

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang akan digunakan dalam klasifikasi American Sign Language menggunakan arsitektur EfficientNetV2B0 dengan *hyperparameter tuning*.

1.1 Latar Belakang

Manusia merupakan makhluk sosial yang saling berinteraksi dan berkomunikasi satu sama lain. Namun, bagi mereka yang mengalami gangguan pendengaran, komunikasi menjadi tantangan besar dalam bersosialisasi. Menurut *World Health Organization* (2024), sekitar 5% dari populasi dunia, atau sekitar 430 juta orang, berjuang melawan gangguan pendengaran, dan angka ini diperkirakan akan terus meningkat seiring waktu. Individu dengan gangguan pendengaran tersebut sebagian besar berkomunikasi dengan mengandalkan bahasa isyarat, yang diekspresikan melalui gerakan tubuh, terutama tangan.

Setiap negara memiliki bahasa isyarat dengan ciri khasnya yang berbeda-beda tergantung pada wilayah dan budaya masing-masing (National Geographic Society, 2024). Menurut *World Population Review* (2024), Bahasa Isyarat Amerika atau yang biasa disebut *American Sign Language* (ASL) merupakan bahasa isyarat yang paling umum digunakan, dengan lebih dari 400 ribu pengguna. ASL merupakan bahasa isyarat yang banyak digunakan tuna rungu di Amerika Utara sebagai bahasa utama dalam berkomunikasi (Alviando dkk., 2020).

Bahasa isyarat memiliki peran penting dalam mempromosikan inklusivitas, memecahkan hambatan komunikasi, serta memberdayakan individu dengan gangguan pendengaran agar dapat berpartisipasi penuh dalam kehidupan sosial dan profesional mereka (Abdillah & Ilyas, 2024). Akan tetapi, sebagian besar masyarakat umum tidak memahami bahasa isyarat, sehingga seringkali mempersulit komunikasi dengan penyandang gangguan pendengaran (Myagila & Kilavo, 2022). Fenomena ini diperburuk oleh fakta bahwa ada beragam bahasa isyarat di dunia, bahkan terdapat lebih dari 150 bahasa isyarat yang masih hidup (Eberhard dkk., 2024).

Ketidaktahuan sebagian besar masyarakat tentang bahasa isyarat mengakibatkan penderita gangguan pendengaran mengalami kesulitan dalam berkomunikasi dengan masyarakat umum. Oleh karena itu, pengenalan bahasa isyarat kepada masyarakat luas perlu dilakukan agar dapat menjembatani kesenjangan komunikasi antara penderita gangguan pendengaran dengan masyarakat umum (Ridwan, 2017). Pengenalan bahasa isyarat diharapkan dapat meningkatkan kesadaran masyarakat serta membantu penderita gangguan pendengaran dalam berinteraksi dengan orang lain (Tao dkk., 2024).

Pengembangan model klasifikasi pembelajaran mesin telah menunjukkan potensi besar untuk mendukung pengenalan bahasa isyarat. Seiring dengan kemajuan teknologi, model tersebut dapat mengenali gerakan bahasa isyarat dan menerjemahkannya ke dalam bentuk yang dapat dipahami oleh orang yang tidak fasih dalam bahasa isyarat (Svendsen & Kadry, 2023). Selain itu, penggunaan pembelajaran mendalam (*deep learning*) dalam klasifikasi bahasa isyarat menawarkan solusi yang efektif, karena memiliki kemampuan untuk mengekstraksi fitur-fitur penting dalam data gerakan yang rumit dan kompleks, yang merupakan tantangan dalam pengenalan bahasa isyarat (Tao dkk., 2024).

Berbagai penelitian terkait klasifikasi ASL dengan bantuan *computer vision* sudah pernah dilakukan sebelumnya. Buttar dkk. (2023) mengembangkan sistem pengenalan ASL menggunakan metode *deep learning*, yakni LSTM dan YOLOv6. Dengan memanfaatkan *MediaPipe* holistic landmarks, model LSTM mencapai akurasi 92%, sedangkan YOLOv6 mencapai akurasi lebih tinggi, yaitu 96%, untuk pengenalan bahasa isyarat statis. Model YOLOv6 terdiri dari sekitar 37 juta parameter. Selain itu, Patel dkk. (2021) mengembangkan model pengenalan Bahasa Isyarat menggunakan metode *Convolutional Neural Networks* (CNN) dengan model *pretrained* VGG16 untuk *transfer learning*. Dataset yang digunakan terdiri lebih dari 87.000 gambar dengan 29 kelas. Hasil dari penelitian menunjukkan bahwa akurasi model meningkat seiring jumlah gambar pelatihan, dengan akurasi tertinggi sebesar 95,33% ketika dilatih menggunakan 2100 gambar per kelas. Model VGG16 memiliki sekitar 138 juta parameter, menjadikannya model yang cukup kompleks dalam hal ukuran.

