

Sistem Pemilihan Menu pada Kursi Roda Pintar menggunakan Suara berbasis MFCC dan CNN pada Jetson TX2 NX

Abdan Idza Hurmuzi¹, Fitri Utamingrum²

Program Studi Teknik Komputer, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹abdan_idza@student.ub.ac.id, ²f3_ningrum@ub.ac.id

Abstrak

Disabilitas adalah kondisi dimana tubuh ataupun mental mengalami gangguan yang dapat membuat seseorang kesulitan dalam melakukan aktivitas ataupun berinteraksi dengan lingkungan sekitar. Orang yang memiliki keterbatasan pada kemampuan berpindah tempat seperti berjalan umumnya menggunakan bantuan kursi roda, yaitu alat yang digunakan untuk mengatasi keterbatasan tersebut. Kursi roda tradisional memiliki kekurangan pada pengguna yang menderita paralisis pada anggota gerak seperti lengan. Kursi roda pintar umumnya dibekali piranti berbasis hand-operated seperti joystick ataupun layar sentuh untuk pengoperasiannya. Akan tetapi terdapat beberapa batasan pada Human Machine Interface tersebut terutama pada penderita tetraplegia, locked-in syndrome dan amyotrophic lateral sclerosis (ALS). Penelitian ini akan berfokus pada penerapan MFCC dan Convolutional Neural Network (CNN) untuk sistem pemilihan menu pada Kursi Roda Pintar menggunakan perintah suara dalam bahasa Indonesia. Pada penelitian ini sebanyak 5,600 data perintah suara telah dikumpulkan. Data tersebut mencakup 14 perintah, yaitu, "Satu", "Dua", "Tiga", "Empat", "Lima", "Enam", "Mati", "Nyala", "Oke", "Tidak", "Info", "Next", "Back", dan "Keluar". Hasil dari pengujian sistem dengan dua jenis gender mendapatkan hasil yang terbilang sangat baik dengan nilai akurasi sebesar 99.64%. Selain itu, besar waktu komputasi adalah 0,2103 detik dalam melakukan proses pengenalan suara.

Kata kunci: ALS, CNN, Disabilitas, MFCC, Perintah Suara.

Abstract

Disability is a condition where the body or mind experiences impairments that can make it difficult for someone to perform activities or interact with their environment. People with mobility limitations, such as difficulty walking, generally use a wheelchair, which is a device used to overcome these limitations. Traditional wheelchairs have limitations for users who suffer from paralysis in limbs, such as the arms. Smart wheelchairs are typically equipped with hand-operated devices like joysticks or touchscreens for operation. However, there are limitations in the Human Machine Interface, especially for individuals with tetraplegia, locked-in syndrome, and amyotrophic lateral sclerosis (ALS). This study focuses on the application of MFCC and Convolutional Neural Network (CNN) for menu selection in a Smart Wheelchair using voice commands in the Indonesian language. In this research, a total of 5,600 voice command data have been collected. The data includes 14 commands, which is, "Satu", "Dua", "Tiga", "Empat", "Lima", "Enam", "Mati", "Nyala", "Oke", "Tidak", "Info", "Next", "Back", dan "Keluar". The system testing yielded excellent results with an accuracy rate of 99.64% when tested with two genders. Additionally, the computation time was 0.2103 seconds for the speech recognition process.

Keywords: ALS, CNN, Disability, MFCC, voice command.

1. PENDAHULUAN

Disabilitas adalah kondisi dimana tubuh ataupun mental mengalami gangguan yang dapat membuat seseorang kesulitan dalam melakukan aktivitas ataupun berinteraksi dengan lingkungan sekitar (Centers for Disease Control

and Prevention, 2020). Diperkirakan terdapat 1.3 juta orang atau 16% dari total populasi dunia pada tahun 2023 menderita disabilitas. Data tersebut mencakup ketidakmampuan pada anggota gerak, seperti cacat tangan, cacat kaki, ataupun ketidakmampuan dalam berjalan. Penderita disabilitas tidak terbatas pada gender,

umur, dan ras (World Health Organization, 2023). Selain itu, diperkirakan terdapat setidaknya 2.2 miliar orang memiliki kecacatan dalam segi penglihatan. Dari 2.2 miliar tersebut, setidaknya terdapat 1 miliar orang yang mengidap kecacatan penglihatan seperti katarak sampai kebutaan (World Health Organization, 2022).

