

BAB I

PENDAHULUAN

1.1. Latar Belakang

Acute Lymphoblastic Leukemia (ALL) adalah penyakit kanker darah yang ditandai oleh penyebaran sel limfoblas abnormal yang akhirnya mengarah ke akumulasi jumlah sel leukemia yang mematikan [1]. Selama bertahun-tahun, melalui perbaikan dalam biologi molekuler, kemajuan substansial telah dicapai dalam memahami *Acute Lymphoblastic Leukemia* (ALL) [2]. *Acute Lymphoblastic Leukemia* merupakan kanker dengan kasus terbanyak pada anak-anak di Indonesia. Total kasus ALL di Indonesia mencapai 2.5 hingga 4.0 dari 100,000 anak dengan perkiraan jumlah kasus 2,000-3,200 per tahun [3]. Oleh karena itu, ALL dapat dijadikan tujuan awal yang logis dalam perkembangan program kanker anak di Indonesia.

Acute Lymphoblastic Leukemia (ALL) dapat ditandai dengan kelelahan, sesak napas, demam, mudah mengalami perdarahan dan memar, berkeringat di malam hari, nyeri tulang dan sendi, dan penurunan berat badan yang tidak dapat dijelaskan [4]. Diagnosis ALL menggunakan citra *peripheral blood smear* (PBS) memainkan peran penting dalam penyaring awal kasus kanker dari non-kanker [5]. Hasil dari diagnosis tersebut terbagi menjadi dua yaitu kanker jinak (*benign*) dan kanker ganas (*malignant*). Kanker ganas sendiri dibagi lagi menjadi beberapa tipe yaitu, *early*, *pre*, dan *pro* [6].

Acute Lymphoblastic Leukemia (ALL) dapat diobati apabila telah dilakukan diagnosis pada tahap awal. Diagnosis biasanya dilakukan dengan perhitungan darah secara lengkap dan analisis morfologi sel di bawah mikroskop oleh seorang ahli medis. Flow cytometry dapat menggantikan pekerjaan manual ini, tetapi membutuhkan peralatan mahal yang tidak tersedia di semua tempat. Oleh karena itu, pembuatan sistem otomatis yang dapat melakukan diagnosis menggunakan citra mikroskop yang relatif murah dapat memberikan keuntungan besar [7]. Sistem otomatis untuk diagnosis menggunakan citra dapat dibuat menggunakan metode CNN, karena memiliki convolution layer yang dapat mengurangi dimensi dari citra tanpa menghilangkan informasi citra.

Pada tahun 2022, Mustafa Ghaderzadeh dkk melakukan penelitian mengenai klasifikasi *Acute Lymphoblastic Leukemia* (ALL) menggunakan metode CNN dengan 10 arsitektur berbeda (MobileNetV3, Xception, InceptionV3, VGG-16, ResNet50V2, NASNetLarge, VGG-19, InceptionResNetV2, EfficientNet, dan DenseNet201) yang digunakan untuk ekstraksi fitur pada data. Dari 10 model tersebut, model yang mendapat nilai performa tertinggi adalah DenseNet201 dengan nilai akurasi, sensitivitas, dan spesifisitas masing-masing 99,85, 99,52, dan 99,89% [8].

Pada penelitian ini, metode klasifikasi yang akan digunakan adalah CNN dengan arsitektur GoogleNet dan metode Deep SMOTE untuk mengatasi imbalance data. Dataset yang digunakan pada penelitian ini adalah data dari kaggle yang berisi citra *Peripheral Blood Smear* (PBS) dari Laboratorium sumsum tulang Rumah Sakit Taleqani (Teheran, Iran). Dalam dataset tersebut, terdapat citra asli dan citra yang telah disegmentasi oleh staf ahli laboratorium. Kelas yang terdapat dalam dataset tersebut antara lain, *benign*, *early*, *pre* dan *pro* dengan jumlah citra setiap kelas berturut-turut 504, 985, 963, 804.

1.2. Rumusan Masalah

Berdasar dari latar belakang penelitian, rumusan masalah yang didapatkan adalah sebagai berikut:

1. Bagaimana mengatasi *imbalance data* menggunakan metode Deep SMOTE?
2. Bagaimana membangun model klasifikasi citra menggunakan CNN dengan arsitektur GoogleNet?
3. Bagaimana perbandingan akurasi antara model klasifikasi yang dilatih menggunakan dataset yang telah *oversampling* menggunakan Deep SMOTE dengan dataset tanpa *oversampling*?

1.3. Batasan Masalah

1. Difokuskan menggunakan dataset dari citra *Peripheral Blood Smear* (PBS) dari Laboratorium sumsum tulang Rumah Sakit Taleqani.
2. Uji coba dataset menggunakan 4 kelas yang terdiri dari *Benign*, *Early*, *Pre*, dan *Pro*.
3. Penelitian ini menggunakan jumlah data yaitu 6512 yang bersumber dari : <https://www.kaggle.com/datasets/mehradaria/leukemia>

1.4. Tujuan Penelitian

1. Mengatasi *imbalance data* menggunakan metode Deep SMOTE
2. Membuat sebuah model klasifikasi menggunakan CNN dengan arsitektur GoogleNet untuk diagnosis *Acute Lymphoblastic Leukemia* (ALL) pada pasien.
3. Mengetahui perbandingan akurasi antara model klasifikasi yang dilatih menggunakan dataset yang telah di-*oversampling* menggunakan Deep SMOTE dengan dataset tanpa *oversampling*.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah dapat memberikan kemudahan terutama dalam bidang Kesehatan untuk mendiagnosis *Acute Lymphoblastic Leukemia* (ALL) secara cepat dan akurat.

1.6. Sistematika Penulisan

Penulisan tugas akhir ini terdiri dari 3 (tiga) bab dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan tentang hal umum dari penelitian tugas akhir, yaitu latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan tentang teori yang berhubungan dengan penelitian yang terdiri dari penjelasan teori darah, teori diagnosis penyakit leukemia, teori sel Leukemia Limfoblastik Akut, teori citra digital, dan *Convolutional Neural Networks* dengan arsitektur Googlenet.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tahap-tahap penelitian tugas akhir yang dilakukan. Mulai dari perumusan masalah, kajian pustaka, pengumpulan data, analisa data, perancangan sistem, implementasi, pengujian, serta kesimpulan dan saran.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil dan pembahasan implementasi dari metode yang diusulkan dan analisis terhadap hasil implementasi.

BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bab yang memuat kesimpulan dari hasil penelitian yang dilakukan serta saran untuk pengembangan pada penelitian selanjutnya berdasarkan hasil analisis yang diperoleh pada bab IV.