFCC

2.3 Raspberry Pi 4 Model B

Pada Gambar 2. Raspberry Pi 4 Model B adalah sebuah komputer papan tunggal (*single-board computer computer*) atau SBC seukuran kartu kredit.



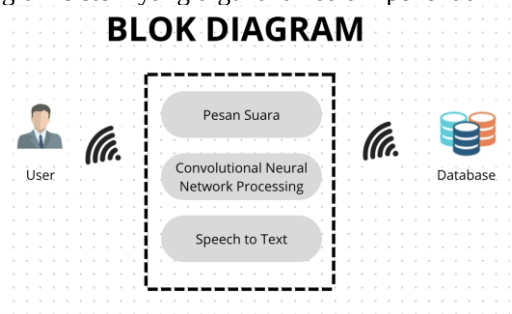
Gambar 2. Raspberry Pi

Pada Gambar 2 Raspberry pi digunakan untuk menjalankan program perkantoran, permainan komputer, dan sebagai pemutar media hingga video beresolusi tinggi. Penggunaan nya memberikan kinerja dekstop yang sebanding dengan sistem. Memory Memiliki Opsi pilihan RAM yaitu, 1GB, 2GB or 4GB LPDDR4-2400 SDRAM. *Connectivity Dual-band* 802.11ac wireless networking, 2.4 GHz and 5.0 GHz IEEE 802.11ac *wireless*.

3. HASIL DAN PEMBAHASAN

Proses yang digambarkan dalam blok diagram alat penerjemah bahasa dimulai dengan input suara dari pengguna, yang dapat berupa ucapan atau percakapan yang disampaikan melalui perangkat input seperti

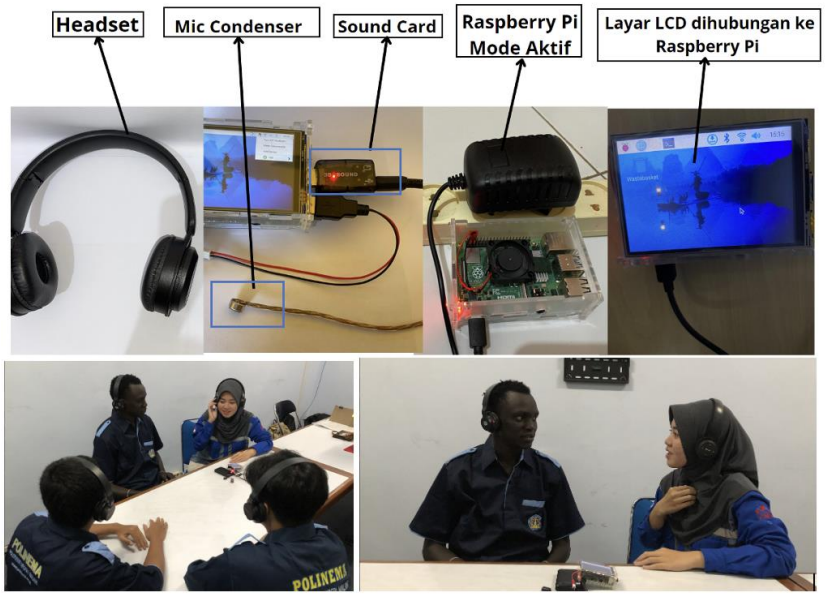
headset atau mikrofon. Sinyal suara ini kemudian diterima oleh sistem untuk diproses dalam tahap Speech to Text, di mana suara yang diucapkan oleh pengguna diubah menjadi teks. Selanjutnya, teks yang dihasilkan akan dianalisis menggunakan Convolutional Neural Network (CNN), yang berfungsi untuk meningkatkan akurasi dalam mengenali kata-kata dan memperbaiki potensi kesalahan yang terjadi selama proses transkripsi. Setelah teks dianalisis, data yang telah diproses akan disesuaikan dengan informasi yang ada dalam database, yang menyimpan berbagai pola kalimat, kamus, dan data terkait lainnya yang mendukung penerjemahan. Akhirnya, hasil terjemahan yang akurat akan diberikan kepada pengguna, baik dalam bentuk teks atau suara, yang memungkinkan komunikasi lintas bahasa berlangsung secara lebih efektif dan real-time. Proses ini memastikan alat penerjemah bekerja dengan akurat dan responsif, memenuhi kebutuhan komunikasi dalam konteks internasional. Blok diagram sistem yang digunakan dalam penelitian ini merujuk pada Gambar 3.