Orang yang memiliki keterbatasan pada kemampuan berpindah tempat seperti berjalan umumnya menggunakan bantuan kursi roda, yaitu alat yang digunakan untuk mengatasi keterbatasan tersebut (Sukerkar, et al., 2018). Kursi roda tradisional memiliki kekurangan pada pengguna yang menderita paralisis pada anggota gerak seperti lengan. Kursi roda pintar umumnya dibekali piranti berbasis hand-operated seperti joystick ataupun layar sentuh untuk pengoperasiannya. Akan tetapi terdapat beberapa batasan pada Human Machine Interface tersebut terutama pada penderita tetraplegia, locked-in syndrome dan amyotrophic lateral sclerosis (ALS). Oleh karena itu, untuk mengatasi permasalahan tersebut, penelitian ini akan berfokus untuk mengimplementasikan penggunaan suara dalam melakukan pemilihan menu pada kursi roda pintar menggunakan bahasa Indonesia dengan menerapkan metode *Mel-Frequency Cepstral Coefficient* (MFCC) dan *Convolutional Neural Network* (CNN) untuk melakukan pengenalan suara pengguna.

Algoritma MFCC sangat populer dan paling banyak digunakan dalam bidang speech recognition. Sebelum adanya MFCC, LPC (Linear Prediction Coefficients) dan LPCC (Linear Prediction Cepstral Coefficients) menjadi algoritma yang paling banyak digunakan bersama dengan Hidden Markov Models untuk pengenalan suara. Algoritma MFCC dibuat untuk mengurangi kompleksitas dari model dan mencapai akurasi yang tinggi (Mahmood & Kose, 2021). Selain itu, algoritma MFCC lebih unggul dari algoritma terdahulu dikarenakan kemampuan dalam meniru kemampuan telinga manusia dalam membedakan frekuensi suara. Pada pengenalan suara modern, metode deep learning digunakan untuk mencapai tujuan ASR (*Automatic Speech Recognition*). Hal tersebut didasari oleh kesulitan yang dihadapi pada saat melakukan pengenalan suara (ASR) seperti, perbedaan pola bicara, kebisingan pada lingkungan, dan sejenisnya (Mahmood & Kose, 2021). Salah satu jenis metode deep learning yang sering

digunakan dalam ASR adalah Convolutional Neural Network (CNN) (Passricha & Aggarwal, 2018). Oleh karena itu, CNN digunakan karena dapat mewujudkan performa yang baik dalam pengenalan suara (Passricha & Aggarwal, 2018). CNN juga lebih unggul apabila digunakan bersamaan dengan MFCC dibanding metode lainnya. Ditambah lagi, akurasi dan presisi yang dihasilkan pada lingkungan bising, terbilang cukup tinggi (Ashar, et al., 2020).

Penelitian ini diharapkan dapat memberikan kenyamanan dan kemudahan bagi pengguna Kursi Roda Pintar dalam melakukan pemilihan menu pada Kursi Roda Pintar menggunakan suara bagi penyandang disabilitas yang memiliki ketidakmampuan pada anggota gerak, seperti cacat tangan, cacat kaki, ataupun ketidakmampuan dalam berjalan.

2. METODOLOGI

2.1 Gambaran Umum Sistem

Sistem Pemilihan Menu pada Kursi Roda Pintar Menggunakan Suara Berbasis MFCC dan CNN pada Jetson TX2 NX akan menangkap perintah suara yang diucapkan oleh pengguna melalui mikrofon yang menghadap kepada mulut pengguna. Suara tersebut nantinya akan disimpan dalam bentuk file yang disimpan pada Jetson TX2 NX. Setelah file suara disimpan, selanjutnya dilakukan proses klasifikasi menggunakan CNN dengan masukan berupa hasil ekstraksi fitur MFCC dari file suara tersebut. Hasil keluaran dari proses klasifikasi tersebut nantinya akan ditampilkan pada Aplikasi Pemilihan Menu dengan mengirimkan data melalui websocket kepada aplikasi tersebut. Nantinya pengguna dapat melihat tampilan dari perintah suara yang diucapkan pada Aplikasi Pemilihan Menu yang ditampilkan pada layar LCD di Kursi Roda Pintar.

