

1. Pendahuluan

Latar Belakang

Ulcerative colitis (UC), atau *Inflammatory Bowel Disease* (IBD), secara signifikan berdampak pada kualitas hidup pasien. Ditandai dengan peradangan dan luka pada rektum dan usus besar [1], penyakit ini mempengaruhi jutaan orang di seluruh dunia dan meningkatkan risiko kanker. Menurut *British Medical Journal* (BMJ), prevalensi IBD meningkat dari 3,32 juta kasus pada tahun 1990 menjadi 4,90 juta kasus pada tahun 2019, yang mencakup peradangan usus besar dan kecil [2]. Penyebab kolitis ulserativa bersifat multifaktorial, yang melibatkan faktor genetik, lingkungan, dan infeksi. Gejalanya meliputi sakit perut, diare berdarah, penurunan berat badan, demam, dan anemia [1], [3]. Meskipun terdapat banyak pilihan terapi, kolitis ulserativa sering dianggap tidak dapat disembuhkan. Pendekatan diagnostik tradisional seperti tes laboratorium, endoskopi, pencitraan medis, dan penilaian klinis memiliki keterbatasan dalam memprediksi perkembangan penyakit dan respons pengobatan dan juga mahal [4], [5], [6], [7]. Metode alternatif yang menjanjikan adalah implementasi teknik pembelajaran mesin untuk meningkatkan akurasi prediksi menggunakan data ekspresi gen microarray.

Sebuah penelitian terkait implementasi machine learning untuk analisis ekspresi gen pada kolitis ulserativa telah dilakukan oleh beberapa peneliti. Pada tahun 2022, Zhang Lin et al. membandingkan empat metode machine learning yaitu Random Forest, Support Vector Machine (SVM), Gradient Boosting Machine, dan Neural Network pada sepuluh ekspresi gen microarray untuk memprediksi kolitis ulserativa. Metode SVM menghasilkan tingkat kesalahan terendah sebesar 0,16% [5]. Pada tahun 2022, Minchun Bu et al. menggunakan kombinasi Regresi Lasso dan SVM-RFE untuk memprediksi diagnosis UC menggunakan biomarker yang diekspresikan secara berbeda, mencapai akurasi dan skor F1 sebesar 95,20%, dan akurasi validasi sebesar 87,4% [8]. Pada tahun 2021, Popa dan timnya mengembangkan model Neural Network untuk memprediksi aktivitas kolitis ulserativa pada pasien yang menjalani terapi nekrosis anti tumor. Model terbaik dari tiga model mencapai akurasi 94,37% pada data uji dan 93,33% pada data validasi [9]. Pada tahun yang sama, Kraszewski dkk. mengembangkan model prediktif untuk UC menggunakan enam metode pembelajaran mesin, dengan metode Random Forest memberikan hasil terbaik [7]. Mossotto dkk. menggunakan Support Vector Machines dan Recursive Feature Elimination untuk mengklasifikasikan pasien ke dalam subkelompok diagnostik berdasarkan temuan endoskopi dan histologis, dengan tingkat akurasi 82,7% [6].

Berdasarkan hasil penelitian ini, metode pembelajaran mesin telah menunjukkan akurasi yang tinggi dalam memprediksi kolitis ulserativa, namun masih ada ruang untuk perbaikan. Tantangan utama dalam pembelajaran mesin termasuk overfitting, set data yang tidak memadai, dan kesulitan dalam memilih parameter kontrol yang optimal untuk tugas-tugas tertentu [10]. Teknik pembelajaran mendalam dapat mengatasi tantangan ini, terutama *Artificial Neural Network* (ANN), yang merupakan inti dari pembelajaran mendalam dan meniru cara manusia belajar. ANN adalah salah satu teknik yang digunakan untuk prediksi dan model komputasi yang terinspirasi oleh struktur dan fungsi jaringan saraf biologis. Sejak perkembangannya pada tahun 1940-an, ANN telah berevolusi menjadi salah satu metode kecerdasan buatan yang paling kuat dan serbaguna. Awalnya ditemukan sebagai model matematika, ANN mendapatkan popularitas pada tahun 1980-an dengan diperkenalkannya algoritma pelatihan backpropagation. Sejak saat itu, ANN telah berkembang pesat, didorong oleh peningkatan daya komputasi. ANN beroperasi mirip dengan otak manusia, dengan tiga lapisan utama yang didasarkan pada pembelajaran mandiri [11]. Namun, dua kekurangan utama dalam ANN termasuk inisialisasi acak bobot dan bias yang mengarah ke hasil yang tidak optimal dan kebutuhan untuk menyesuaikan jaringan saraf untuk setiap masalah, menentukan jumlah neuron yang optimal di lapisan tersembunyi [11], [12], [13]. Diperlukan optimasi pada pemodelan dalam menyelesaikan permasalahan tersebut.