Arsitektur-arsitektur di atas memiliki performa yang baik, namun jumlah parameter yang dihasilkan model cukup besar. Semakin besar parameter yang dihasilkan maka ukuran model yang dihasilkan juga semakin besar dan proses komputasi yang dibutuhkan juga semakin lama (Julianto dkk., 2022). Saat ini terdapat arsitektur CNN yang memiliki jumlah

parameter yang kecil dengan performa yang baik yaitu arsitektur EfficientNet. Arsitektur ini terdiri dari 8 model, yaitu B0 hingga B7 yang menandakan jumlah parameter yang semakin banyak dari 5.3 juta hingga 66 juta parameter dan akurasi yang lebih tinggi (Tan & Le, 2019a). EfficientNet telah menunjukkan akurasi yang lebih tinggi dengan jumlah parameter yang lebih sedikit dan lebih cepat dibandingkan beberapa *pre-trained* model lainnya (Sasongko & Amrullah, 2023). Selain itu terdapat pula arsitektur EfficientNetV2, yang merupakan keluarga baru dari *deep learning* yang lebih unggul dalam pelatihan serta efisiensi parameter (Citra dkk., 2023). Penelitian menggunakan EfficientNet pernah dilakukan oleh Putri dkk. (2024) untuk klasifikasi bahasa isyarat ASL. Penelitian tersebut menggunakan dataset ASL yang terdiri dari 14.740 gambar. Hasil dari penelitian menunjukkan bahwa model berbasis EfficientNet mencapai akurasi pelatihan dan validasi sebesar 98,40%, dengan keunggulan berupa efisiensi ukuran model dan waktu inferensi. Model EfficientNet-B0, yang digunakan dalam penelitian ini, terdiri dari sekitar 5,3 juta parameter, sehingga lebih ringan dibandingkan model *pretrained* seperti VGG16.

Model CNN mampu memberikan performa yang cukup baik, tetapi masih dapat ditingkatkan untuk mencapai hasil yang lebih optimal. Salah satu cara untuk meningkatkan performanya adalah dengan melakukan *hyperparameter tuning*. *Hyperparameter tuning* memiliki peran penting dalam mengoptimalkan dan meningkatkan kinerja CNN. Beberapa *hyperparameter*, seperti *learning rate*, *dropout rate*, *batch size*, dan *dense layer*, sering kali disesuaikan untuk meningkatkan akurasi dan menghindari *overfitting* (Goodfellow dkk., 2016). Selain *hyperparameter tuning*, augmentasi data juga dapat meningkatkan performa model dalam tugas klasifikasi dengan memperbesar keragaman data pelatihan tanpa menambah jumlah sampel baru. Teknik augmentasi seperti rotasi, *flipping*, *zooming*, dan *shifting* dapat membantu model menjadi lebih invarian terhadap perubahan kecil dalam input, sehingga meningkatkan generalisasi terhadap data baru (Shorten & Khoshgoftaar, 2019).

Penelitian ini berfokus pada klasifikasi American Sign Language (ASL) dengan menggunakan arsitektur CNN EfficientNetV2B0. Peningkatan kinerja model dilakukan melalui augmentasi data serta *tuning hyperparameter* pada beberapa variabel utama, seperti *learning rate*, *dropout rate*, *batch size*, dan jumlah *dense layer*. Evaluasi model dilakukan dengan menggunakan *confusion matrix* untuk mengukur performa klasifikasi berdasarkan metrik akurasi, *precision*, *recall*, dan F1-score. Hasil penelitian ini diharapkan dapat

menghasilkan model klasifikasi bahasa isyarat yang optimal, yang nantinya dapat menjadi dasar dalam pengembangan teknologi pendukung komunikasi bagi penyandang gangguan pendengaran dan masyarakat luas.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka dapat dirumuskan beberapa permasalahan sebagai berikut:

1. Bagaimana performa arsitektur CNN EfficientNetV2B0 dalam klasifikasi ASL?
2. Bagaimana pengaruh *hyperparameter tuning* terhadap akurasi klasifikasi ASL menggunakan EfficientNetV2B0?
3. Bagaimana kontribusi teknik augmentasi data dalam meningkatkan performa model klasifikasi ASL?

