

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit paru-paru, khususnya Tuberkulosis (TB), telah kembali menjadi penyebab utama kematian akibat agen infeksius tunggal di dunia, menggantikan COVID-19 setelah tiga tahun pandemi. Menurut *Global Tuberculosis Report 2024* yang dirilis oleh *World Health Organization* (WHO), diperkirakan 10,8 juta orang jatuh sakit akibat TB pada tahun 2023, dan penyakit ini menyebabkan sekitar 1,25 juta kematian secara global [1]. Tingginya angka morbiditas dan mortalitas ini menuntut adanya deteksi dini yang cepat dan akurat. Meskipun pemeriksaan sinar-X dada (*Chest X-ray*) menjadi standar medis utama untuk skrining awal, interpretasinya sangat bergantung pada subjektivitas dan keahlian radiolog yang rentan terhadap kelelahan visual (*visual fatigue*) akibat tingginya volume pasien, yang dapat berujung pada kesalahan diagnosis [2], [3].

Teknologi *Deep Learning*, khususnya *Convolutional Neural Network* (CNN), telah terbukti efektif dalam mengekstraksi fitur visual kompleks untuk analisis citra medis [4], [5]. Namun, penerapan CNN menghadapi kendala utama berupa keterbatasan ketersediaan dataset medis yang seimbang karena alasan privasi pasien [2]. Minimnya variasi data latih ini menyebabkan model rentan mengalami *overfitting*, yaitu kondisi di mana model memiliki akurasi tinggi saat pelatihan namun gagal mengenali pola penyakit pada data baru [6].

Guna mengatasi kelangkaan data dan mencegah *overfitting*, teknik *Data Augmentation* diterapkan sebagai solusi efisien untuk memperkaya variasi data secara sintesis tanpa perlu menambah pasien fisik. dalam penelitian [3], penerapan skenario augmentasi yang tepat dapat meningkatkan metrik evaluasi model secara signifikan. Melalui transformasi geometri seperti rotasi, pergeseran, dan pembesaran citra, teknik ini memanipulasi data latih agar model mempelajari fitur yang lebih invarian terhadap perubahan posisi dan kondisi, yang dinilai krusial untuk meningkatkan ketahanan model pada klasifikasi multi-kelas yang melibatkan struktur sensitif.

Dalam konteks pengembangan sistem kecerdasan buatan untuk medis, sekedar mengetahui pengaruh dari suatu teknik belumlah cukup untuk menghasilkan alat bantu diagnosis yang aplikatif. Tantangan utamanya terletak pada bagaimana menerapkan teknik *Data Augmentation* tersebut secara terintegrasi ke dalam siklus pengembangan model CNN yang utuh. Pemilihan arsitektur *MobileNetV2* dalam penelitian ini didasarkan pada efisiensi komputasinya, namun model yang ringan ini sangat membutuhkan regulasi data yang kuat agar terhindar dari *overfitting* saat dihadapkan pada data yang terbatas. Oleh karena itu, *Data Augmentation* tidak lagi diposisikan sekedar sebagai variabel yang dianalisis secara statistik, melainkan sebagai komponen rekayasa fundamental yang wajib diterapkan untuk membangun arsitektur deteksi penyakit paru yang stabil, *robust*, dan memiliki kemampuan generalisasi tinggi.

Berdasarkan urgensi klinis dan kendala teknis tersebut, penelitian ini bermaksud melakukan analisis pengaruh teknik *Data Augmentation* terhadap akurasi model CNN (khususnya arsitektur *MobileNetV2*) dalam mengklasifikasikan citra *X-ray* paru ke dalam tiga kelas kondisi: *Normal*, *Pneumonia*, dan *Tuberculosis*. Alur penelitian dimulai dengan pengumpulan dataset sekunder, prapemrosesan (*resizing* dan *rescaling*), penerapan augmentasi data dinamis pada data latih, pelatihan model menggunakan metode *transfer learning*, dan diakhiri dengan evaluasi performa menggunakan *confusion matrix* untuk memastikan akurasi model memenuhi standar medis. Harapan dalam penelitian ini adalah mampu menghasilkan model diagnostik yang akurat, *robust*, dan memiliki kemampuan generalisasi yang baik pada dataset terbatas.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana menerapkan teknik *data augmentation* pada dataset citra X-Ray agar dapat mengatasi keterbatasan jumlah dan variasi data?
2. Bagaimana mengembangkan arsitektur model *Convolutional Neural Network* (CNN) untuk klasifikasi deteksi penyakit paru menggunakan dataset yang telah diaugmentasi??