2.2 Teknik Pengumpulan Data

Data yang nantinya akan digunakan untuk pelatihan model pada *deep learning* untuk dapat melakukan klasifikasi, didapatkan dari *dataset* yang diambil secara mandiri (*dataset* primer). Data yang akan digunakan terdiri dari 14 kelas yaitu, "Satu", "Dua", "Tiga", "Empat", "Lima", "Enam", "Mati", "Nyala", "Oke", "Tidak", "Info", "Next", "Back", dan "Keluar".

Dataset diambil pada mahasiswa/i kelas pemrosesan sinyal A tahun ajar 2023 sebanyak 40 anak (7 Perempuan, 33 Laki-Laki), dosen

pembimbing skripsi (Perempuan), dan beberapa teman peneliti. Setiap orang mengucapkan 10 ucapan tiap kelasnya dimana 5 adalah intonasi normal dan 5 lainnya adalah intonasi yang berbeda-beda. Hal tersebut bertujuan agar data lebih variatif. Pengambilan data tersebut dilakukan menggunakan mikrofon webcam Logitech C270 HD yang sama dengan sistem pada saat proses implementasi dengan kondisi yang cukup bersih dari noise agar pengucapan dapat terdengar dengan jelas. Pengambilan dataset dilakukan di Ruang Grup Riset Computer Vision, Ruang 9.5, Fakultas Ilmu Komputer, Universitas Brawijaya, Kota Malang, Jawa Timur.

3. PERANCANGAN DAN IMPLEMENTASI

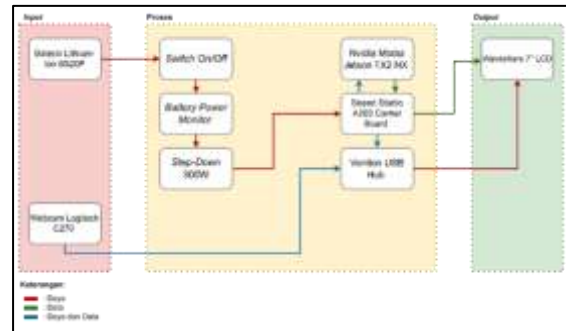
3.1. Perancangan dan Impelementasi Perangkat Keras

Dalam perancangan perangkat keras pada sistem ini, terdapat tiga komponen utama, yaitu input, proses, dan output yang ditunjukkan pada Gambar 1.

Pada bagian *input* terdiri atas komponen yang bertanggungjawab untuk menyediakan *input* bagi sistem. Komponen tersebut adalah Baterai Lithium Ion yang berfungsi untuk sumber daya utama sistem. Selain itu, adalah *Webcam* Logitech C270 yang digunakan untuk menangkap perintah suara yang diucapkan pengguna menggunakan mikrofon yang terdapat pada *Webcam* tersebut.

Pada bagian *proses* terdiri atas komponen yang bertugas untuk mengolah masukan yang diterima sistem dan menghasilkan output pada sistem. *Webcam* Logitech C270 dan LCD Waveshare 7 inci akan terhubung dengan Seeed Studio A203 Carrier Board. Selain itu, Seeed Studio A203 Carrier Board juga digunakan untuk menjalankan modul Nvidia Jetson TX2 NX dan memberi daya kepada seluruh komponen yang terhubung melalui *port* yang tersedia pada Seeed Studio A203 Carrier Board. Pada modul Nvidia Jetson TX2 NX, modul tersebut digunakan untuk menerapkan proses yang berhubungan dengan perangkat lunak sistem.

Pada bagian *output* hanya terdiri dari LCD Waveshare 7 inci untuk menampilkan keluaran dari proses pengenalan suara yang telah dilakukan oleh perangkat lunak pada Nvidia Modul Jetson TX2 NX pada Aplikasi Pemilihan Menu.



Gambar 1. Diagram Blok Perancangan Perangkat Keras.

Implementasi perangkat keras diimplementasikan pada Kerangka Kursi Roda Pinar yang telah dibuat untuk dengan mempertimbangkan kerangka kursi roda yang bersifat ringan, kokoh, portabel dan memberi kenyamanan bagi pengguna. Gambar 2 menunjukkan penampakan Kursi Roda Pinar saat mode penggunaan dan saat dilipat.