Gambar 3. Blok Diagram Alat Penerjemah Bahasa

CNN telah membuktikan kekuatannya dalam pemrosesan gambar, namun penerapannya dalam pemrosesan ucapan juga menarik perhatian karena kemampuannya mengenali fitur kompleks dalam sinyal audio. CNN, yang awalnya dirancang untuk tugas pengenalan gambar, dapat diadaptasi untuk memproses spektogram atau representasi visual dari data ucapan. Pada dasarnya CNN mampu mengekstrak fitur-fitur penting dari suara, seperti pola frekuensi atau keberadaan elemen audio tertentu.Lapisan konvolusi di CNN memainkan peran penting dalam mendeteksi fitur-fitur ini dengan cara yang mirip dengan mendeteksi pola visual dalam gambar. Namun, dalam konteks suara, lapisan konvolusi dapat mengidentifikasi pola spektral atau temporal, seperti harmonik pada frekuensi tertentu atau perubahan intensitas suara yang cepat. Selain itu, lapisan pengumpulan di CNN membantu mengurangi dimensi data suara, menyimpan informasi penting, dan mengurangi kompleksitas komputasi. CNN dalam konteks ucapan sering digunakan untuk tugas-tugas seperti pengenalan ucapan, di mana model dapat “mempelajari” pola yang berkaitan dengan fonem atau kata. Oleh karena itu, meskipun CNN terutama dikembangkan untuk tugas pengenalan gambar, kemampuannya untuk menangani representasi visual data ucapan menjadikannya alat yang ampuh untuk tugas pemrosesan ucapan yang kompleks.

Hasil dari pembuatan perangkat yang digunakan pada penelitian ini ada pada Gambar 4:



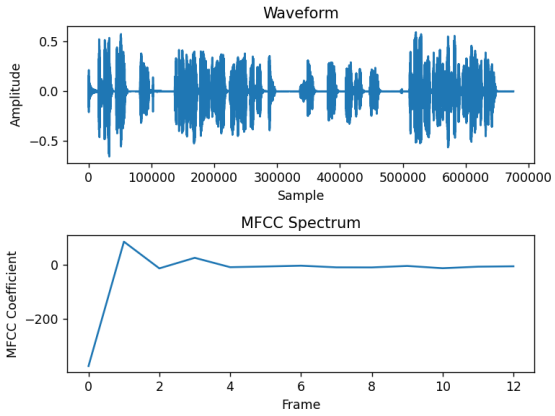
Gambar 4. Pengujian Alat

Sistem ini dirancang untuk memfasilitasi penerjemahan suara ke teks secara real-time dengan menggunakan beberapa komponen utama. Headset digunakan oleh peserta untuk menangkap input audio, memastikan suara yang jelas dan fokus untuk proses penerjemahan. Mic condenser berfungsi untuk meningkatkan kualitas audio, mengurangi kebisingan latar belakang, dan meningkatkan kejernihan suara yang ditangkap. Sound card menghubungkan mikrofon ke Raspberry Pi, yang bertindak sebagai unit pemrosesan utama untuk sistem ini. Raspberry Pi yang ditampilkan dalam mode aktif menjalankan perangkat lunak yang diperlukan untuk memproses suara yang tertangkap dan melakukan transkripsi serta penerjemahan secara real-time. Selain itu, layar LCD yang terhubung ke Raspberry Pi menampilkan output, seperti teks terjemahan atau status sistem, memungkinkan pengguna untuk memantau perkembangan penerjemahan.