Penelitian ini bertujuan untuk mengembangkan model prediktif untuk kolitis ulserativa menggunakan ANN yang dioptimalkan dengan *Firefly Algorithm* (FA). FA melakukan optimasi dengan mensimulasikan perilaku berkedip kunang-kunang untuk menemukan solusi terbaik melalui daya tarik. Pengoptimalan ini menentukan jumlah neuron yang optimal di lapisan tersembunyi serta bobot dan bias awal [14], [15], [16]. Dengan mengintegrasikan variabel acak dan adaptif, algoritme ini menekankan eksplorasi dan eksploitasi ruang pencarian dalam berbagai tahap optimasi. Kontribusi dari penelitian ini sangat penting dalam bidang diagnostik medis dan pemodelan prediktif [17]. Dengan mengintegrasikan FA dengan ANN, bertujuan untuk meningkatkan akurasi dan ketahanan model prediktif untuk kolitis ulserativa, yang berpotensi mengubah hasil klinis dan perawatan pasien. Pendekatan ini menawarkan metode yang hemat biaya dan dapat diandalkan untuk diagnosis dini dan penilaian respons pengobatan pada pasien kolitis ulserativa [4]. Selain itu, metodologi baru untuk mengoptimalkan ANN dengan FA dapat diterapkan pada tugas prediksi medis lainnya, yang menunjukkan keserbagunaan dan keefektifannya dalam menangani kumpulan data yang kompleks. Penelitian ini tidak hanya membahas keterbatasan saat ini dalam pemodelan prediktif kolitis ulserativa, tetapi juga menetapkan dasar untuk penelitian di masa depan untuk mengeksplorasi teknik pengoptimalan lanjutan dalam aplikasi medis.

Topik dan Batasannya

Penelitian ini membahas masalah utama prediksi penyakit melalui pemanfaatan pemodelan ANN yang ditingkatkan melalui penggabungan FA. Penelitian ini berkaitan dengan perbandingan akurasi model ANN telah dioptimalkan dengan FA dalam konteks prediksi penyakit berdasarkan data ekspresi gen. Penelitian ini dibatasi

oleh penggunaan data ekspresi gen yang diperoleh dari situs Gene Expression Omnibus (GEO), khususnya dataset seri GSE11223. Dataset ini terdiri dari data ekspresi gen berbasis microarray dari biopsi epitel kolon dari 129 pasien UC dan 73 donor kontrol yang sehat. Kumpulan data ini mencakup 30.957 gen yang diekspresikan di tiga wilayah saluran pencernaan: usus besar distal, usus besar proksimal, dan ileum terminal.

Tujuan

Tujuan dari penelitian ini adalah pertama, untuk mencari nilai akurasi tertinggi dari metode jaringan saraf tiruan menggunakan optimasi FA, serta mengetahui tingkat akurasi, F1-score, *presicion*, dan *recall* dalam mengidentifikasi penyakit ulkeratis kolitis pada data *microarray gene expression*.

Organisasi Tulisan

Dalam laporan penelitian ini, struktur organisasi penulisan disusun dengan beberapa bagian. Bagian pertama, Pendahuluan yang mencakup latar belakang, topik dan batasan, tujuan, serta organisasi penulisan. Bagian kedua, Studi Terkait yaitu memaparkan penelitian-penelitian sebelumnya yang relevan. Bagian Sistem yang Dibangun menjelaskan detail sistem atau model yang dikembangkan. Bagian Evaluasi menyajikan hasil pengujian dan

analisis kinerja sistem. Terakhir, bagian Kesimpulan dan Saran menyimpulkan temuan utama dan memberikan rekomendasi untuk penelitian selanjutnya.