Gambar 2. Penampakan Kursi Roda saat Mode Pengguna (a) dan Mode Dilipat (b).

3.2. Perancangan Perangkat Lunak

Pada perancangan perangkat lunak, akan terbagi menjadi 4 bagian, yaitu pembuatan model CNN, sistem pengenalan suara, REST-API Server, dan Aplikasi Pemilihan Menu.

Bagian pertama, yaitu pembuatan model CNN adalah langkah untuk mendapatkan model dari hasil *training* terhadap *dataset* yang telah dikumpulkan sebelumnya. Pembuatan model tersebut dilakukan dari dua langkah. Langkah pertama adalah melakukan ekstraksi fitur MFCC dari tiap data suara pada *dataset*. Setelah seluruh data suara telah diekstraksi, langkah selanjutnya adalah melakukan *training* dari hasil ekstraksi tersebut menggunakan metode CNN sehingga menghasilkan model yang nantinya akan ditanamkan pada sistem pengenalan suara.

Bagian Kedua, yaitu sistem pengenalan suara adalah langkah untuk melakukan proses pengenalan suara dari *input* perintah suara yang diucapkan oleh pengguna. Perintah suara pengguna yang telah ditangkap oleh mikrofon

pada sistem, nantinya akan dilakukan proses ekstraksi fitur dan klasifikasi menggunakan model yang telah dibuat dan ditanamkan sebelumnya. Data suara yang telah diekstraksi menggunakan MFCC selanjutnya akan menjadi masukan pada model untuk dilakukan proses klasifikasi sehingga menjadi *output* dari data suara tersebut.

Bagian ketiga, yaitu REST-API Server adalah langkah untuk menyimpan data menu pada Kursi Roda Pintar sekaligus bertanggungjawab untuk menjalankan ataupun menghentikan program menu yang dipilih sebelumnya oleh pengguna menggunakan perintah suara. Pada saat Aplikasi Pemilihan Menu dijalankan, REST-API Server akan mengirimkan data menu yang ada pada Kursi Roda Pintar yang akan ditampilkan pada Aplikasi Pemilihan Menu tersebut. Selain itu, pada saat pengguna memilih menu yang ada pada Kursi Roda Pintar melalui Aplikasi Pemilihan Menu, aplikasi tersebut akan mengirimkan data ke REST-API Server untuk menjalankan ataupun menghentikan program menu yang telah dipilih oleh pengguna.

Bagian empat, yaitu Aplikasi Pemilihan Menu adalah langkah yang berfungsi sebagai antarmuka pengguna dengan Kursi Roda Pintar untuk menjalankan menu yang ada. Aplikasi tersebut akan menampilkan tampilan visual yang interaktif agar pengguna dapat dengan mudah memahami proses yang sedang dilakukan pada Kursi Roda Pintar. Aplikasi tersebut bergantung pada sistem pengenalan suara dan REST-API untuk dapat menjalankan prosesnya.

Pada ketergantungan dengan sistem pengenalan suara, aplikasi tersebut berfungsi untuk menerima data perintah suara yang diucapkan oleh pengguna dari proses pengenalan suara yang dilakukan oleh sistem pengenalan suara. Selain itu, pada ketergantungan dengan REST-API Server, aplikasi tersebut berfungsi untuk mendapatkan data menu yang tersedia pada Kursi Roda Pintar sekaligus berfungsi untuk mengirimkan data untuk menjalankan ataupun menghentikan program menu pada Kursi Roda Pintar ke REST-API Server.

4. PENGUJIAN DAN ANALISIS

Pada tahap pengujian dan analisis ini, peneliti menguji sekaligus menganalisis kinerja berdasarkan hasil implementasi sistem pemilihan menu pada Kursi Roda Pintar dengan mengucapkan perintah suara dengan jelas dan

cukup lantang sebanyak 1 kali tiap 14 perintah suara pada mikrofon sistem yang terdiri dari suara perempuan dan laki-laki sebanyak 8 orang.

Berdasarkan hasil pengujian tersebut, analisis dilakukan menggunakan *Classification Metrics* yang terdiri dari *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Berdasarkan komponen-komponen tersebut, didapatkan kesimpulan terhadap kinerja klasifikasi dari model yang telah dibuat menggunakan arsitektur CNN. Berdasarkan hasil pada Tabel 1, model dinilai mendapatkan hasil yang bisa dikatakan sangat bagus berdasarkan nilai yang ada pada komponen *Classification Metrics*.

