

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Tuberkulosis merupakan penyakit menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Penyakit ini biasanya menyerang paru-paru, tetapi bisa juga mengganggu bagian tubuh lainnya dalam beberapa situasi. Tuberkulosis menjadi salah satu tantangan kesehatan di seluruh dunia yang sangat serius. Pada tahun 2023, diperkirakan terdapat 10,8 juta kasus baru tuberkulosis secara global, meningkat dari 10,7 juta pada tahun 2022 dan jauh di atas 10,1 juta pada tahun 2020 [1]. Mulai tahun 2024, pola ini menunjukkan perubahan yang lebih teratur, dengan perkiraan angka kejadian tuberkulosis mulai menurun. Sementara itu, jumlah kematian akibat tuberkulosis pada tahun 2023 menurun menjadi sekitar 1,25 juta, yang terdiri dari 1,09 juta di antara mereka yang tidak memiliki HIV dan 161 ribu di antara mereka yang terinfeksi HIV, jika dibandingkan dengan 1,32 juta pada tahun 2022[2].

Meskipun terdapat penurunan jumlah kasus dan kematian akibat tuberkulosis pada tahun 2023 dan 2024, pencapaian target strategis WHO untuk menurunkan angka kematian hingga 75% pada tahun 2025 masih belum tercapai. Pengurangan kumulatif selama periode 2015 sampai 2023 hanya mencapai 23%[2]. Salah satu tantangan utama dalam penanganan tuberkulosis adalah keterlambatan diagnosis dan pengobatan, terutama di negara-negara dengan beban tuberkulosis yang tinggi. Saat ini, diagnosis tuberkulosis umumnya dilakukan menggunakan X-ray dada. Sebuah penelitian menunjukkan bahwa X-ray dada dapat mengidentifikasi 40-79% kasus tuberkulosis pada pasien yang tidak menunjukkan gejala tradisional [3]. Namun, metode ini memiliki sejumlah keterbatasan. Penilaian hasil X-ray oleh tenaga kesehatan sering dipengaruhi oleh perbedaan individu, sensitivitas yang rendah pada tahap awal penyakit, serta potensi kelelahan yang dapat mengurangi ketepatan dan konsentrasi selama diagnosis. Dengan keterbatasan-keterbatasan tersebut, teknologi berbasis *machine learning* dapat menjadi solusi yang dapat meningkatkan efisiensi dan keakuratan dalam deteksi tuberkulosis pada citra X-ray dada. *Machine learning* tradisional seperti Support Vector Machine (SVM), K-Nearest Neighbor (KNN), dan Random Forest telah digunakan untuk analisis data medis dalam waktu yang lama. Metode ini mengandalkan ekstraksi fitur manual (seperti tekstur, bentuk, atau

intensitas piksel) yang memerlukan keahlian dibidang terkait untuk memahami parameter relevan. Keunggulan utama dari *machine learning* tradisional adalah kebutuhan akan data yang lebih sedikit dan hasil yang lebih mudah dipahami. Namun, metode ini kurang efisien dalam mengatasi kompleksitas pola yang tidak terlihat dalam citra medis. Akurasi dari *machine learning* tradisional sangat tergantung pada kualitas fitur yang dipilih, sehingga dapat terpengaruh oleh bias dari manusia[4].

CNN (Convolutional Neural Network) adalah suatu jenis jaringan saraf tiruan yang dirancang agar mampu mengolah data berbentuk grid seperti gambar. Berbeda dengan *machine learning* tradisional, CNN mampu mengekstraksi fitur secara otomatis dari data mentah melalui operasi konvolusi berstruktur. CNN dapat secara efisien dan akurat mengenali pola, bentuk, atau suatu struktur dari gambar yang diinput. Oleh karena itu, CNN mampu secara efisien mendeteksi pola-pola halus yang sering kali tidak terlihat oleh pengamatan manusia, termasuk tanda-tanda awal tuberkulosis pada citra X-ray [5]. Meskipun CNN telah menunjukkan hasil yang menjanjikan, tetapi tantangan besar tetap ada dalam mengoptimalkan model untuk mencapai kinerja yang maksimal. Penyesuaian *hyperparameter* yang tepat sangat diperlukan selama proses pelatihan model, karena hal ini secara langsung mempengaruhi cara kerja model dan kemampuannya dalam memproses serta memahami data dengan baik. Dalam hal ini, *Hyperband* dapat menjadi salah satu cara untuk mengoptimalkan model. *Hyperband* adalah metode pencarian *hyperparameter* yang efisien, yang dapat secara otomatis menemukan kombinasi *hyperparameter* terbaik untuk model CNN. Metode ini bekerja dengan menggunakan pendekatan berbasis pemrograman bandit untuk mengalokasikan sumber daya komputasi secara optimal dalam pencarian *hyperparameter* [6]. *Hyperband* memiliki keunggulan dibandingkan metode pencarian *hyperparameter* lainnya karena mampu mengevaluasi lebih banyak kombinasi *hyperparameter* dalam waktu lebih singkat melalui mekanisme *successive halving*. Pendekatan ini membuat *Hyperband* lebih efisien dibandingkan grid search, random search, maupun Bayesian optimization, terutama saat menangani model CNN yang membutuhkan waktu pelatihan cukup lama. Dengan menggunakan *Hyperband*, pengoptimalan model CNN dapat lebih cepat dan efisien, sehingga menghasilkan model yang lebih akurat dan dapat diandalkan untuk deteksi tuberkulosis.

