

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa dekade terakhir kecerdasan buatan atau yang disebut dengan *Artificial Intelligence* (AI) berkembang pesat dan memiliki dampak besar di berbagai bidang. Dalam konteks ini, AI melibatkan penggunaan algoritma dan model matematika untuk memungkinkan komputer serta sistem lainnya belajar dari data, lalu mengenali pola, dan akhirnya dapat membuat keputusan yang cerdas. Konsep - konsep penting dalam AI mencakup pembelajaran mesin (*Machine Learning*) dan *Deep Learning*.((Russell & Norvig, 2010);(Goodfellow et al., 2016)).

Pada masa kini teknologi interaksi antar manusia dan komputer pun telah mengalami kemajuan yang signifikan.(Bingol & Aydogmus, 2020). Salah satunya adalah robot interaktif berbasis pengenalan suara, yang telah banyak dikembangkan untuk berbagai kebutuhan industri dan kehidupan sehari - hari. (Moriya et al., 2019)

Penelitian kali ini berfokus dengan pengenalan suara menggunakan metode *Deep Learning* untuk mengontrol pergerakan robot. Suara merupakan alat komunikasi yang paling sering digunakan manusia karena struktur alaminya. Oleh karena itu, pengenalan suara ini telah menjadi topik penelitian selama bertahun - tahun. (Moriya et al., 2019). Sudah banyak sekali contoh aplikasi yang menggunakan teknologi *Machine Learning* dan *Deep Learning* untuk mengenali dan memproses perintah suara, seperti siri pada iPhone, *google assistant* pada perangkat android dan alexa pada perangkat amazon echo yang telah berhasil diimplementasikan.

Beberapa penelitian sebelumnya telah membuktikan model machine learning dalam pengenalan suara, diantaranya yaitu Rudregowda et al. (2024) tentang *Audio-visual speech recognition based on a deep convolutional neural network*. Penelitian ini berfokus pada pengembangan sistem pengenalan suara audiovisual berbasis kombinasi antara model CNN dan LSTM dengan penggabungan fitur suara dan visual. Jumlah data yang digunakan sebanyak 80 videos yang terdiri dari 8 kata dan 16 orang. Terbagi kedalam 75 persen data training dan 25 persen data validasi dan hasil akhir akurasi yang didapatkan adalah 91,53 persen. Selanjutnya penelitian oleh Tu et al. (2019) tentang *An iterative mask estimation approach to deep learning based multi-channel speech recognition*. Penelitian ini berfokus pada pendekatan *iterative mask estimation* untuk meningkatkan suara multi saluran berbasis beamforming dengan kombinasi *deep learning* model DNN dan LSTM. Sumber data yang digunakan berasal dari chime-4 challenge yang terdiri dari data percakapan dengan 6 mikrofon dalam berbagai noise, terdapat data suara asli dan data suara simulasi. Dengan hasil yang didapatkan yaitu penurunan error sebesar 20 - 30 persen dibandingkan metode gaussian mixture model (GMM) tradisional. Selanjutnya terdapat penelitian oleh Fatehi et al. (2025) tentang *An overview of high-resource automatic speech recognition methods and their empirical evaluation in low-resource environments*. Penelitian ini berfokus pada evaluasi berbagai teknik *automatic speech recognition* (ASR). Sumber dataset yang digunakan berasal dari librispeech dan uaspeech. Jurnal ini mengevaluasi model end to end ASR seperti DNN, LSTM, CNN, transformers based models, self supervised learning dan transfer learning. Dengan jumlah data yang digunakan 541 suara didapatkan hasil error dari model DNN yaitu 19,25 persen, kombinasi model antara CNN dan LSTM didapatkan hasil error sebesar 27,13 persen, dan yang terakhir untuk model transformers based model, SSL dan transfer learning sebesar 20 persen.

Adapun penelitian mengenai kontrol robot diantaranya penelitian Dokuz & Tufekci (2021) tentang *Mini-batch sample selection strategies for deep learning based speech recognition*. Penelitian ini menggunakan dataset besar yang bersumber dari dataset VCTK speech corpus. Dengan menggunakan model LSTM dan optimasi yang dilakukan dengan mini-batch gradient descent. Selanjutnya penelitian Orooso et al. (2025) tentang *Transforming English language learning: Advanced speech recognition with MLP-LSTM for personalized education*. Penelitian ini menggunakan gabungan model antara multilayer perceptron (MLP) dan LSTM. Dengan hasil akhir akurasi mencapai 98,25 persen dan error sebesar 0,875. Selanjutnya terdapat

penelitian Bingol & Aydogmus (2020) tentang *Performing predefined tasks using the human robot interaction on speech recognition for an industrial robot*. Penelitian ini menggunakan sumber data dari Kuka KR Agilus KR6 R900. dengan model yang digunakan berupa kombinasi antara DNN dengan fuzzy logic controller. Dan menurut hasil penelitian model DNN lebih akurat dibandingkan model klasik dalam menangani variasi bahasa atau dialek dengan nilai akurasi mencapai 90,37 persen.

Berdasarkan beberapa penelitian tersebut, dapat disimpulkan bahwa pemilihan arsitektur model sangat mempengaruhi kinerja sistem pengenalan suara. Berbagai arsitektur seperti Deep Neural Network (DNN), Recurrent Neural Network (RNN), serta kombinasi DNN-LSTM telah banyak digunakan. Namun pendekatan tersebut masih memiliki keterbatasan, terutama dalam hal efisiensi ekstraksi fitur spasial dan ketahanan terhadap noise lingkungan. Sedangkan model gabungan Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM) lebih baik karena memiliki keunggulan dalam mengekstraksi pola spasial sekaligus mampu memodelkan urutan temporal. CNN bertugas mengekstrak fitur penting dari data suara yang direpresentasikan dalam bentuk spektrogram sedangkan LSTM memproses urutan fitur secara temporal sehingga meningkatkan akurasi sistem pengenalan suara.