Tabel 1. *Classification Metrics* dari Pengujian pada Model dengan Arsitektur CNN

| Kelas | <i>Accuracy</i> (%) | <i>Precision</i> | <i>Recall</i> | <i>F1-score</i> |
|-----------|---------------------|------------------|---------------|-----------------|
| “info” | 100 | 1 | 1 | 1 |
| “tiga” | 97,32 | 1 | 0,62 | 0,77 |
| “mati” | 100 | 1 | 1 | 1 |
| “enam” | 99,11 | 0,89 | 1 | 0,94 |
| “satu” | 100 | 1 | 1 | 1 |
| “empat” | 98,21 | 0,8 | 1 | 0,89 |
| “next” | 100 | 1 | 1 | 1 |
| “keluar” | 96,43 | 0,83 | 0,62 | 0,71 |
| “nyala” | 98,21 | 0,8 | 1 | 0,89 |
| “dua” | 98,21 | 1 | 0,75 | 0,86 |
| “oke” | 100 | 1 | 1 | 1 |
| “tidak” | 95,54 | 0,64 | 0,88 | 0,74 |
| “back” | 99,11 | 0,89 | 1 | 0,94 |
| “lima” | 98,21 | 1 | 0,75 | 0,86 |
| Rata-rata | 98,60 | 0,91 | 0,90 | 0,90 |

5. KESIMPULAN

Kesimpulan yang didapatkan dari penelitian ini adalah, pengujian yang telah dilakukan menyimpulkan bahwa walaupun penelitian ini sudah mendapatkan hasil yang cukup baik pada pengujian terhadap sistem, terdapat hal yang perlu diperhatikan dalam memilih perintah suara yang nantinya akan digunakan pada sistem. Pada penelitian ini, terdapat 14 perintah suara yaitu, “Satu”, “Dua”, “Tiga”, “Empat”, “Lima”, “Enam”, “Mati”, “Nyala”, “Oke”, “Tidak”, “Info”, “Next”, “Back”, dan “Keluar”.

Dari 14 perintah suara tersebut, terdapat 2 perintah suara yang cukup mirip dalam pelafalan katanya. Perintah suara tersebut yaitu, “Tiga” dan “Tidak” yang disimpulkan dari nilai yang ada pada *Classification Metrics*. Dengan demikian, peneliti memberi saran untuk penelitian serupa selanjutnya agar pemilihan perintah suara yang akan digunakan pada sistem dapat lebih diperhatikan, terutama dalam pelafalan katanya. Hal tersebut bertujuan untuk

menghindari bias antara kelas yang satu dengan kelas yang lainnya.

6. DAFTAR PUSTAKA

- Centers for Disease Control and Prevention, 2020. cdc.gov. [Online] [Diakses 28 Februari 2023].
- World Health Organization, 2022. Blindness and vision impairment. [Online] Tersedia di: <https://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment>. [Diakses 25 Februari 2023].
- World Health Organization, 2023. Disability. [Online] Tersedia di: [https://www.who.int/news-room/fact-sheets/detail/disability-and-health#:~:text=Key%20facts,earlier%20han%20those%20without%20disabilities](https://www.who.int/news-room/fact-sheets/detail/disability-and-health#:~:text=Key%20facts,earlier%20than%20those%20without%20disabilities). [Diakses 25 Februari 2023].
- Sukerkar, K. et al., 2018. Smart Wheelchair: A Literature Review. International Journal of Informatics and Communication Technology (IJ-ICT), hal. 63.
- Mahmood, A. & Kose, U., 2021. Speech recognition based on Convolutional neural networks and MFCC. Advances in Artificial Intelligence Research (AAIR), Volume 1, hal. 6-7.
- Passricha, V. & Aggarwal, R. K., 2018. Convolutional Neural Network for Raw Speech Recognition. In: From Natural to Artificial Intelligence: Algorithms and Applications. London, United Kingdom: IntechOpen, hal. 33.
- Ashar, A., Bhatti, M. S. & Mushtaq, U., 2020. Speaker Identification Using a Hybrid CNN-MFCC. 2020 International Conference on Emerging Trends in Smart Technologies (ICETST).