1.3 Tujuan dan Manfaat

Berdasarkan rumusan masalah yang telah diidentifikasi, tujuan penelitian ini adalah sebagai berikut:

1. Mengklasifikasikan American Sign Language menggunakan CNN dengan arsitektur EfficientNetV2B0.
2. Menganalisis pengaruh *hyperparameter tuning* terhadap performa model klasifikasi ASL.
3. Mengevaluasi efektivitas teknik augmentasi data dalam meningkatkan akurasi klasifikasi ASL.

Manfaat dari penelitian ini diharapkan dapat memberikan kontribusi ilmiah dalam pengembangan metode *deep learning*, khususnya pada penerapan *hyperparameter tuning* dan augmentasi data dalam klasifikasi bahasa isyarat ASL. Selain itu, hasil penelitian ini diharapkan dapat menjadi referensi bagi pengembangan sistem pengenalan bahasa isyarat yang lebih akurat dan efisien.

1.4 Ruang Lingkup

Ruang lingkup mempunyai tujuan memberikan batasan dari penelitian ini agar tidak menyimpang dan sesuai dengan target yang diinginkan. Ruang lingkup dalam penelitian ini adalah sebagai berikut:

1. Penelitian dilakukan menggunakan bahasa pemrograman Python dan menggunakan Google Colaboratory.
2. Dataset yang digunakan dalam penelitian ini adalah dataset American Sign Language oleh Mavi & Dikle (2022) yang diunduh dari Kaggle.
3. Citra pada dataset merupakan citra statis, sehingga model yang dikembangkan tidak dapat menangani gerakan dinamis.
4. Jumlah citra yang digunakan sebanyak 22.801 citra dan terdiri dari 27 kelas, yang mencakup angka, huruf, dan ekspresi, serta satu kelas tambahan yaitu "NULL".
5. Data dibagi menjadi 80% untuk data latih, 10% untuk data validasi, dan 10% untuk data uji, dengan pembagian yang dilakukan hanya sekali di awal.
6. Arsitektur model yang digunakan dalam penelitian ini adalah EfficientNetV2B0.
7. Penelitian membandingkan performa model dengan beberapa skenario, meliputi: pelatihan menggunakan data tanpa augmentasi dan data dengan augmentasi, serta pengaruh *hyperparameter tuning* pada *learning rate*, *batch size*, *dropout*, dan *dense layer*.
8. Model memproses setiap data masukan yang diberikan, termasuk data baru, untuk melakukan prediksi atau klasifikasi sesuai dengan pola yang telah dipelajari selama pelatihan.

1.5 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penyusunan penelitian ini terbagi dalam beberapa pokok bahasan, yaitu:

BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang, rumusan masalah, tujuan dan manfaat penelitian, ruang lingkup penelitian, serta sistematika penulisan dalam penyusunan laporan skripsi mengenai klasifikasi American Sign Language menggunakan arsitektur EfficientNetV2B0 dengan *hyperparameter tuning*.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan hasil studi pustaka mengenai penelitian sebelumnya serta dasar teori mengenai bahasa isyarat, citra digital, pemrosesan citra, *one-hot encoding*, klasifikasi citra, dataset, *deep learning*, CNN,

activation function, adam optimizer, hyperparameter, early stopping, EfficientNetV2B0, dan evaluasi model yang digunakan dalam pelaksanaan serta penyusunan laporan skripsi mengenai klasifikasi American Sign Language menggunakan arsitektur EfficientNetV2B0 dengan hyperparameter tuning.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan metodologi penelitian yang digunakan pada penelitian klasifikasi American Sign Language menggunakan arsitektur EfficientNetV2B0 dengan *hyperparameter tuning*, meliputi garis besar penyelesaian masalah, pengumpulan data, pra-pemrosesan data, augmentasi data, pembagian data, pelatihan dan validasi model, pengujian model serta evaluasi model.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari tahap-tahap dalam penelitian meliputi hasil pengumpulan data, pra-pemrosesan data, augmentasi data, pembagian data, pelatihan dan validasi model, pengujian model, dan evaluasi model dalam klasifikasi American Sign Language menggunakan arsitektur EfficientNetV2B0 dengan *hyperparameter tuning*.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian mengenai klasifikasi American Sign Language menggunakan arsitektur EfficientNetV2B0 dengan *hyperparameter tuning* serta saran untuk pengembangan lebih lanjut.