

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Manusia adalah makhluk sosial yang saling berkomunikasi dengan menggunakan bahasa. Bentuk komunikasi ini biasanya dilakukan secara verbal/lisan, yang artinya komunikasi dengan menggunakan kata-kata. Akan tetapi *The World Health Organization* (WHO) memberikan pernyataan bahwa lebih dari 5% populasi dunia atau 466 juta orang mengalami gangguan pendengaran. Diperkirakan pada tahun 2050 lebih dari 900 juta orang atau satu dari setiap sepuluh orang akan mengalami gangguan pendengaran (Bridget Shield, 2019). Dengan informasi tersebut dapat disimpulkan bahwa akan terjadi peningkatan populasi yang mengalami gangguan pendengaran. Komunitas yang mengalami gangguan pendengaran disebut sebagai tunarungu. Komunitas ini memiliki cara berkomunikasi sendiri tanpa menggunakan bahasa lisan yaitu dengan bahasa isyarat. Sedangkan pada umumnya bahasa isyarat sulit dipahami oleh masyarakat dan membuat komunitas tersebut merasa terasingkan di lingkungannya. Bahasa isyarat itu sendiri merupakan bahasa yang digunakan untuk berkomunikasi dengan menggunakan gerak bibir dan bahasa tubuh, termasuk ekspresi wajah dan pandangan mata. Mengatasi keterbatasan dalam hal berkomunikasi, teknologi *computer vision* bisa menjadi solusi. Teknologi ini mampu menjembatani komunikasi antar manusia melalui mesin. *Computer vision* merupakan salah satu bidang yang memanfaatkan kecerdasan buatan untuk melatih komputer dalam menafsirkan dan memahami dunia visual. Teknologi ini memanfaatkan data dari kamera atau sensor yang kemudian diolah menggunakan model *deep learning*. Sehingga mesin mampu secara akurat mengidentifikasi dan mengklasifikasikan objek kemudian bereaksi terhadap apa yang mesin “lihat”.

Penelitian ini dilakukan dengan pendekatan berbasis *computer vision*. Pada pendekatan tersebut digunakan data yang berasal dari kamera untuk menangkap gerakan bahasa isyarat. Selanjutnya melakukan praproses seperti mengumpulkan data mentah dengan menggunakan kamera, deteksi objek, segmentasi, konversi warna, *reshaping*, *thresholding* dan operasi morfologi. Setelah data berhasil

didapatkan, kemudian melakukan normalisasi untuk pembuatan model dataset. Normalisasi ini memastikan bahwa data yang berhasil dimodelkan bisa dengan efektif dan efisien digunakan oleh sistem cerdas. Semua praproses tersebut dilakukan untuk memisahkan objek yang sangat penting dari *noise*. Sampai pada akhirnya sistem komputer dengan teknologi kecerdasan buatan mampu memanfaatkan model dari dataset tersebut dengan maksimal.

Untuk membuat kecerdasan buatan. Salah satu pendekatan yang bisa digunakan adalah menggunakan Jaringan Syaraf Tiruan yang terinspirasi dari jaringan syaraf pada manusia. Konsep tersebut kemudian dikembangkan lebih lanjut dalam *Deep Learning*. *Deep Learning* telah menjadi salah satu topik hangat dalam dunia *Machine Learning* karena kapabilitasnya yang signifikan dalam memodelkan berbagai data kompleks seperti citra dan suara. Metode *Deep Learning* yang saat ini memiliki hasil paling signifikan dalam pengenalan citra adalah *Convolutional Neural Network* (CNN). Hal tersebut dikarenakan CNN berusaha meniru sistem pengenalan citra pada visual *cortex* manusia sehingga memiliki kemampuan mengolah informasi citra.

