

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Leukemia adalah jenis penyakit kanker yang menyerang sel darah putih dan berasal dari sumsum tulang. Penyakit ini ditandai dengan adanya sel-sel abnormal dalam darah yang menyebar secara luas. Leukemia akut pada anak-anak merupakan jenis kanker yang paling umum dan mencakup sekitar 30-40% dari seluruh kasus kanker pada anak-anak di bawah usia 15 tahun [1]. Menurut *World Health Organization* (WHO), pada tahun 2020 terdapat sekitar 474.519 kasus baru leukemia di seluruh dunia. Sementara itu, angka kematian akibat leukemia mencapai sekitar 311.594 pada tahun yang sama. Jumlah ini meliputi semua jenis leukemia, baik yang akut maupun kronis, dan semua kelompok usia. Dalam klasifikasi leukemia, terdapat empat kategori besar yang didasarkan pada tingkat kematangan sel dan asal sel, yaitu *Acute Lymphoblastic Leukemia* (ALL), *Acute Myeloid Leukemia* (AML), *Chronic Lymphocytic Leukemia* (CLL), dan *Chronic Granulocytic Leukemia* (CGL) [2]. Leukemia memiliki tingkat kesembuhan yang tinggi melalui deteksi dini, namun memiliki tingkat kematian yang tinggi akibat terlambat didiagnosis. Karena termasuk dalam jenis tumor cair, leukemia sulit terdeteksi secara fisik, dan gejala-gejala umumnya tidak terlihat hingga dilakukan tes darah yang tepat. Tes darah melalui citra mikroskopis adalah salah satu cara untuk melihat sel-sel abnormal pada leukemia [3].

Acute Lymphoblastic Leukemia (ALL) adalah salah satu jenis kanker darah yang umumnya ditemukan pada anak-anak. Lebih dari 80% penderita ALL berusia 1-4 tahun, dan keganasannya berkembang dengan cepat. Jika tidak segera diobati, kondisi penderita dapat memburuk dalam beberapa bulan dan berakibat fatal. Oleh karena itu, deteksi dini sangat penting untuk memastikan langkah pengobatan yang lebih efektif [4].

Diagnosis leukemia berdasarkan analisis citra sel-sel mikroskopis secara manual memerlukan ahli yang terampil dalam bidangnya, dan hal ini dapat memakan waktu lama dan meningkatkan risiko kesalahan [3]. Berbagai kemajuan di bidang pengolahan citra medis dan sistem diagnosis yang terbantu komputer diharapkan dapat mengurangi pengaruh subjektivitas dan membantu dokter dalam proses diagnosis [4]. Teknik *deep*

learning dapat digunakan untuk melakukan klasifikasi pada dataset besar. *Deep learning* adalah sebuah algoritma jaringan saraf tiruan yang mengambil data acak sebagai *input*, kemudian memprosesnya melalui beberapa lapisan tersembunyi (*hidden layer*) untuk melakukan transformasi non-linier dari data tersebut dan menghitung nilai *output* [5].

Penelitian terkait yang pernah dilakukan pada tahun 2018, Moh Hamim Zajuli Al Faroby melakukan sebuah studi yang berkaitan dengan pengelompokkan jenis penyakit Leukemia menggunakan metode *Support Vector Machine* (SVM) yang merupakan salah satu jenis pembelajaran terawasi (*supervised learning*). SVM mengelompokkan data penyakit leukemia ke dalam empat kategori, yaitu *Acute Lymphocytic Leukemia* (ALL), *Acute Myelogenous Leukemia* (AML), *Chronic Lymphocytic Leukemia* (CLL), dan *Chronic Myelogenous Leukemia* (CML). Ada dua jenis data yang digunakan dalam metode ini, yaitu data *training* sebagai acuan dan data *testing* sebagai pengujian hasil metode. SVM menggunakan tiga jenis kernel, yaitu kernel linear, RBF, dan Polynomial, yang mempengaruhi performa metode. Dalam SVM, terdapat dua parameter, yaitu parameter C dan Gamma, yang berpengaruh signifikan pada hasil training dan prediksi data *testing*. Parameter C digunakan untuk membentuk vektor pendukung, dan semakin besar nilainya, semakin banyak data *training* yang akan menjadi vektor pendukung. Sementara itu, parameter Gamma digunakan untuk membentuk *hyperplane* pemisah antara kedua kelas. Semakin besar nilai Gamma, maka *hyperplane* yang terbentuk semakin presisi terhadap setiap vektor pendukung. Dalam studi ini, kernel linear memberikan hasil terbaik pada rentang nilai C antara 100 sampai 1000 dengan memprediksi 24 data tes dengan benar. Kernel RBF memberikan hasil terbaik pada rentang nilai C antara 20 sampai 200 dan rentang nilai Gamma antara 10 hingga 200 dengan memprediksi 25 data tes dengan benar. Sedangkan, kernel Polynomial memberikan hasil terbaik pada nilai C sebesar 0.5 dan rentang nilai Gamma antara 20 sampai 100 dengan memprediksi 25 data tes dengan benar. [6].

