

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Robot telah menjadi bagian penting dari kehidupan manusia, robot sendiri sudah membantu dalam berbagai aspek seperti produksi industri (Riyadi & Nugroho, 2020), pembersihan rumah tangga (Utama et al., 2020), dan sebagai hiburan. Dalam hal komunikasi data Robot terbagi menjadi 2 golongan. Yang pertama, robot yang dikomunikasikan melalui kabel, yang di mana kelemahan utama bentuk ini adalah pergerakan robot yang dibatasi pada ukuran kabel. Yang kedua nirkabel, menggunakan teknologi seperti *Bluetooth* dan *Wi-Fi* memungkinkan robot bisa beroperasi tanpa batasan ukuran kabel. Komunikasi nirkabel ini memiliki berbagai bentuk, termasuk penggunaan tombol pada perangkat kendali jarak jauh, maupun menggunakan aplikasi pada smartphone menggunakan interaksi fisik. Suara manusia juga memiliki potensi untuk dijadikan sumber komunikasi pada robot, yang dimana bisa memberikan interaksi yang lebih fleksibel antara manusia dengan robot dikarenakan tidak memerlukan interaksi fisik seperti menekan tombol, Komunikasi seperti ini bisa membantu dalam media pembelajaran anak-anak (Aguilera et al., 2024), dan orang yang mempunyai keterbatasan fisik (Ntoa et al., 2023).

Manusia memiliki karakteristik yang bersifat unik, termasuk ciri fisik, genetik, maupun suara. Suara manusia memiliki perbedaan anatomi vokal, yang mempengaruhi frekuensi suara (SingWise, 2020.). Misalnya, mengutip sebuah jurnal dikatakan bahwa suara anak laki-laki memiliki frekuensi antara 210 hingga 270 Hz, sedangkan suara anak perempuan antara 200 hingga 290 Hz. Pada orang dewasa, pitch suara laki-laki berkisar antara 120 hingga 150 Hz, sementara perempuan antara 200 hingga 280 Hz. Di usia lanjut, pitch suara laki-laki berubah menjadi 100 hingga 140 Hz, dan perempuan 160 hingga 200 Hz (Ramba, 2020). Tantangannya adalah jika suara manusia akan dijadikan sumber komunikasi pada robot. Diperlukan sebuah sistem cerdas yang bisa mendeteksi berbagai karakteristik suara manusia tersebut.

Teknologi komunikasi dan informasi semakin berkembang, salah satunya banyaknya aplikasi berbasis *Artificial Intelligence* (AI). Menurut John Mccarthy pada buku (*Artificial Intelligence: 2022*) atau biasa dikenal sebagai bapak AI

mendefinisikan bahwa AI adalah suatu langkah untuk menciptakan komputer, robot, aplikasi, atau program yang bekerja secara cerdas, layaknya seperti manusia. AI bekerja dengan cara memproses informasi, belajar dari data, mengenali pola, dan membuat keputusan. AI memerlukan komponen penting dalam membangunnya, salah satunya adalah *Machine Learning* (ML). Untuk membangun sebuah AI, diperlukan data yang rapi dan terstruktur, data yang dilabeli, serta perancangan algoritma. ML, sebagai sub-bidang dari AI, berfokus pada menangani dan memprediksi data berskala besar dengan memanfaatkan algoritma pembelajaran yang diterapkan pada data-data tersebut. (Akmaluddin & Dewayanto, 2023). Untuk melatih data dalam jumlah besar dan kompleks, dibutuhkan jaringan saraf tiruan dengan banyak lapisan, yang mampu melakukan tugas-tugas seperti pengenalan Gambar dan pemrosesan suara dengan akurasi tinggi, yang dikenal sebagai *Deep Learning*.

Peran *Deep Learning* (DL) membantu dalam mengenali pola suara dan memahami konteks percakapan. Dengan menggunakan jaringan saraf tiruan, DL dapat menganalisis dan mensintesis suara manusia dengan akurasi tinggi (Raup et al., 2022). Salah satu model yang digunakan adalah *Deep Learning Convolutional Neural Network* (DL-CNN). CNN, sebagai jenis jaringan saraf tiruan cukup efektif dalam mengenali dan memproses data visual dan audio, dapat menganalisis spektrum frekuensi dari sinyal suara, dengan mengidentifikasi pola. Seperti pada Studi sebelumnya yang dilakukan oleh (Ramba, 2020) menunjukkan bahwa model CNN berhasil mencapai akurasi 100% pada ruangan senyap, 67,67% pada ruangan dengan intensitas noise 42dB, dan 51,67% pada ruangan bising, untuk 5 jenis perintah. Penelitian lain oleh (Karenina et al., 2023) menemukan bahwa model CNN mencapai akurasi 92% dalam mengklasifikasikan rentang usia dan gender berdasarkan suara, menunjukkan kemampuan membedakan suara anak-anak, remaja, dan dewasa.

