

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komputasi dalam dua dekade terakhir telah mengalami kemajuan yang sangat pesat. Salah satu kemajuan paling signifikan adalah teknologi kecerdasan buatan (*Artificial Intelligence/AI*). AI kini telah berevolusi dari yang sebelumnya hanya berfungsi sebagai alat bantu otomasi proses menjadi teknologi yang mampu meniru proses kognitif manusia, seperti belajar, bernalar, memahami konteks, dan mengambil keputusan secara otonom. Kehadiran AI telah menjadi fondasi utama revolusi industri 4.0 yang berperan penting dalam mendukung transformasi digital di bidang ekonomi, manufaktur, transportasi, pertahanan, hingga kesehatan melalui aplikasinya yang luas seperti rekomendasi konten digital, pengenalan wajah dan suara, *autonomous driving*, hingga pengambilan keputusan strategis berbasis *big data* (Jordan dan Mitchell, 2015).

Salah satu cabang AI yang mengalami kemajuan pesat adalah *computer vision*, yaitu bidang yang memungkinkan komputer untuk “melihat” dan memahami informasi visual dari citra atau video sebagaimana manusia memproses penglihatan. Sejak keberhasilan *Convolutional Neural Network* (CNN) dalam kompetisi *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) oleh Krizhevsky *et al.* (2012), *computer vision* telah mengalami perkembangan signifikan. Kemajuan pesat ini didorong oleh peningkatan kemampuan komputasi,

ketersediaan dataset dalam jumlah besar, serta inovasi algoritma *deep learning* yang mampu mengekstrak dan merepresentasikan fitur kompleks secara otomatis.

CNN mengekstrak fitur lokal seperti tepi, tekstur, dan bentuk dari citra melalui operasi konvolusi sehingga memiliki keterbatasan dalam menangkap konteks global dari sebuah citra. Hal ini disebabkan oleh prinsip kerja CNN yang berbasis pada *local receptive field*, yaitu area terbatas pada citra yang diproses oleh setiap kernel konvolusi, sehingga informasi spasial dan hubungan kontekstual antarbagian citra sering kali tidak tertangkap secara optimal (Khan *et al.*, 2022). Keterbatasan ini menjadi kendala serius ketika analisis visual membutuhkan pemahaman konteks global dan hubungan spasial yang luas, seperti pada citra medis yang sering kali memerlukan informasi pola distribusi fitur patologis secara menyeluruh (Sarvamangala dan Kulkarni, 2022). Salah satu contohnya dapat ditemukan pada kasus *diabetic retinopathy* (DR), di mana informasi spasial global dari citra retina memegang peran penting dalam penentuan tingkat keparahan penyakit secara menyeluruh.

Diabetic retinopathy merupakan komplikasi mikrovaskular dari diabetes mellitus yang ditandai oleh kerusakan pembuluh darah retina akibat hiperglikemia kronis dan dapat menimbulkan gangguan fungsi penglihatan (Teo *et al.*, 2021). Secara umum, tingkat keparahan *diabetic retinopathy* diklasifikasikan bertahap mulai dari *mild*, *moderate*, *severe non-proliferative diabetic retinopathy* (NPDR), hingga tahap *proliferative diabetic retinopathy* (PDR), yang masing-masing menunjukkan progresivitas kerusakan vaskular retina (Wilkinson *et al.*, 2003). Menurut *International Diabetes Federation* (IDF), prevalensi diabetes secara global mencapai lebih dari 537 juta jiwa pada tahun 2021 dan diperkirakan akan

meningkat menjadi 643 juta jiwa pada tahun 2030 dengan sekitar sepertiga penderita mengalami *diabetic retinopathy* dalam berbagai tingkat keparahan (Magliano dan Boyko, 2021). Deteksi dini *diabetic retinopathy* menjadi langkah penting untuk memungkinkan pemberian intervensi lebih awal guna mencegah progresi penyakit ke tahap yang lebih berat. Salah satu cara yang dapat dilakukan adalah melalui analisis citra retina dari *fundus camera* untuk mengidentifikasi perubahan patologi pada retina.