Penelitian lain terkait deteksi penyakit Leukemia pernah dilakukan oleh M. Dzikru Suratin et al pada tahun 2015. Dalam penelitian ini, dilakukan identifikasi sel *Acute Lymphoblastic Leukemia* (ALL) dengan menggunakan ciri-ciri morfologi. Meskipun

segmentasi berdasarkan warna dapat memisahkan komponen dalam citra sediaan apus darah, perlu diperhatikan bahwa kualitas pewarnaan slide dapat mempengaruhi proses tersebut. Untuk menghindari keadaan *over-segmentation*, metode pengaplikasian median filter, *K-means clustering*, dan operasi morfologi dilakukan sebelum *watershed transform*. Metode *Support Vector Machine* menunjukkan hasil yang menjanjikan dalam klasifikasi sel darah putih normal dan sel ALL, dengan akurasi pengenalan sebesar 95,45% untuk semua citra uji. Untuk pengembangan lebih lanjut, perlu dilakukan pengukuran terhadap obyek sitoplasma dan klasifikasi subtype dari tipe ALL [4].

Pada tahun 2022, Abdul Mahatir Najar et al juga melakukan penelitian untuk mengidentifikasi jenis Leukemia. Penelitian ini menggunakan lima metode *machine learning* yang sering digunakan untuk klasifikasi, yaitu *Support Vector Machine* (SVM), KNN Classifier, *Multilayer Perceptron Classifier*, *Decision Tree Classifier*, dan *Gaussian Naive Bayes*. Hasilnya menunjukkan bahwa *Decision Tree Classifier* memberikan akurasi terbaik sebesar 0,833, yang berarti tingkat kebenarannya mencapai 83,3%. Agar proses identifikasi kanker menjadi lebih akurat, untuk penelitian selanjutnya, kolaborasi dengan peneliti dari bidang lain seperti biologi, kimia, atau farmasi perlu dilakukan [7].

Penelitian lainnya dilakukan oleh Ahmed T. Shalol, et al. pada tahun 2020. Berdasarkan penelitian tentang klasifikasi penyakit leukemia menggunakan *deep learning* metode CNN dengan arsitektur VGG-19 menggunakan dataset ALL-IDB2 menghasilkan nilai akurasi sebesar 96.11% dengan nilai *F1 score* sebesar 97.9%. akurasi dengan arsitektur VGG-19 yang didapat lebih tinggi dari arsitektur lainnya seperti ResNet-18 dan ResNext dengan dataset yang sama [8].

Kemudian, untuk penelitian terkait lainnya dilakukan pada tahun 2022 yang dilakukan oleh Fajar Sarasati et al. Berdasarkan hasil penelitian, berhasil diklasifikasikan penyakit pada tanaman jagung dengan menggunakan metode *deep learning* yang mengadopsi arsitektur EfficientNet-B0. Metode ini menghasilkan akurasi terbaik sebesar 96% dan loss sebesar 0,12, serta val loss sebesar 0,13 dan val accuracy sebesar 95%. Oleh karena itu, setiap kelas dapat terprediksi dengan benar

dengan tingkat akurasi di atas 90%, dan evaluasi menunjukkan bahwa plot grafik good fit [9].

Penelitian lainnya juga dilakukan pada tahun 2022 yang dilakukan oleh Kasuma Indra Suryawan. Berdasarkan hasil eksperimen menggunakan arsitektur EfficientNet-B0, ditemukan bahwa model terbaik diperoleh dengan membagi data menjadi 90% data untuk pelatihan dan 10% data untuk validasi menggunakan *optimizer* AdaGrad, batch size sebesar 128, dan fungsi aktivasi Leaky ReLU. Meskipun model yang dihasilkan lebih sederhana dalam hal komputasi, namun model tersebut berhasil mencapai tingkat akurasi sebesar 94%. Hasil dari penelitian pendahuluan sebelumnya hanya mencapai akurasi 82,1%, sedangkan penelitian sebelumnya yang menggunakan arsitektur AlexNet mencapai tingkat akurasi 93% namun memiliki kompleksitas yang lebih tinggi [10].

