

BAB I

PENDAHULUAN

Leukemia adalah jenis kanker darah yang disebabkan oleh pertumbuhan sel darah putih yang tidak terkendali di sumsum tulang [1]. Penyakit ini dapat menyerang siapa saja, baik anak-anak maupun dewasa, dan dianggap mematikan jika tidak terdeteksi dini. Menurut Organisasi Kesehatan Dunia (WHO), leukemia termasuk dalam sepuluh jenis kanker paling mematikan di dunia [2]. Di Indonesia, berdasarkan data Globocan 2020, leukemia menempati peringkat ke-8 dengan lebih dari 11.000 kasus baru setiap tahun [3]. Oleh karena itu, upaya deteksi dini yang cepat dan akurat sangat penting untuk mengurangi angka kematian dan meningkatkan efektivitas pengobatan. Deteksi leukemia saat ini sangat bergantung pada metode laboratorium seperti morfologi sel darah dengan mikroskop, sitogenetika, dan imunofenotiping, yang seringkali subjektif, invasif, dan memerlukan biaya serta waktu yang signifikan [4, 5]. Di sisi lain, keterbatasan ketersediaan spesialis, terutama di daerah terpencil, menjadi tantangan besar dalam menetapkan diagnosis dini. Selain itu, kompleksitas data medis yang melibatkan berbagai parameter hematologi dan gambar mikroskop seringkali mempersulit proses analisis manual [6].

Kemajuan dalam kecerdasan buatan (AI), terutama machine learning (ML), telah membuka peluang signifikan dalam pengembangan sistem prediksi berbasis data untuk aplikasi medis. Berbagai algoritma telah diterapkan untuk mendeteksi dan memprediksi penyakit, termasuk leukemia. Dua algoritma yang sering digunakan dalam klasifikasi penyakit adalah Random Forest (RF) dan Logistic Regression (LR) [7, 8]. Random Forest dikenal karena kemampuannya menangani dataset besar dan kompleks dengan kinerja tinggi, sementara Logistic Regression menawarkan interpretabilitas model yang baik, terutama untuk data linear [9].

Meskipun kedua metode telah digunakan dalam berbagai studi, perbandingan langsung antara kinerja RF dan LR dalam prediksi leukemia masih terbatas. Beberapa studi telah menunjukkan bahwa RF memiliki akurasi tinggi dalam mendeteksi leukemia berdasarkan fitur hematologi atau gambar mikroskopis, dengan akurasi mencapai lebih dari 95% [10, 11]. Dengan pendekatan hibrida, Narayanan dkk. (2025) mengembangkan sistem untuk mengidentifikasi leukemia akut yang menggabungkan Random Forest (RF) untuk klasifikasi dan Fuzzy C-Means (FCM) untuk segmentasi gambar. Model mereka mencapai akurasi

99,06%, sensitivitas 99,4%, dan spesifisitas 97,8% dengan memproses dataset sekitar 8. 637 gambar mikroskopis. Kemampuan RF untuk mengelola kompleksitas gambar dan interaksi fitur secara efektif didukung oleh temuan ini [12]. Menggunakan gambar mikroskopis, Khan dkk. (2021) meneliti penerapan metode pembelajaran mesin untuk klasifikasi leukemia. Menurut studi oleh Fernández-Delgado dkk. yang diterbitkan di BMC Bioinformatics pada 2018, model Random Forest (RF) outperformed model Logistic Regression (LR) pada 243 dataset aktual, menunjukkan potensinya untuk digunakan dalam pengaturan klinis [13]. Mereka menemukan bahwa model berbasis Random Forest menawarkan akurasi yang lebih tinggi daripada model konvensional, menunjukkan potensinya yang menjanjikan dalam aplikasi klinis. Hasil benchmark menunjukkan bahwa RF unggul dalam sekitar 69% situasi. Skor Brier berkurang sebesar -0.027, AUC meningkat sebesar +0.041, dan selisih akurasi rata-rata sebesar +0.029, semua menunjukkan bahwa RF berkinerja jauh lebih baik daripada model lain [14]. Namun, Logistic Regression masih dapat diterapkan dalam beberapa skenario, terutama ketika digunakan pada dataset yang lebih kecil dan lebih bersih [15]. Mahmood & Kadir (2025) menggunakan beberapa variasi Logistic Regression dengan regularisasi Ridge, Lasso, dan khususnya ElasticNet pada data ekspresi gen (16.383 gen, 281 sampel, tujuh subtype leukemia). Dengan akurasi yang lebih tinggi, AUC, dan kemampuan seleksi gen yang lebih efektif, model ElasticNet terbukti menjadi yang terbaik. Khususnya pada data berdimensi tinggi, hasil ini menyarankan penggunaan Logistic Regression yang kuat dan dapat diinterpretasikan [16]. Namun, untuk menilai kinerja keduanya secara langsung dalam kasus prediksi leukemia berdasarkan gambar sel darah mikroskopis, studi perbandingan yang lebih komprehensif masih diperlukan secara keseluruhan.

Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan membandingkan kinerja metode Random Forest dan Logistic Regression dalam mendeteksi leukemia berdasarkan dataset medis yang tersedia. Evaluasi akan dilakukan menggunakan berbagai metrik kinerja seperti akurasi, sensitivitas, spesifisitas, dan AUC (Area Under Curve). Hasil penelitian ini diharapkan dapat menjadi dasar untuk pengembangan sistem pendukung keputusan medis yang dapat membantu dalam diagnosis leukemia yang dini, cepat, dan akurat, terutama di daerah dengan sumber daya terbatas.