3. Bagaimana performa (akurasi, *precision*, *recall*, *f1-score*) dari model CNN yang telah dikembangkan dalam mendeteksi penyakit paru?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan fokus, penulis menetapkan batasan masalah sebagai berikut:

1. Data: Dataset yang digunakan adalah data sekunder publik (sumber: Kaggle/rupeshmahanty) yang terdiri dari citra X-ray toraks dengan tiga label kelas: *Normal*, *Pneumonia*, dan *Tuberculosis*.
2. Metode: Algoritma *Deep Learning* yang digunakan adalah *Convolutional Neural Network* (CNN) dengan arsitektur *MobileNetV2*.
3. Teknik Augmentasi: Teknik *Data Augmentation* yang dianalisis dibatasi pada transformasi geometri, yaitu: *Rotation* (rotasi), *Zoom* (perbesaran), *Shear*, dan *Flip*.
4. Evaluasi: Parameter evaluasi performa model menggunakan *Confusion Matrix* yang meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, serta analisis grafik *Loss* dan *Accuracy* untuk mendeteksi *overfitting*.
5. Lingkup: Penelitian ini berfokus pada analisis komputasi dan performa algoritma, tidak mencakup implementasi sistem ke dalam aplikasi klinis atau diagnosis medis secara langsung kepada pasien.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membangun model klasifikasi penyakit paru berbasis CNN pada dataset tiga kelas sebagai *baseline* (tolak ukur) awal.
2. Menganalisis dan membandingkan performa model CNN sebelum dan sesudah diterapkan teknik *Data Augmentation*.
3. Membuktikan efektivitas teknik *Data Augmentation* dalam meningkatkan kemampuan generalisasi model (mengurangi *overfitting*) pada klasifikasi citra medis X-ray.

1.5 Manfaat Penelitian

1. Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi pada literatur ilmu komputer, khususnya di bidang *Computer Vision* dan *Medical Imaging*, mengenai dampak spesifik teknik augmentasi data terhadap akurasi klasifikasi multi-kelas pada citra toraks.

2. Manfaat Praktis

Hasil model terbaik dari penelitian ini diharapkan dapat menjadi referensi dasar bagi pengembangan sistem *Computer Aided Diagnosis* (CAD) yang dapat membantu tenaga medis atau radiolog sebagai opini kedua (*second opinion*) dalam mendiagnosis kelainan paru secara lebih cepat dan akurat.

1.6 Sistematika Penulisan

BAB 1 PENDAHULUAN Bab ini membahas latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA Bab ini berisi teori-teori penunjang yang relevan, seperti konsep *Deep Learning*, *CNN*, *Data Augmentation*, Citra X-ray, serta penelitian-penelitian terdahulu yang sejenis.

BAB 3 METODOLOGI PENELITIAN Bab ini menjelaskan tahapan penelitian mulai dari pengumpulan dataset, pra-pemrosesan data, skenario augmentasi, perancangan arsitektur model, hingga metode evaluasi yang digunakan.

BAB 4 HASIL DAN PEMBAHASAN Bab ini memaparkan hasil eksperimen pelatihan model dengan berbagai skenario, analisis grafik *training*, perbandingan akurasi, dan pembahasan mengenai pengaruh augmentasi.

BAB 5 PENUTUP Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran untuk pengembangan penelitian selanjutnya.