Berbagai penelitian sebelumnya menunjukkan bahwa CNN telah berhasil

diimplementasikan dalam deteksi citra medis. Sebagai contoh, sebuah penelitian menggunakan data National Library of Medicine dan berhasil membangun model CNN dengan akurasi klasifikasi 84% pada data uji untuk mendeteksi tuberkulosis [7]. Penelitian lainnya mengulas sistem diagnostik berbantuan komputer (CAD) berbasis CNN yang mampu mempercepat diagnosis tuberkulosis melalui analisis otomatis citra X-ray dada, yang mencakup arsitektur seperti AlexNet, VGGNet, dan ResNet [8]. Di sisi lain, beberapa penelitian terdahulu turut menegaskan peran krusial tuning hyperparameter adaptif untuk meningkatkan kinerja model CNN dalam klasifikasi citra medis[9]. Lebih lanjut, penelitian lainnya mengoptimalkan CNN dengan menggunakan algoritma *Hyperband*, yang meningkatkan akurasi dan sensitivitas diagnosis berbasis CT scan hingga 97% untuk COVID-19, menunjukkan potensi metode ini dalam domain medis yang lebih luas [10].

Berdasarkan permasalahan yang telah diuraikan sebelumnya, penelitian ini bertujuan untuk mengembangkan model Convolutional Neural Network (CNN) untuk klasifikasi tuberkulosis yang dioptimalkan menggunakan metode *Hyperband*. Selain itu, penelitian ini juga akan membandingkan kinerja model sebelum dan sesudah optimasi guna mengevaluasi pengaruh *Hyperband* terhadap kinerja model. Selain itu, penelitian ini akan menganalisis dampak perbedaan rasio pembagian dataset terhadap kinerja model untuk menentukan konfigurasi data yang paling optimal. Melalui pengembangan model ini, diharapkan dapat mencapai hasil yang memuaskan, dan dapat mendukung otomatisasi diagnosis penyakit tuberkulosis serta berkontribusi dalam upaya penanganan dan pengendalian tuberkulosis.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang yang telah dijelaskan sebelumnya, maka dirumuskan beberapa permasalahan dalam penelitian ini.

- a. Bagaimana merancang model Convolutional Neural Network untuk klasifikasi citra X-ray dada ke dalam kelas Normal dan Tuberculosis?
- b. Bagaimana pengaruh optimasi Hyperband dan variasi rasio pembagian dataset terhadap kinerja model Convolutional Neural Network?

## **1.3 Tujuan Penelitian**

Berdasarkan latar belakang di atas, berikut merupakan tujuan dilakukannya penelitian ini.

- a. Mengembangkan model Convolutional Neural Network untuk mengklasifikasikan citra X-ray dada ke dalam kelas Normal atau Tuberculosis.
- b. Mengevaluasi pengaruh optimasi Hyperband dan variasi rasio pembagian dataset terhadap kinerja model Convolutional Neural Network.

#### **1.4 Manfaat Penelitian**

Penelitian ini diharapkan dapat berperan sebagai referensi baru dalam penerapan Convolutional Neural Network dengan optimasi *Hyperband* untuk klasifikasi citra medis, khususnya pada X-ray dada dalam mendiagnosis tuberkulosis. Selain itu, hasil penelitian ini diharapkan dapat diterapkan secara nyata untuk mendukung otomatisasi diagnosis penyakit tuberkulosis serta berperan dalam upaya penanganan dan pengendalian tuberkulosis.

#### **1.5 Batasan Masalah**

Adapun batasan masalah dalam menjalankan penelitian ini antara lain:

- a. Penelitian ini dibatasi pada penggunaan arsitektur Convolutional Neural Network (CNN) dasar dan tidak menggunakan arsitektur modern lain, seperti ResNet maupun VGG.
- b. Metode optimasi yang digunakan dalam penelitian ini terbatas pada Hyperband, tanpa dilakukan perbandingan dengan metode optimasi lainnya.
- c. Dataset yang digunakan berupa citra X-ray dada dengan dua kelas, yaitu Normal dan Tuberculosis, yang diperoleh dari tiga sumber data di platform Kaggle.
- d. Evaluasi kinerja model hanya dilakukan dengan menggunakan metrik akurasi, presisi, recall, dan F1-score.