Pada bagian bawah Gambar 4, terlihat para peserta yang sedang berinteraksi dengan sistem. Mereka menggunakan headset sambil berbicara, yang memungkinkan sistem untuk memproses ucapan mereka menjadi teks dan memberikan terjemahan secara real-time, seperti yang ditunjukkan dalam pengaturan ini. Seluruh sistem dirancang untuk memastikan bahwa bahasa pemograman python mampu berjalan dengan stabil, responsif, dan sesuai dengan spekasi yang telah ditetapkan. Dengan mengimplementasikan Raspyberry Pi 4 Type B sebagai platform perangkat keras miniatur yang mampu memproses dari input suara sehingga menghasilkan output, terdapat soundcard sebagai penghubung antara mic condeser berfungsi perangkat menangkap suara dari pembicara dan akan diproses pada sistem menggunakam *Library Speech Recognition*, selanjutnya ditambahkan ekstraksi fitur MFCC yang akan menampilkan hasil spektrum audio dari suara pembicara metode dan menghasilkan output berupa nilai vektor selanjutnya akan diproses pada metode *Convolutional Neural Network* (CNN) agar menghasilkan hasil yang lebih akurat, kemudian diterjemahkan menggunakan translator DeepL API, dan dikonversi *Text to Speech* yang hasil akhir berupa suara hasil terjemahan yang mampu didengarkan.

3.1. Pengujian Accuracy dan Delay Speech to Text Menggunakan Metode CNN

Pengujian dilakukan untuk mengetahui tentang *accur*acy dan *delay speech to text* menggunakan metode CNN. Audio yang masuk akan diproses oleh model CNN yang dilatih pada data audio dan teks. Sehingga hasilnya akan digunakan evaluasi untuk meningkatkan algoritma dan model *speech to text.* Berdasarkan hasil pengujian sistem yang telah dilakukan, dapat disimpulkan bahwa penggunaan Speech-to-Text berbasis Deep Learning dalam pengenalan suara dan penerjemahan bahasa, khususnya menggunakan Convolutional Neural Networks (CNN), telah menunjukkan kinerja yang memuaskan. Hasil uji coba yang dilakukan pada beberapa input suara menunjukkan tingkat akurasi yang sangat baik, dengan rata-rata akurasi mencapai 98,42%. Sistem ini mampu mendeteksi dan mentranskripsi suara dengan tingkat kesalahan yang rendah, yang terlihat dari nilai loss yang konsisten rendah pada setiap percakapan yang diuji. Beberapa contoh frasa dalam bahasa Inggris berhasil diubah menjadi teks dengan akurasi tinggi, seperti dalam kalimat "How are you?" yang memiliki akurasi 98%, dan "See you next week" yang mencapai 100%. Ini menunjukkan bahwa sistem ini dapat bekerja dengan baik dalam real-time untuk mengkonversi suara menjadi teks dan mendukung komunikasi antarbahasa secara efisien. Dengan menggunakan teknologi Raspberry Pi sebagai penghubung antara headset dan sistem, serta integrasi dengan DeepL API untuk penerjemahan, sistem ini dapat memberikan hasil terjemahan yang dapat langsung didengar, memfasilitasi komunikasi lintas bahasa dalam pengaturan internasional dengan cepat dan akurat.



Gambar 5. Grafik Spektrum Input Suara

Gambar 5 menunjukkan dua grafik yang terkait dengan ekstraksi fitur Mel-Frequency Cepstral Coefficients (MFCC) dari sinyal suara, yang digunakan dalam proses Speech-to-Text pada sistem penerjemahan suara ke teks. Grafik pertama adalah waveform yang menggambarkan sinyal suara mentah yang diambil sebagai input. Pada sumbu x terlihat sampel dari sinyal suara, sementara sumbu y menunjukkan

amplitudo sinyal tersebut. Waveform ini memberikan gambaran visual dari fluktuasi amplitudo suara pada berbagai titik waktu, yang merupakan langkah awal dalam proses pengenalan suara.

Selanjutnya, grafik kedua menunjukkan MFCC spectrum, yang menggambarkan koefisien MFCC dari sinyal suara yang telah diproses. Pada sumbu x terdapat frame yang menunjukkan pembagian sinyal suara menjadi potongan-potongan kecil untuk analisis lebih lanjut, sedangkan pada sumbu y terlihat nilai koefisien MFCC yang mewakili karakteristik frekuensi penting dari sinyal suara. Proses ekstraksi MFCC ini penting karena mengubah sinyal suara menjadi representasi numerik yang lebih mudah dianalisis oleh sistem, sesuai dengan cara pendengaran manusia. Fitur-fitur yang diekstraksi ini kemudian digunakan dalam Convolutional Neural Networks (CNN) untuk meningkatkan akurasi dalam pengenalan suara, yang pada gilirannya mendukung penerjemahan suara ke teks secara lebih akurat dan efisien.