Analisis citra retina memerlukan pemahaman spasial yang menyeluruh karena perbedaan tingkat keparahan *diabetic retinopathy* seringkali tidak hanya ditentukan oleh keberadaan lesi, yaitu perubahan patologi pada jaringan retina, tetapi juga oleh distribusi, pola penyebaran, dan hubungan antar-lesi pada skala global. Proses ini mencakup identifikasi jenis lesi, seperti adanya *microaneurysms* (titik-titik merah kecil), *hemorrhages* (perdarahan), *hard exudates* (bercak kuning), dan *cotton wool spots*, dan jumlah lesi yang muncul. Aspek lain yang sangat penting adalah lokasi dan distribusi lesi, yaitu di mana lesi tersebut berada, apakah terpusat di dekat makula (pusat retina) atau tersebar luas. Keberadaan *neovascularization* (pembuluh darah baru yang abnormal) di dekat diskus optikus juga menjadi penanda bahwa penyakit telah memasuki tahap yang lebih parah (Yang *et al.*, 2022). Karakteristik seperti ini sulit ditangkap secara menyeluruh oleh CNN karena sifatnya yang terbatas dalam memodelkan relasi antarbagian citra. Hal tersebut menyebabkan informasi spasial terkait distribusi lesi dan pola global sering kali terabaikan sehingga dapat menurunkan akurasi dalam penentuan tingkat keparahan penyakit.

Vision Transformer (ViT) merupakan arsitektur *deep learning* yang dikembangkan untuk mengatasi keterbatasan CNN dalam memahami konteks spasial secara menyeluruh. Berbeda dengan CNN yang mengandalkan operasi konvolusi, ViT memanfaatkan mekanisme *self-attention* untuk menangkap hubungan antar seluruh bagian citra sehingga mampu merepresentasikan pola distribusi dan keterkaitan lesi dalam skala global secara lebih komprehensif (Dosovitskiy *et al.*, 2021). Kemampuan ini memungkinkan model mengenali pola patologis kompleks, termasuk interaksi antarlesi yang berjauhan, yang sangat krusial dalam proses analisis citra retina. ViT menawarkan pendekatan yang lebih efektif dibandingkan CNN dalam tugas identifikasi tingkat keparahan *diabetic retinopathy*, karena mampu memanfaatkan informasi spasial secara optimal untuk menghasilkan hasil analisis yang lebih tepat (He *et al.*, 2022).

Permasalahan lain dalam klasifikasi tingkat keparahan *diabetic retinopathy* yaitu struktur labelnya yang bersifat ordinal. Label tersebut memiliki urutan atau tingkatan yaitu *mild*, *moderate*, *severe*, hingga *proliferative*. Pendekatan klasifikasi *multi-class* biasa kurang optimal karena mengabaikan informasi urutan variabel target, sedangkan pendekatan regresi kurang tepat karena perbedaan antarlevel ordinal tidak dapat dikuantifikasikan secara numerik. Pendekatan *extended binary classification* oleh Li dan Lin (2007) dengan memecah tugas prediksi menjadi beberapa *binary classification tasks* dapat menyebabkan inkonsistensi prediksi antarlevel kelas. Metode *Consistent Rank Logits* (CORAL) oleh Cao *et al.* (2020) mengatasi masalah inkonsistensi tersebut dengan menerapkan batasan *weight-sharing* pada lapisan terakhirnya sehingga berpotensi membatasi kapasitas representasi model.

Conditional Ordinal Regression for Neural Networks (CORN) dikembangkan oleh Shi *et al.* (2023) dengan mempertahankan konsistensi peringkat antarlevel kelas tanpa menerapkan batasan *weight-sharing*. CORN dimodelkan melalui serangkaian *conditional probabilities* yang saling bergantung secara berurutan. CORN menggunakan prinsip *conditional training subset* dan *chain rule of probability* untuk memastikan bahwa probabilitas prediksi bersifat monoton menurun. Pendekatan ini memungkinkan model untuk secara adaptif mempelajari hubungan hierarkis antarlevel kelas tanpa mengurangi kapasitas representasinya, sehingga lebih sesuai dalam menangani struktur label ordinal pada tugas identifikasi tingkat keparahan *diabetic retinopathy*.

