

## 1. Pendahuluan

### Latar Belakang

Komunikasi adalah sesuatu yang lumrah dilakukan antar sesama manusia. Komunikasi adalah bentuk penyampaian informasi dari seseorang atau beberapa orang, kelompok, organisasi, atau masyarakat. Fungsi dari komunikasi itu sendiri adalah untuk menyampaikan informasi, menyampaikan pendapat, bentuk interaksi terhadap orang lain, menjaga jalinan hubungan antar orang agar tetap baik, dan untuk mengekspresikan diri. Komunikasi dikatakan berhasil jika kedua pihak atau lebih yang melakukan komunikasi bisa menerima dengan benar atas komunikasi yang disampaikan. Ada beberapa jenis komunikasi yang biasa digunakan menurut cara penyampaian, yaitu komunikasi secara verbal, non verbal, dan tertulis.

Orang-orang difabel yang tidak bisa berkomunikasi secara verbal biasanya berkomunikasi secara non-verbal. Komunikasi non verbal yang biasa digunakan orang-orang difabel adalah bahasa isyarat. Namun, banyak sekali orang yang tidak bisa memahami bahasa isyarat. Hal itu membuat terjadinya keterbatasan dalam berkomunikasi. Oleh karena itu, diperlukan sebuah sistem untuk mengatasi keterbatasan dalam berkomunikasi tersebut. Sistem yang perlu dibangun dari permasalahan itu adalah sistem yang bisa menerjemahkan bahasa isyarat. Hal itu adalah ide yang mendasari teretusanya Tugas Akhir ini, penulis termotivasi untuk membuat komunikasi menjadi lebih mudah untuk siapapun. Sistem yang dibangun adalah sistem yang bisa mengenali bahasa isyarat huruf alfabet yang diperagakan menggunakan gestur tangan secara statis.

Dalam penelitian yang dilakukan sebelumnya tentang *static hand gesture* oleh Atoany Nazareth Fierro Radilla dan Karina Ruby Perez Daniel yang berjudul “*Siamese Convolutional Neural Network for ASL Alphabet Recognition*”[1]. Dalam penelitian tersebut menggunakan metode *CNN* dengan arsitektur *Siamese* untuk membedakan kemiripan dari gestur di setiap kelasnya dengan mengalokasikan 30 *epoch* untuk melakukan proses *training*. Dalam penelitian itu dihasilkan hasil evaluasi performa sistem yang cukup bagus yaitu dengan tingkat akurasi sebesar 96%. Kesalahan hasil prediksi sering terjadi pada kelas M, kelas N, kelas R dan kelas U. Banyak prediksi yang harusnya memprediksi kelas N, tetapi memprediksi kelas M, begitu juga sebaliknya. Banyak prediksi yang harusnya memprediksi kelas M, tetapi memprediksi kelas N. Hal yang sama terjadi pada kelas R dan U. Penelitian serupa juga dilakukan oleh Shweta Upadhyay, Dr. R.K. Sharma, Dr. Prashant Singh Rana pada tahun 2020 dari Thapar Institute of Engineering and Technology, Patiala yang berjudul “*Sign Language Recognition with Visual Attention*”[2]. Pada penelitian tersebut, penulis melakukan pengenalan bahasa isyarat menggunakan *dataset ASL Alphabet* yang terdiri dari 29 kelas, yaitu kelas alfabet A-Z ditambah dengan kelas ‘hapus’, ‘spasi’, dan ‘no gesture’. Untuk keperluan pengujian, penulis melakukan pembagian dataset ke dalam *training set* dan *testing set* dengan perbandingan 9:1. Pendekatan yang dilakukan adalah dengan menggunakan metode *Faster-RCNN*. Dari pendekatan tersebut hasil evaluasi mencapai tingkat akurasi sebesar 94%.

Lalu ada penelitian lain yang sudah dilakukan oleh Akm Ashiquzzaman, Hyunmin Lee, Kwangki Kim, Hye-Young Kim, Jaehyung Park, dan Jinsul Kim, dengan judul “*Compact Spatial Pyramid Pooling Deep Convolutional Neural Network Based Hand Gestures Decoder*”[3]. Dalam penelitian tersebut, penulis menggunakan metode *CNN* namun dengan menambahkan *Spatial Pyramid Pooling* dalam arsitektur *CNN*-nya untuk mengenali bahasa isyarat menggunakan *dataset ASL Alphabet*. Pada penelitian tersebut didapatkan hasil evaluasi secara keseluruhan dengan tingkat akurasi, *recall*, dan *f1-score* mencapai 94%, presisi 95%, dengan waktu proses prediksi untuk 1 data adalah 0.013 s. Misklasifikasi paling banyak terjadi pada kelas M dengan banyak memprediksi kelas N. misklasifikasi terjadi karena gestur M dan N terlihat sangat mirip. Penelitian serupa juga pernah dilakukan oleh P. V. S. M. S. Kartik, Konjeti B. V. N. S. Sumanth, V. N. V. Sri Ram, P. Prakash pada tahun 2020 dengan judul “*Sign Language Recognition with Visual Attention*”[4]. Pada penelitian tersebut, dataset yang digunakan adalah *ASL Alphabet* yang terdiri dari 29 kelas antara lain 26 kelas alfabet huruf A-Z ditambah dengan kelas ‘hapus’, ‘spasi’, dan ‘no gesture’. Untuk keperluan training dan testing penulis membagi dataset tersebut kedalam training set dan testing set dengan perbandingan 9:1. Dalam penelitian tersebut penulis menggunakan pendekatan *deep learning* dengan *convolutional neural networks* dengan arsitektur ResNet-50. Hasil evaluasi yang diperoleh dari penelitian tersebut adalah tingkat akurasi yang mencapai 98,67%.

