

BAB I PENDAHULUAN

1.1 Latar Belakang

Pneumonia merupakan salah satu penyakit paru yang menjadi penyebab utama morbiditas dan mortalitas di dunia, terutama pada kelompok rentan seperti anak-anak dan lansia. Tingginya angka kejadian dan kematian akibat pneumonia menunjukkan bahwa deteksi dini penyakit ini sangat penting untuk mempercepat proses diagnosis serta meningkatkan peluang kesembuhan pasien. Dalam praktik klinis, salah satu modalitas pencitraan medis yang paling umum digunakan untuk diagnosis penyakit paru adalah citra *X-ray* dada, karena bersifat non-invasif, relatif murah, dan mudah diakses di berbagai fasilitas pelayanan kesehatan [1]. Selain pneumonia, penyakit paru-paru secara umum juga tercatat sebagai penyumbang signifikan terhadap beban penyakit global, khususnya pada kelompok penyakit pernapasan kronis [2].

Dalam praktik radiologi, citra *X-ray* dada sering kali menampilkan lebih dari satu temuan patologis secara bersamaan, seperti *cardiomegaly*, *edema*, *atelectasis*, atau *pneumothorax*. Kondisi ini menunjukkan bahwa diagnosis berbasis citra *X-ray* memiliki tingkat kompleksitas yang tinggi, karena satu citra dapat merepresentasikan berbagai kondisi patologis paru secara simultan. Oleh karena itu, permasalahan ini tidak dapat direpresentasikan sebagai klasifikasi satu kelas tunggal, melainkan memerlukan pendekatan klasifikasi *multi-label* (*multi-criteria*) untuk mengakomodasi seluruh temuan patologis dalam satu citra [3].

Dataset *CheXpert-v1.0-small* yang digunakan dalam penelitian ini dirancang untuk merepresentasikan kompleksitas tersebut dengan menyediakan klasifikasi empat belas kategori kondisi paru, yaitu *No Finding*, *Enlarged Cardiomeastinum*, *Cardiomegaly*, *Lung Opacity*, *Lung Lesion*, *Edema*, *Consolidation*, *Pneumonia*, *Atelectasis*, *Pneumothorax*, *Pleural Effusion*, *Pleural Other*, *Fracture*, dan *Support Devices* [4]. Kompleksitas dataset ini diperkuat oleh fakta bahwa proses interpretasi citra *X-ray* secara manual masih sangat bergantung

pada keahlian radiolog, sehingga berpotensi menimbulkan subjektivitas dan variasi antar-pengamat. Kondisi tersebut dapat berdampak pada perbedaan hasil diagnosis serta keterlambatan dalam pengambilan keputusan klinis.

Seiring dengan perkembangan teknologi, pengembangan sistem otomatis berbasis *deep learning* sebagai bagian dari *data mining* telah menunjukkan potensi besar sebagai sistem pendukung keputusan klinis. Pendekatan ini dinilai mampu membantu proses diagnosis secara lebih cepat, konsisten, dan akurat dalam analisis citra medis [5]. Berbagai arsitektur *deep learning* telah dikembangkan untuk tugas klasifikasi citra medis, termasuk *Convolutional Neural Network* (CNN) yang terbukti efektif dalam menangkap fitur lokal pada citra *X-ray*. Namun demikian, arsitektur CNN memiliki keterbatasan dalam menangkap ketergantungan jangka panjang dan konteks global pada citra *X-ray* beresolusi tinggi, karena sifat lokal dari operasi konvolusi. Keterbatasan ini mendorong eksplorasi arsitektur alternatif yang lebih mampu merepresentasikan hubungan global antar bagian citra [6].

Salah satu pendekatan yang berkembang pesat adalah *Vision Transformer* (ViT), yang mengadaptasi mekanisme *self-attention* untuk memproses citra sebagai urutan *patch*, sehingga memungkinkan model menangkap hubungan global antar *patch* citra secara lebih efektif dibandingkan pendekatan konvolusional tradisional [7]. Sejumlah penelitian menunjukkan bahwa ViT memiliki potensi yang menjanjikan dalam analisis citra medis, termasuk pada citra *X-ray* dada, terutama ketika dikombinasikan dengan teknik *transfer learning* yang mampu mempercepat proses pelatihan model dan meningkatkan performa pada data terbatas [8]. Meskipun demikian, penerapan ViT masih menghadapi tantangan praktis, seperti kebutuhan sumber daya komputasi yang relatif tinggi serta keterbatasan efisiensi ketika diterapkan pada citra beresolusi tinggi dan skenario *multi-label*.

Sebagai pengembangan dari arsitektur *Transformer*, *Swin Transformer* diperkenalkan untuk meningkatkan efisiensi komputasi tanpa mengorbankan kemampuan representasi fitur. *Swin Transformer* menerapkan mekanisme *shifted window-based self-attention* yang memungkinkan pemodelan konteks lokal dan global secara hierarkis, sehingga lebih efisien dalam memproses citra beresolusi

