

1. Pendahuluan

Latar Belakang

Ulcerative Colitis (UC) adalah penyakit radang usus kronis yang memengaruhi usus besar, ditandai dengan periode peradangan mukosa yang kambuh secara berulang yang berdampak signifikan pada kehidupan pasien yang terkena dampaknya. Sebuah analisis menyeluruh tentang *Inflammatory Bowel Disease* (IBD) di Asia, berdasarkan data dari Global Burden of Disease Study 2019, mengungkapkan peningkatan yang mencolok dalam *age-standardized prevalence rate* (ASPR) IBD dari 29,81 per 100.000 orang pada tahun 1990 menjadi 39,37 per 100.000 orang pada tahun 2019 dengan insiden dan penyebaran tertinggi di antara semua benua [1]. Gejala klinis UC bervariasi, dengan pasien sering mengalami kelelahan, ketidaknyamanan perut, diare berdarah, dan inkontinensia tinja [2]. Penyebab UC masih belum sepenuhnya dipahami, tetapi secara umum diyakini bahwa faktor genetik, faktor lingkungan, dan kondisi psikologis berkontribusi pada gangguan ekosistem mikroba usus, yang menyebabkan timbulnya penyakit [3].

Diagnosis kolitis ulseratif tidak dapat ditentukan secara pasti hanya dengan pemeriksaan diagnostik tunggal. Hal ini didasarkan analisis keseluruhan dari gejala-gejala klinis, tes laboratorium, dan temuan endoskopi, histologis, dan radiologis [4]. Diagnosis dini dan strategi pencegahan sangat penting untuk mengelola UC secara efektif, seperti yang disoroti oleh berbagai studi klinis [5]. Meskipun ada kemajuan dalam teknik pencitraan, penggunaan CT dan MRI dalam mendiagnosis UC memiliki keterbatasan yang signifikan. Cara tersebut mungkin menunjukkan penebalan kolon, namun tidak cukup sensitif atau spesifik untuk menjadi alat diagnostik. Kesulitan dalam mengenali dan memberikan diagnosis yang tepat, disebabkan oleh gambaran klinis yang tidak selalu jelas, dan kurangnya parameter laboratorium yang spesifik untuk mendeteksi peradangan usus, termasuk *ulcerative colitis* [6]. Untuk mengatasi kekurangan tersebut, beberapa pendekatan otomatis telah dipublikasikan dalam literatur terbaru sebagai alternatif dengan mengimplementasikan *machine learning* untuk meningkatkan kinerja prediktif yang akan digunakan pada data ekspresi gen [7]. *Machine learning* memberikan potensi yang besar untuk mengenali pola-pola dalam dataset yang kompleks dan berdimensi tinggi, yang merupakan ciri khas dari profil ekspresi gen pada UC.

Sejumlah penelitian telah dilakukan dalam kaitannya dengan penerapan *machine learning* pada data ekspresi gen guna mengidentifikasi UC. Pada tahun 2020, Khorasani dan koleganya mengimplementasikan *developed feature selection algorithm* dan digabungkan dengan *Support Vector Machine* (SVM) yang akan membedakan antara subjek sehat dan subjek dengan KU berdasarkan ekspresi gen. Penelitian tersebut mendeteksi dengan sempurna semua kasus aktif dan memiliki presisi rata-rata 0.62 pada kasus tidak aktif [3]. Pada tahun 2022, Zhang Lin dan koleganya melakukan penilaian gen untuk memprediksi KU dengan menggunakan enam metode *machine learning*. Penilaian tersebut menunjukkan hasil SVM lebih baik dalam memprediksi KU, dengan *error rate* 0.16% [8]. Pada tahun 2021, Popa dan timnya mengembangkan model *Neural Network* untuk memprediksi aktivitas *ulcerative colitis* 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 2022, Minchun Bu dan rekannya menggunakan kombinasi Regresi Lasso dan SVM-RFE untuk memprediksi diagnosis UC menggunakan biomarker yang diekspresikan secara berbeda. Mereka mencapai akurasi dan skor F1 sebesar 95,20% dan akurasi validasi sebesar 87,4% [10]. Pada tahun 2021, Huang dan rekannya menggunakan tiga metode *machine learning* untuk mendiagnosis penyembuhan mukosa pada pasien UC. Algoritma *k-nearest neighbors* memberikan hasil terbaik dengan akurasi 93,4% (sensitivitas 86,2%, spesifisitas 95,8%) [11].

Berdasarkan survei literatur, sedikit penelitian yang menggunakan seleksi fitur untuk mengatasi masalah dimensi dalam kasus *ulcerative colitis*. Karakteristik dari ekspresi gen adalah tingginya jumlah dimensi data, sehingga seleksi fitur menjadi penting untuk mengurangi dimensi tersebut. Pemilihan fitur ini dapat ditingkatkan dengan menggunakan metode *meta-heuristik*, dan salah satu metode yang dapat digunakan adalah *Bat Algorithm* [12, 13].

Penelitian ini bertujuan untuk mendeteksi *ulcerative colitis* berbasis data ekspresi gen menggunakan metode *Bat Algorithm-Ensemble*. *Bat Algorithm* merupakan salah satu algoritma optimasi yang dapat digunakan untuk seleksi fitur. Algoritma ini dapat membantu dalam menangani masalah data dengan dimensi tinggi dengan memilih *subset* fitur yang paling relevan, hal ini akan mengurangi kompleksitas perhitungan dan mempercepat proses analisis. Metode *ensemble* merupakan salah satu metode *machine learning* yang digunakan untuk meningkatkan akurasi prediksi. Algoritma ini memberikan keuntungan, diantaranya dapat mengatasi risiko *overfitting*, dan menghasilkan model yang mudah diinterpretasi.

Topik dan Batasannya

Permasalahan yang diangkat pada penelitian ini membahas tiga permasalahan utama. Pertama, bagaimana efektivitas seleksi fitur dengan menggunakan metode *Bat Algorithm*. Kedua, bagaimana efektivitas optimasi *hyperparameter tuning* menggunakan metode *ensemble*. Ketiga, Bagaimana hasil performa deteksi menggunakan metode *Bat Algorithm* dan metode *ensemble*. Ruang lingkup penelitian dalam tugas akhir ini terfokus pada deteksi penyakit *ulcerative colitis* dengan memanfaatkan data ekspresi gen, yang akan dianalisis dengan penerapan metode *Bat Algorithm* dan *Ensemble*.

Tujuan

Tujuan dari penelitian ini adalah pertama, Menerapkan metode *Bat Algorithm* untuk melakukan seleksi fitur. Kedua, Menerapkan metode *ensemble* untuk menentukan optimasi *hyperparameter tuning*. Ketiga, Mengetahui hasil performa deteksi menggunakan metode *Bat Algorithm* dan metode *ensemble*.

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.