Meninjau penelitian yang sudah ada, peneliti masih melihat beberapa keterbatasan penelitian, Salah satunya adalah tidak membahas pengenalan karakter suara per-individu. Mengetahui permasalahan tersebut, penulis bermaksud mengembangkan teknologi tersebut dengan menambahkan konsep *speaker verification*. Sehingga apabila di komunikasikan pada sebuah robot atau sebuah sistem, seseorang yang tidak memiliki hak sebagai navigtor tidak bisa menjadi

faktor atau variabel yang bisa mengganggu mobilitas robot tersebut. Seperti pada penelitian (Afif Ma'rif et al., 2022) yang bisa membuat sebuah sistem presensi kehadiran menggunakan verifikasi suara berbasis pembelajaran model klasifikasi *Learning Vector Quantization (LVQ)*. adapun penelitian oleh (Sidik Permana et al., 2018) mengimplementasikan metode *Dynamic Time Warping (DTW)* dalam pengenalan jenis suara pria dan wanita.

Maka dari itu, Penulis mengembangkan sebuah *Prototype* robot beroda yang dikomunikasikan secara nirkabel berbasis aplikasi *mobile* yang dikendalikan melalui suara, menggunakan teknologi model *Deep Learning Convolutional Neural Network* sebagai *Voice Command*, serta Model *Learning Vector Quantization* dan Metode perbandingan *Dynamic Time Warping* sebagai *Voice Verification*. Dengan memanfaatkan ekstraksi fitur suara *Mel Frequency Cepstral Coefficient*, sistem ini mampu mengenali perintah suara pengguna dan memverifikasi suara pengguna, sehingga bisa meningkatkan efisiensi pengendalian robot maupun segi keamanan.

1.2 Rumusan Masalah

Penerapan model pembelajaran *Convolutional Neural Network* sebagai pengenalan perintah suara pengguna dan model *Learning Vector Quantization* serta metode perbandingan *Dynamic Time Warping* sebagai pengenalan suara pengguna menimbulkan beberapa pertanyaan. Berikut adalah beberapa rumusan masalah yang mungkin relevan untuk penelitian ini:

- a. Bagaimana cara agar suara bisa diklasifikasikan oleh model dengan memanfaatkan ekstraksi fitur *Mel Frequency Cepstral Coefficient*?
- b. Bagaimana cara mengembangkan sistem yang bisa memverifikasi pengguna menggunakan *Learning Vector Quantization* dan *Dynamic Time Warping*?
- c. Bagaimana cara mengembangkan sistem yang bisa mengenali perintah suara pengguna menggunakan *Convolutional Neural Network*?

1.3 Tujuan

Penelitian ini memiliki tujuan yang akan dicapai, yaitu:

1. Mengekstraksi fitur suara menggunakan *Mel Frequency Cepstral Coefficient* sebagai bahan untuk klasifikasi dan perbandingan
2. Membandingkan performa model *Learning Vector Quantization* dan metode perbandingan *Dynamic Time Warping* dalam memverifikasi suara pengguna.
3. Mengevaluasi performa model klasifikasi *Convolutional Neural Network* dalam mengenali perintah suara pengguna.

1.4 Manfaat

Penelitian ini bermanfaat sebagai landasan dalam mengembangkan sistem yang bisa mengenali perintah suara dan memverifikasi suara pengguna menggunakan model pembelajaran dan metode perbandingan.

1.5 Batasan Masalah

Dalam penelitian ini, beberapa batasan perlu diberlakukan untuk memfokuskan lingkup penelitian dan memudahkan pelaksanaannya. Berikut adalah beberapa batasan yang mungkin relevan untuk penelitian ini:

- a. Penelitian ini hanya menggunakan metode *Mel Frequency Cepstral Coefficient* dalam mengekstraksi ciri suara.
- b. Penelitian ini hanya berfokus dalam pengembangan sistem pengenalan perintah dan verifikasi suara pengguna tanpa membandingkan dengan metode apapun.
- c. Dataset yang digunakan dalam penelitian ini hanya bersifat primer yang hanya dalam ruang lingkup penelitian.