tinggi. Pendekatan ini telah menunjukkan kinerja yang menjanjikan pada tugas klasifikasi *multi-label* citra *X-ray* dada, termasuk pada dataset penyakit paru [9].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan studi perbandingan eksperimental terhadap kinerja *Vision Transformer* (ViT) dan *Swin Transformer* melalui proses *fine-tuning* pada dataset *CheXpert-v1.0-small*. Penelitian ini difokuskan pada tugas klasifikasi *multi-label* penyakit paru, dengan perhatian khusus pada deteksi pneumonia. Hasil penelitian diharapkan dapat memberikan gambaran mengenai efektivitas dan efisiensi kedua arsitektur *Transformer* sebagai sistem pendukung diagnosis pneumonia yang akurat dan andal.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Seberapa besar pengaruh penerapan metode *fine-tuning* terhadap peningkatan kinerja model *Vision Transformer* (ViT) dan *Swin Transformer* dalam klasifikasi *multi-label* citra *X-ray* paru-paru pada dataset *CheXpert-v1.0-small*?
2. Bagaimana kinerja model *Vision Transformer* (ViT) dan *Swin Transformer* dalam melakukan klasifikasi *multi-label* citra *X-ray* paru-paru berdasarkan metrik evaluasi multi-kriteria, seperti akurasi, sensitivitas, dan spesifisitas?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini ditetapkan untuk menjaga fokus dan keterbatasan penelitian sehingga hasil yang diperoleh dapat lebih terukur. Batasan masalah penelitian ini adalah sebagai berikut:

1. Penelitian hanya menggunakan dataset *CheXpert-v1.0-small* yang terdiri dari citra *sinar-X* dada, dengan 14 kategori label (*CHEXP14*), yaitu *No Finding*, *Enlarged Cardiomeastinum*, *Cardiomegaly*, *Lung Opacity*, *Lung Lesion*, *Edema*, *Consolidation*, *Pneumonia*, *Atelectasis*, *Pneumothorax*, *Pleural Effusion*, *Pleural Other*, *Fracture*, dan *Support Devices*.

2. Penelitian hanya membandingkan dua arsitektur model berbasis *transformer*, yaitu *Vision Transformer (ViT)* dan *Swin Transformer*.
3. Evaluasi performa model dibatasi pada klasifikasi multi-label dengan pendekatan evaluasi multi-kriteria, meliputi akurasi, sensitivitas, dan spesifisitas.
4. Data tambahan seperti informasi klinis pasien, demografi, atau variabel medis lain tidak digunakan dan dianggap parameter konstan.
5. Proses pelatihan model dilakukan menggunakan metode fine-tuning, tanpa mengembangkan arsitektur baru atau memodifikasi algoritma inti dari model *transformer*.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai dalam penelitian ini adalah sebagai berikut:

1. Menganalisis dan membandingkan performa model *Vision Transformer (ViT)* dan *Swin Transformer* dalam melakukan klasifikasi multi-label citra X-ray paru-paru untuk mendeteksi pneumonia.
2. Mengevaluasi pengaruh penerapan metode fine-tuning terhadap peningkatan akurasi, sensitivitas, dan spesifisitas kedua model pada dataset *CheXpert-v1.0-small*.
3. Memvalidasi penerapan pendekatan klasifikasi multi-label dan evaluasi berbasis multi-kriteria dalam menilai kinerja model sehingga diperoleh informasi yang komprehensif dan terukur terkait kemampuan deteksi pneumonia secara otomatis.

1.5 Manfaat Penelitian

Manfaat penelitian ini dapat dijelaskan sebagai berikut:

1. Manfaat Teoritis

Penelitian ini memberikan wawasan mengenai penerapan arsitektur *Vision Transformer (ViT)* dan *Swin Transformer* pada citra X-ray paru-paru untuk deteksi pneumonia. Selain itu, penelitian ini dapat menjadi referensi dalam pengembangan pendekatan klasifikasi multi-label dan evaluasi berbasis

multi-kriteria pada klasifikasi penyakit paru berbasis *deep learning*. Penelitian ini juga memperluas pemahaman mengenai penggunaan metode fine-tuning dalam meningkatkan performa model transformer pada data medis.

2. Manfaat Praktis

Penelitian ini dapat membantu tenaga medis dan institusi kesehatan dalam pengembangan sistem deteksi pneumonia otomatis yang lebih cepat dan akurat. Selain itu, hasil penelitian ini dapat menjadi acuan bagi peneliti atau pengembang lain dalam implementasi model transformer pada citra medis serupa, serta menjadi dasar pengembangan sistem berbasis kecerdasan buatan untuk meningkatkan efisiensi diagnosis dan kualitas layanan radiologi digital.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran umum alur penelitian dan isi masing-masing bab:

BAB I PENDAHULUAN berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian. Bab ini menjelaskan alasan, konteks, dan tujuan dilakukannya penelitian.

BAB II TINJAUAN PUSTAKA berisi teori dan konsep dasar mengenai pneumonia, dataset *CheXpert-v1.0-small*, *deep learning*, *Vision Transformer* (ViT), *Swin Transformer*, serta penelitian terdahulu yang relevan sebagai acuan pengembangan model.

BAB III METODE PENELITIAN Membahas desain penelitian, metode fine-tuning pada model ViT dan Swin Transformer, tahapan *preprocessing* data, strategi pelatihan, evaluasi model menggunakan metrik multi-kriteria, serta alur kerja eksperimen secara sistematis.

BAB IV HASIL DAN PEMBAHASAN Menyajikan hasil eksperimen, perbandingan performa model Vision Transformer dan Swin Transformer, interpretasi hasil, serta analisis akurasi, sensitivitas, dan spesifisitas dalam deteksi

pneumonia.

BAB V PENUTUP Berisi kesimpulan penelitian, jawaban atas rumusan masalah, saran untuk penelitian selanjutnya, serta rekomendasi implementasi model dalam konteks klinis.