Sebernarnya konsep penelitian dan pengembangan sistem komputer vision untuk penerjemah bahasa isyarat sudah ada beberapa yang telah dilakukan antara lain : Sistem Pengenalan Bahasa Isyarat Indonesia Dengan Menggunakan Metode Fuzzy K-Nearest Neighbor (Gafar, 2017), Pengenalan Angka Sistem Isyarat Bahasa Indonesia Dengan Menggunakan Metode Convolutional Neural Network (Bakti., 2019), Model Penerjemah Bahasa Isyarat Indonesia (Bisindo) Menggunakan Convolutional Neural Network (Fadillah, 2020). Dari beberapa penelitian tersebut masih ditemukan keterbatasan dan kekurangan yang bisa diperbaiki seperti kemampuan sistem yang hanya bisa menerjemahkan bahasa isyarat sederhana, input data yang hanya berupa foto dari model peraga yang diambil secara manual atau tidak *realtime* berasal dari video kamera/webcam, tidak melakukan penghilangan *area* wajah dan leher didalam proses pengolahan citra untuk deteksi objek tangan, dan hasil akurasi yang rendah dalam menerjemahkan bahasa isyarat.

Sebagai perbaikan dan pengembangan sistem lebih lanjut, penelitian ini akan menggunakan Bahasa Isyarat Indonesia (BISINDO) sebagai jenis bahasa isyarat

yang dipakai didalam dataset. Selanjutnya peneliti akan menggunakan framework *MediaPipe* untuk mendeteksi objek tangan manusia dan mendapatkan landmark, hal ini bertujuan untuk meningkatkan kemampuan dan akurasi dalam mendeteksi bahasa isyarat. Kemudian pengambilan citra akan dilakukan dengan menggunakan satu kamera *webcam (selfie)* yang nantinya data citra akan diterjemahkan secara *realtime* ketika objek tangan terdeteksi oleh sistem. Pada proses pengolahan data citra dari kamera akan dilakukan proses pengaturan kecerahan gambar, membalik (*flip*) gambar, konversi warna, *reshaping* citra dan *cropping* data citra. Algoritma yang akan digunakan pada penelitian ini adalah CNN yaitu model arsitektur *Convolutional Neural Network*, algoritma ini dipilih berdasarkan tingkat kemampuan dan akurasi untuk menerjemahkan abjad dan nomor, hal ini dibuktikan dari penelitian berjudul “Automatic recognition of Colombian car license plates using convolutional neural networks and Chars74k database” dengan objek input berupa gambar karakter yang diambil dari pelat mobil Kolombia yang memiliki akurasi di atas 98% (D E Arroyo-Pérez., 2020).

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang, terdapat beberapa permasalahan yang bisa dirumuskan yaitu sebagai berikut :

1. Bagaimana implementasi *framework MediaPipe* dalam mendeteksi objek tangan pada data citra dari kamera video?
2. Bagaimana proses pengolahan dan normalisasi data citra digital?
3. Bagaimana implementasi metode *Deep Learning Model Convolutional Neural Network* dalam menerjemahkan gerakan bahasa isyarat?

Batasan masalah yang ada dalam penelitian ini yaitu sebagai berikut :

1. Objek penelitian adalah daerah anggota badan manusia dari pergelangan sampai ujung jari tangan.
2. Pencahayaan ruangan harus terang. Objek penelitian berupa tangan harus masuk dalam frame kamera.
3. Kamera yang digunakan minimal memiliki resolusi sebesar 480 x 360 *pixels*.

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut :

1. Untuk mengetahui implementasi *framework MediaPipe* dalam mendeteksi objek tangan.
2. Untuk mengetahui teknik pengolahan dan normalisasi data citra yang tepat.
3. Untuk mengetahui hasil metode *Deep Learning Model Convolutional Neural Network* dalam memahami dan menerjemahkan gerakan bahasa isyarat.

1.4 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut :

1. Meningkatkan efisiensi komputasi pada sistem sehingga memiliki tingkat fungsional dan akurasi yang baik didalam menerjemahkan bahasa isyarat.
2. Memberikan data yang sesuai dengan lingkungan arsitektur dari machine learning dalam proses data training dan prediksi gerakan bahasa isyarat.
3. Mampu mengembangkan software untuk memudahkan masyarakat luas dalam memahami bahasa isyarat dengan bantuan machine learning.