3.2. Pengujian Spell Checking Translator

Pengujian akan dilakukan dengan menggunakan kumpulan data teks dalam Bahasa Jerman yang akan diterjemahkan ke Bahasa Indonesia. Kumpulan data akan diproses dengan *translator DeepL* kemudian akan menghasilkan berbagai jenis kesalahan ejaan atau hasil terjemah. Hasil yang didapatkan nantinya digunakan evaluasi agar meningkatkan fitur dan efektivitas spell checking pada *translator DeepL*.

Tabel 1. Hasil Pengujian Spell Checking

No	Terjemahan	Text	Informasi
1	Bagaimana Kabarmu	Text: How are you Translation: Apa Kabarmu	Tidak Sesuai
2	Berapa umurmu Saya berumur 20 tahun	Text: How old are you I am 20 years old Translation: Berapa umurmu Saya berumur 20 tahun	Sesuai
3	Bisakah kamu membantu saya, tentu saja	Text : Can you help me of course Translation: Bisakah kamu membantu saya, tentu saja	Sesuai
4	Dari mana asalmu Saya berasal	Text : Where do you come from I come from in Translation: Dari mana asalmu Saya berasal	Sesuai
5	Bagaimana kabarmu, aku baik-baik saja, begitu juga kamu	Text : How are you, I'm doing well and you Translation : Bagaimana kabarmu, aku baik-baik saja, begitu juga kamu	Sesuai
6	Sampai jumpa minggu depan	Text : See you next week Translation : Sampai jumpa minggu depan	Sesuai
7	Senang bertemu dengan anda	Text : Nice to meet her Translation: Saya senang bertemu dengan anda	Tidak Sesuai

Berdasarkan Tabel 1 menggunakan kata dari pengujian sebelumnya hasil Translator DeepL menunjukkan dengan 7 kali pengujian didapatkan 5 hasil terjemah “Sesuai” yang diinginkan, dan terdapat 2 hasil “Tidak Sesuai”, berdasarkan hasil yang didapatkan terjemahan yang dianggap sesuai adalah "Berapa umurmu Saya berumur 20 tahun" untuk "How old are you I am 20 years old". Terjemahan ini langsung mencerminkan makna yang tepat dan alami dalam bahasa Indonesia. Namun, terjemahan yang tidak sesuai adalah "Saya senang bertemu dengan anda" untuk "Nice to meet her". Terjemahan ini salah karena menggunakan "anda", yang lebih formal, sementara dalam konteks ini yang dimaksud adalah "her" (perempuan), sehingga seharusnya terjemahan yang lebih tepat adalah "Senang bertemu dengan dia".. Hal ini menunjukkan algoritma sistem perlu diperbaiki dan kesulitan memahami konteks kalimat atau frasa yang kompleks.

3.3. Pengujian Text To Speech

Pengujian *text to speech* bertujuan untuk mengevaluasi kinerja dan kualitas sistem TTS dalam menghasilkan suara yang akurat, jelas, dan natural

Tabel 2. Hasil Pengujian Text to Speech

No	Input Text	Output Speech	Akurasi
1	Apa Kabarmu	Apa Kabarmu	100 %
2	Berapa umurmu Saya berumur 20 tahun	Berapa umurmu Saya berumur 20 tahun	100 %
3	Bisakah kamu membantu saya, tentu saja	Bisakah kamu membantu saya, tentu saja	100 %
4	Dari mana asalmu Saya berasal	Dari mana asalmu Saya berasal	100 %
5	Bagaimana kabarmu, aku baik-baik saja, begitu juga kamu	Bagaimana kabarmu, aku baik-baik saja, begitu juga kamu	100 %
6	Sampai jumpa minggu depan	Sampai jumpa minggu depan	100 %
7	Saya Senang bertemu dengan Anda	Saya senang bertemu dengan Anda	100 %

Berdasarkan Tabel 2 Hasil Pengujian *Text to Speech* (TTS) menunjukkan dengan 8 kali percobaan yang menginputkan teks kemudian diubah menjadi suara dapat berjalan dengan sukses. Sehingga didapatkan akurasi *text to speech* sebesar 100%, hal ini menunjukkan sistem mampu bekerja dengan baik adapun algoritma yang digunakan sesuai dengan keinginan, terlebih lagi dapat diimplementasi kan pada alat yang berupa *headset*.