Karena penulis berpandangan bahwa untuk terjadinya komunikasi yang efektif maka diperlukan suatu sistem yang dalam evaluasinya akurasi yang hampir mendekati 100%, sehingga dalam sisi akurasi penulis berpandangan masih ada peluang untuk meningkatkannya agar menjadi seakurat mungkin. Penulis mengusulkan sebuah metode yang masih baru yaitu *CNN* dengan arsitektur *Efficient-Net B4* yang dirumuskan oleh Mingxing Tan dan Quoc V. Le pada tahun 2019[5]. Pendekatan tersebut dipilih karena metode tersebut dalam penelitian Mingxiang Tan dan Quoc V. Le[5], *Efficient-Net B4* mendapatkan hasil evaluasi yang lebih baik dari sisi akurasi dibandingkan dengan *ResNet-50* bahkan *ResNet-152*. Pada penelitian ini sistem dibangun untuk mengenali gestur tangan untuk mengenali dan menerjemahkan bahasa isyarat. Pada penelitian ini penulis membuat 4 skenario untuk mengenali gestur tangan sekaligus untuk mengetahui pengaruh perbedaan *input size* pada model dan penggunaan *learning rate* yang dinamis menggunakan *Cyclical Learning rate* atau *CLR* terhadap hasil evaluasi performa sistem dan efisiensi waktu. *Cyclical Learning Rate* pertama kali dipublikasikan oleh Leslie N. Smith pada tahun 2017[6]. Dimana pada penelitian tersebut, Leslie N. Smith melakukan perbandingan dengan melakukan training menggunakan model dengan macam-macam arsitektur. Setiap arsitekturnya dilakukan pengujian dengan

menggunakan *CLR* dan tidak menggunakan *CLR*. Hasilnya pada semua model yang digunakan, pengujian yang menggunakan *CLR* mendapatkan hasil evaluasi yang lebih baik dan mengalami konvergen pada nilai *epoch* yang lebih kecil dibandingkan dengan pengujian tanpa menggunakan *CLR*.

### **Topik dan Batasannya**

Berdasarkan latar belakang yang dijelaskan, permasalahan yang dapat diangkat dalam tugas akhir ini, yaitu membangun sebuah sistem yang dapat mengenali bahasa isyarat dengan gestur tangan statis menggunakan *CNN* dengan arsitektur *Efficient Net B4*. Performa sistem dalam mengenali bahasa isyarat dengan gestur tangan statis. Pengaruh variasi *input size* dan *CLR* terhadap performa sistem yang dibangun.

Batasan dari sistem yang dibangun dalam tugas akhir ini adalah sistem hanya mengenali huruf alphabet ditambah dengan isyarat untuk spasi, hapus, dan no gesture yang ada pada *dataset ASL Alphabet*(29 kelas). Penelitian ini tidak menerapkan *hand detection*, sehingga data inputan yang dimasukan sudah berupa citra tangan yang memperagakan bahasa isyarat dengan format citra .jpg. *Dataset* diperagakan dengan tangan kanan dari satu orang saja. Latar dalam dataset yang digunakan tidak terlalu beragam.

### **Tujuan**

Tujuan yang ingin diraih dari tugas akhir ini adalah untuk membangun sebuah sistem yang bisa mengenali dan menerjemahkan bahasa isyarat dengan gestur tangan statis dengan menggunakan pendekatan *CNN* dengan arsitektur *Efficient-Net B4* dan menguji pengaruh variasi *input size* dan *CLR* terhadap performa dan efisiensi sistem.

### **Organisasi Tulisan**

Pada bagian pertama menjelaskan pendahuluan, pada bagian kedua menjelaskan studi terkait, pada bagian ketiga menjelaskan sistem yang dibangun, pada bagian keempat menjelaskan evaluasi dari sistem yang dibangun dan pada bagian kelima menjelaskan tentang kesimpulan dan saran.