Penelitian mengenai ViT dan CORN telah dilakukan secara terpisah. Penelitian yang dilakukan oleh Matsoukas *et al.* (2021) membandingkan kinerja *Convolutional Neural Network* (CNN) dan *Vision Transformer* (ViT) pada tiga dataset citra medis, yaitu APTOS (berisi gambar retina untuk deteksi *diabetic retinopathy*), ISIC (berisi gambar *dermatoscopic* kulit untuk klasifikasi lesi kulit dan deteksi melanoma), dan CBIS-DDSM (berisi gambar *mammogram* untuk diagnosis kanker payudara). Hasil penelitian menunjukkan bahwa ViT mengungguli CNN pada ketiga dataset ketika menggunakan *pretrained model* dari *ImageNet*. Selain itu, Penelitian yang dilakukan oleh Shi *et al.* (2023) membandingkan beberapa pendekatan klasifikasi ordinal, yaitu *multi-class classification* (CE-NN), *extended binary classification* (OR-NN), *Consistent Rank Logits* (CORAL), dan *Conditional Ordinal Regression for Neural Network* (CORN) pada dataset MORPH-2 (berisi gambar wajah untuk estimasi usia dan pengenalan wajah), AFAD (berisi gambar wajah untuk estimasi usia), dan Fireman (berisi gambar terkait pemadam kebakaran

untuk deteksi api dan klasifikasi situasi). Hasil penelitian menunjukkan bahwa CORN memberikan performa terbaik dengan nilai *mean absolute error* terendah pada dataset MORPH-2 dan AFAD, serta seimbang dengan OR-NN pada dataset Fireman.

Penelitian ini bertujuan untuk menggabungkan *pretrained model Vision Transformer* (ViT) dan *Conditional Ordinal Regression for Neural Network* (CORN) pada dataset RetinaMNIST. Integrasi tersebut diharapkan dapat meningkatkan kinerja model dalam mendeteksi tingkat keparahan *diabetic retinopathy* dengan memanfaatkan konteks spasial global citra retina sekaligus mempertahankan hubungan ordinal antarlevel kelas, sehingga mendukung deteksi dini yang lebih akurat dan pengambilan keputusan klinis yang lebih tepat. *Benchmark* resmi MedMNIST menjadikan ResNet-18 dan ResNet-50 (Yang *et al.*, 2021) sebagai model *baseline* dengan performa terbaik, sehingga arsitektur tersebut akan digunakan sebagai pembandingan dalam penelitian ini.

1.2 Rumusan Masalah

Berdasarkan penjelasan pada latar belakang, maka permasalahan yang dapat diidentifikasi adalah bagaimana implementasi *Vision Transformer* dengan pendekatan *Conditional Ordinal Regression* untuk tugas klasifikasi ordinal tingkat keparahan *diabetic retinopathy* dapat meningkatkan akurasi prediksi dibandingkan model *benchmark* ResNet-18 dan ResNet-50.

1.3 Batasan Masalah

Batasan-batasan masalah dalam penelitian ini dapat dituliskan sebagai berikut:

1. Dataset yang digunakan adalah RetinaMNIST yang merupakan bagian dari *MedMNIST v2 dataset collection* oleh Yang *et al.* (2022), dengan fokus tugas klasifikasi ordinal tingkat keparahan *diabetic retinopathy* berdasarkan citra fundus retina.
2. Model yang dikembangkan terbatas pada *Vision Transformer* dengan pendekatan *Conditional Ordinal Regression*, dan hasilnya dibandingkan dengan *benchmark* model ResNet-18 dan ResNet-50 berdasarkan *metrics* yang bersesuaian.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan model *Vision Transformer* dengan pendekatan *Conditional Ordinal Regression* untuk mendeteksi tingkat keparahan *diabetic retinopathy* serta membandingkan hasilnya dengan *benchmark* model ResNet-18 dan ResNet-50 untuk menilai akurasi prediksi yang dihasilkan.