3.4. Pengujian Kuisoner

Hasil dari penyebaran kuisioner dengan mencoba dan menganalisis kinerja sistem translator terdapat responden yang berjumlah 30 responden yang terdiri dari mahasiwa Jurusan Bahasa Jerman Universitas Negeri Malang, dan teman-teman disekitar, menunjukkan beberapa pandangan dan pengalaman yang menarik terkait efektivitas serta keakuratan sistem tersebut, dimana sebagian besar responden memberikan umpan balik balik positif mengenai kemudahan pengguna, sementara beberapa lainnya menyampaikan mengenai kesalahan terjemahan. Sehingga hal ini nantinya menandakan perlunya pengembangan atau perbaikan untuk kedepannya agar dapat memenuhi kebutuhan pengguna.

Tabel 3. Hasil Pengujian Kuisoner

Uji	Hasil
Penggunaan Bahasa Jerman	53,70% responden "Sangat setuju dan setuju" bahwa mereka sering menggunakan bahasa Jerman dalam kehidupan sehari-hari; 15,00% tidak setuju.
Waktu Terlambat Penerjemahan	55% responden mengalami keterlambatan waktu lebih dari 10 detik untuk hasil terjemahan.
Kepuasan dengan Penerjemahan DeepL	61,90% "Sangat setuju" dan 28,00% "Setuju" dengan kualitas terjemahan dari DeepL Translator.
Penerjemahan yang Tidak Tepat oleh DeepL	47% "Tidak setuju" dengan penerjemahan yang tidak tepat oleh DeepL, sementara 42% mengalami kata/ungkapan yang diterjemahkan dengan salah.
Kebisingan pada Hasil Suara Terjemahan	40,40% "Sangat setuju dan setuju" bahwa sistem dapat menyaring kebisingan dengan efektif; 58% "Tidak setuju dan sangat tidak setuju."
Kelayakan dan Kepuasan	98% responden "Sangat setuju dan setuju" mengenai kualitas terjemahan yang tinggi dan akurat yang diberikan.

4. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, beberapa kesimpulan dapat diambil. Pertama, alat penerjemah yang dirancang menggunakan perangkat Raspberry Pi melalui proses pengujian sebanyak delapan kali menunjukkan bahwa waktu yang dibutuhkan untuk menghasilkan hasil transkripsi teks adalah 2 detik 4ms per langkah. Meskipun waktu tersebut terbilang cukup lama, hal ini disebabkan oleh metode CNN yang digunakan dalam proses pengenalan ejaan kata, yang membutuhkan beberapa kali pengulangan bobot atau epoch sebanyak 10 kali. Meskipun prosesnya memerlukan waktu yang cukup panjang, sistem berhasil menghasilkan terjemahan yang cukup akurat, dengan rata-rata akurasi mencapai 98%. Proses penerjemahan yang relatif lama ini disebabkan oleh kompleksitas dari algoritma CNN yang membutuhkan waktu lebih lama untuk memproses input suara dan mengoptimalkan hasil transkripsi. Dalam penelitian ini, meskipun keterlambatan tersebut terjadi, kualitas terjemahan yang dihasilkan tetap memadai untuk mendukung komunikasi real-time dalam konteks internasional.

Kedua, sistem Speech-to-Text yang menggunakan Metode CNN, DeepL Translator, dan Text-to-Speech dapat berfungsi dengan baik sesuai tujuan, serta memberikan hasil terjemahan yang akurat dan berkualitas

tinggi. Berdasarkan data dari 30 responden mahasiswa kelas internasional, 98% menunjukkan tingkat kelayakan dan kepuasan yang tinggi terhadap kualitas terjemahan yang dihasilkan. Walaupun terdapat sedikit keterlambatan dalam proses penerjemahan, sistem ini tetap dapat berfungsi dengan baik dalam skenario real-time, berkat akurasi yang tinggi dalam pengenalan suara dan terjemahan yang dihasilkan."