Selain itu, penelitian ini menggunakan data suara yang direkam secara mandiri agar sistem dapat dilatih menggunakan data suara yang lebih relevan dengan kebutuhan sistem dalam kontrol robot.

Dari observasi yang telah dilakukan baik secara langsung ataupun tidak langsung, didapat aspek penting dalam sistem kendali robot berbasis arduino mega. adalah penggunaan *software* Python dalam pengambilan data, pengolahan data, dan pengujian hasil menggunakan *Deep Learning* dengan jenis pemodelan *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) . Selain hal tersebut untuk merancang input data masuk kedalam sistem kendali robot digunakan Arduino Mega. Dari kedua hal tersebut penulis mengambil Langkah untuk mengintegrasikan dalam komunikasi serial antar dua *software* agar mendapatkan hasil yang diharapkan. Dengan pendekatan ini, diharapkan sistem yang dibangun tidak hanya memiliki akurasi tinggi, tetapi juga mampu bekerja secara responsif dan optimal dalam skenario pengendalian robot berbasis perintah suara.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas maka permasalahan yang perlu dirumuskan adalah sebagai berikut:

1. Bagaimana cara membuat program pengolahan database rekaman suara menjadi perintah pada Robot Sosial?
2. Bagaimana cara membuat program *deep learning* model *Convolutional Neural Network* (CNN) dan Long Short-Term Memory (LSTM)
3. Bagaimana model *Convolutional Neural Network* (CNN) dan Long Short-Term Memory (LSTM) dapat diimplementasikan dalam sistem kendali Robot Sosial berbasis pengenalan suara (speech recognition)?
4. Bagaimana perbandingan perintah pada kontrol robot yang menggunakan model *Convolutional Neural Network* (CNN) dengan model gabungan *Convolutional Neural Network* (CNN) dan Long Short-Term Memory (LSTM)

1.3 Batasan Masalah

Agar tidak menyimpang jauh dari permasalahan, maka penelitian ini mempunyai batasan masalah sebagai berikut:

1. Penelitian ini menggunakan penggabungan konsep *Deep Learning* dengan metode *Convolutional Neural Network* (CNN) . dan *Long Short-Term Memory* (LSTM).
2. Penelitian ini bekerja berdasarkan dataset yang telah dibuat, diluar itu robot akan mendeteksi hal lain yang inputnya mendekati.
3. Data set yang dibuat hanya berfokus pada gerakan tangan robot dan roda. Yaitu berupa angka satu sampai sepuluh, gerakan tangan seperti dadah dan salam dan gerakan roda maju, mundur, kanan, dan kiri.
4. Penelitian ini menggunakan komunikasi data serial antara program *Deep Learning* python dan program mikrokontroler arduino mega.
5. Penelitian ini menggunakan dataset yang terdiri dari 4.200 data dengan 21 kelas kata yang dikumpulkan dari 4 orang. Jumlah data dan partisipan yang

terbatas dapat mempengaruhi generalisasi model dalam mengenali variasi pengucapan yang lebih luas.

1.4 Tujuan Penelitian

Adapun tujuan dilakukannya penelitian ini yaitu:

1. Membuat program pengolahan database rekaman suara menjadi perintah pada Robot Sosial.
2. Membuat program deep learning model *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM)
3. Dapat mengimplementasikan model *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) dalam sistem kendali Robot Sosial berbasis pengenalan suara (*speech recognition*).
4. Membandingkan perintah pada kontrol robot yang menggunakan model *Convolutional Neural Network* (CNN) dengan model gabungan *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM)

1.5 Manfaat Penelitian

Penelitian ini akan memberikan manfaat yaitu diharapkan Rancang Bangun sistem dapat meningkatkan efisiensi dan akurasi pengenalan suara, terutama dalam menghadapi *noise* dan memungkinkan sistem untuk beroperasi secara real time.

1.6 Sistematika Penulisan

Penulisan Skripsi ini disusun ke dalam lima buah bab dengan sistematika sebagai berikut.

BAB I : PENDAHULUAN

Memberikan pemahaman mengapa penelitian ini dilakukan, dengan menguraikan latar belakang yang memicu penelitian, rumusan masalah yang menjadi fokus, batasan-batasan yang mengapit penelitian, tujuan yang ingin dicapai, manfaat

yang diharapkan dari hasil penelitian, dan rangkuman singkat tentang struktur keseluruhan penulisan.

BAB II : TINJAUAN PUSTAKA

Merinci pengetahuan terkait dengan penelitian ini. Ini mencakup ulasan literatur yang relevan dan konsep-konsep yang menjadi dasar bagi penelitian ini, serta menjelaskan prinsip dasar dari Robot Sosial dalam *Speech Recognition*.

BAB III : METODOLOGI PENELITIAN

Menggambarkan secara detail tentang bagaimana penelitian ini dilakukan. Ini mencakup lokasi dan waktu penelitian, daftar peralatan dan bahan yang digunakan, serta langkah-langkah prosedur penelitian.

BAB IV : HASIL DAN PEMBAHASAN

Menjelaskan hasil olah data, pengujian dan pembahasan metode *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) *Deep Learning* secara *real time* untuk mengontrol gerak Robot Sosial.

BAB V : PENUTUP

Merangkum hasil penelitian dan mengidentifikasi potensi kekurangan dalam penelitian ini. Rekomendasi akan diberikan untuk pengembangan penelitian lebih lanjut di masa depan.