Pada penelitian ini dilakukan uji coba perbandingan penggunaan nilai *learning rate* dan nilai *epoch* untuk mendapatkan nilai akurasi tertinggi menggunakan metode *Convolutional Neural Network* (CNN) dengan menggunakan arsitektur EfficientNet-B0 untuk mengklasifikasikan penyakit *Acute Lymphoblastic Leukemia* (ALL).

1.2 Rumusan Masalah

1.2.1 Permasalahan

Bagaimana pengaruh nilai *learning rate* dan nilai *epoch* untuk menghasilkan nilai akurasi dalam klasifikasi mengklasifikasikan penyakit *Acute Lymphoblastic Leukemia*?

1.2.2 Solusi Permasalahan

Solusi yang diusulkan dalam penelitian ini adalah mengimplementasikan model berbasis *deep learning* menggunakan arsitektur EfficientNet-B0 dengan melakukan uji coba skenario terhadap nilai *learning rate* dan *epoch* untuk mengetahui pengaruh nilai *learning rate* dan nilai *epoch* terhadap akurasi yang dihasilkan.

1.2.3 Pertanyaan Penelitian

1. Bagaimana pengaruh perubahan nilai *learning rate* dan *Epoch* dalam mengklasifikasikan penyakit kanker darah jenis *Acute Lymphoblastic Leukemia* (ALL)?
2. Berapa akurasi yang didapat dari model terbaik arsitektur EfficientNet-B0 dalam mengklasifikasikan kanker darah jenis *Acute Lymphoblastic Leukemia* (ALL)?

1.3 Batasan Masalah

1. Klasifikasi citra *Acute Lymphoblastic Leukemia* (ALL) pada penelitian ini dilakukan dengan 2 kelas target yaitu positif ALL dan normal.
2. Dataset yang digunakan adalah citra C-NMC 2019 *dataset* yang berisi data citra mikroskopis darah putih sebanyak 12.528 data citra.
3. Nilai *learning Rate* yang digunakan yaitu 0.01, 0.001, dan 0.0001[11]
4. Nilai *Epoch* yang digunakan yaitu 30, 40, 50 *Epoch*
5. *Optimizer* yang digunakan adalah Adamax [11].

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

1. Untuk mengetahui pengaruh nilai *learning rate* dan *epoch* dalam klasifikasi penyakit *Acute Lymphoblastic Leukemia* (ALL) menggunakan arsitektur EfficientNet-B0.
2. Untuk memperoleh nilai akurasi dari model terbaik klasifikasi penyakit *Acute Lymphoblastic Leukemia* (ALL) menggunakan arsitektur EfficientNet-B0.

1.4.2 Manfaat

Implementasi arsitektur EfficientNet-B0 pada penyakit ALL untuk menghasilkan klasifikasi yang baik, sehingga dapat membantu proses deteksi penyakit melalui citra sel darah putih yang terinfeksi ALL.

1.5 Sistematika Proposal

Sistematika penulisan proposal Tugas Akhir ini meliputi sebagai berikut:

- **BAB I PENDAHULUAN**

Pada bab ini dijelaskan mengenai gambaran umum dan batasan masalah, usulan solusi untuk mengatasi permasalahan, tujuan dan manfaat dari solusi tersebut, serta struktur keseluruhan proposal.

- **BAB II KAJIAN PUSTAKA**

Pada bab ini disajikan pembahasan teori terkait citra, *Convolutional Neural Network*, model arsitektur EfficientNet serta penelitian terkait yang telah dilakukan sebelumnya.

- **BAB III METODE USULAN**

Pada bab ini disajikan pembahasan tentang dataset yang digunakan dalam penelitian, arsitektur sistem training dan testing serta skenario uji coba.

- **BAB IV HASIL DAN PEMBAHASAN**

Bab ini menjelaskan mengenai lingkungan uji coba, tahapan pembuatan program, dan hasil uji coba.

- **BAB V PENUTUP**

Bab ini berisi kesimpulan dan saran berdasarkan penelitian yang telah dilakukan.