Terakhir, perancangan headset penerjemah dengan menggunakan Metode CNN menunjukkan hasil yang memadai, dengan algoritma yang efektif dalam mengubah data suara menjadi teks, yang tercermin dari rata-rata akurasi sebesar 98%. Namun, untuk mengoptimalkan kecepatan proses, beberapa strategi, seperti pengurangan jumlah epoch dalam pelatihan CNN atau penggunaan perangkat keras dengan kemampuan pemrosesan lebih cepat, dapat diimplementasikan untuk mengurangi latensi lebih lanjut dan memastikan penerjemahan yang lebih cepat dan lebih efisien.

5. DAFTAR PUSTAKA

[1] "kemendikbud.go.id," 4 November 2021. [Online]. Available: https://www.kemdikbud.go.id/main/blog/2021/11/konsorsium-program-internasional-upaya-percepatan-internasionalisasi-di-lingkup-ptv.. [Accessed 24 Juni 2024].

[2] R. Arvyanda, E. Fernandito and P. Landung, "Analisis Pengaruh Perbedaan Bahasa dalam Komunikasi," *JURNAL HARMONI NUSA BANGSA,* vol. 1, no. 1, pp. 1-14, 2023.

[3] M. Atik Yuliyani, "Tantangan Kerja Sama Internasional Bidang Pendidikan Tinggi Islam (Studi Kasus: UIN Syarif Hidayatullah Jakarta,UIN Walisongo Semarang, dan UIN Sunan KalijagaYogyakarta)," *Jurnal Pendidikan* , vol. 12, no. 02, pp. 1131-1154, 2023.

[4] M. Penke, "dw.com," 10 09 2021. [Online]. Available: https://www.dw.com/id/hubungan-bilateral-indonesia-jerman/a-58979889. [Accessed 16 Juni 2024].

[5] B. R. Setiawan, A. Aranta and B. Irmawati, "SPEECH TO TEXT BAHASA SASAK MENGGUNAKAN EXTRAKSI FITUR MEL-FREQUENCY CEPSTRAL COEFFICIENTS DAN KLASIFIKASI CONVOLUTIONAL NEURAL NETWORKS," *Jurnal Teknologi Informasi, Komputer dan Aplikasinya (JTIKA),* vol. 5, pp. 21-32, 2023.

[6] S. T. Maria Isabel Rivas Ginel, "Machine Translation and Gender Biases in Video Game Localisation: A," *Journal of Data Mining and Digital Humanities,* Vols. 2416-5999, pp. 1-10, 2022.

[7] R. Setiawan, dicoding, 8 September 2021. [Online]. Available: https://www.dicoding.com/blog/apa-itu-internet-of-things/. [Accessed 24 Desember 2023].

[8] A. N. Sarirotul Ilahiyah, *Jurnal Sistem & Teknologi Informasi Indonesia,* vol. 3, no. Implementasi Deep Learning, pp. 49-56, 2018.

[9] W. B. Pratiwi, A. Aranta and G. S. Nugraha, "Analisis Kebutuhan Dataset Algoritma Speech to Text Bahasa Sasak Menggunakan Perbandingan Data Suara Bahasa Inggris Pada Metode CNN," *J-COSINE (Journal of Computer Science and Informatics Engineering),* vol. 7, pp. 165-173, 2023.

[10] Paramitha, Pradnja, I. Arya, H. B. Kusnawan and M. Ernawati, "Performance Comparison of Deep Learning Algorithm for Speech Emotion Recognition," *J Cosine Journal,* vol. 6, no. 2, pp. 15-24, 2022.

[11] O. D. Nurhayati, D. Eridani and M. H. Tsalavin, "SISTEM ISYARAT BAHASA INDONESIA (SIBI) METODE," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK),* vol. 6, pp. 819-826, 2022.

[12] D. Frese, "pixabay.com," pixabay, [Online]. Available: https://pixabay.com/id/service/about/. [Accessed 29 06 2024].

[13] R. Hadyan, "hypeabis.id," 10 Agustus 2021. [Online]. Available: https://hypeabis.id/read/3124/simak-dulu-perbedaan-headphone-headset-earphone-dan-handsfree-jangan-tertukar. [Accessed 21 Desember 2023].

[14] S. S. H. Emad, "folio3.ai," folio3, 29 02 2024. [Online]. Available: https://www.folio3.ai/blog/how-to-integrate-google-speech-to-text-api-into-your-applications/. [Accessed 29 06 2024].