

## BAB 1 PENDAHULUAN

### 1.1. Latar Belakang

*Acute Lymphoblastic Leukemia (ALL)* merupakan salah satu jenis penyakit kanker darah yang umum terjadi pada anak-anak dengan rentang usia tiga hingga lima tahun, akan tetapi juga dapat menyerang orang dewasa [1]. Menurut *update* terakhir *American Cancer Society*, perkiraan jumlah kasus baru di Amerika Serikat (AS) pada tahun 2022 tercatat sekitar 6.660 kasus termasuk anak-anak dan dewasa dan 1.560 telah terkonfirmasi mengalami kasus kematian akibat ALL [2]. Diantara semua kasus, faktanya sekitar 60% ALL terjadi pada anak-anak dengan rentang usia sebelum usia 15 tahun dan setelah 50 tahun pada orang dewasa [3]. ALL terjadi ketika sumsum darah mulai memproduksi limfoblas, yaitu sel-sel putih yang tidak sepenuhnya terbentuk. Sel abnormal ini memadati sel-sel darah merah dan putih yang sehat serta trombosit darah dan sumsum tulang sehingga sulit bagi tubuh untuk melawan infeksi dan penyakit. Jumlah sel abnormal (leukemia) tumbuh dengan cepat dan menyebar ke organ dan jaringan lain seperti otak, hati, ginjal, dan limpa [4][5].

Mengenai faktor risiko dan penyebab seseorang dapat terkena ALL belum diketahui secara pasti, namun terdapat beberapa faktor risiko yang dapat meningkatkan peluang seseorang terkena ALL yaitu paparan radiasi, paparan zat kimia, infeksi virus, sindrom genetik, usia, jenis kelamin, dan memiliki kembar identik yang terkena ALL dapat meningkatkan risiko terkena ALL. Faktor lainnya yang juga memungkinkan adalah paparan medan elektromagnetik (seperti tinggal di dekat saluran listrik atau menggunakan ponsel), paparan diesel, bensin, pestisida di tempat kerja, dan bahan kimia tertentu lainnya, dan rokok. Namun sejauh ini, tidak satu pun dari faktor-faktor ini yang dapat dikaitkan secara pasti dengan ALL, tetapi penelitian di bidang ini masih terus berlanjut [6]. Seseorang yang terkena ALL akan mengalami gejala-gejala tertentu seperti kelelahan ekstim pada minggu pertama, kemudian terjadi anemia akibat sel darah merah tidak diproduksi karena pusat tulang yang dipenuhi dengan sel-sel leukemik, perdarahan dan/atau memar, nyeri pada tulang sendi, demam atau infeksi, nyeri perut, pembekakan kelenjar getah bening, dan kesulitan bernafas [7].

Tanda dan gejala tertentu yang muncul memang dapat menunjukkan bahwa seseorang mungkin menderita ALL, tetapi tes diperlukan untuk mengkonfirmasi diagnosis. Diagnosis ALL

yang dilakukan sejak dini sangat penting untuk pemulihan dan perawatan pasien. Diagnosis ALL umumnya dilakukan dengan menggunakan tes hitung darah lengkap atau *complete blood count* (CBC) untuk mengukur jumlah komponen sel darah merah, sel darah putih, dan trombosit. Selain itu dilakukan juga analisis morfologis sel di bawah mikroskop oleh seorang ahli medis untuk mendapatkan gambaran detail mengenai bentuk sel darah. Sebagian besar pasien dengan ALL memiliki jumlah eritrosit dan bentuk sel leukosit yang tidak normal [8][9]. Akan tetapi proses identifikasi yang dilakukan masih tergolong tradisional sehingga memakan waktu yang lama dalam prosesnya dan membutuhkan biaya yang tidak sedikit serta rawan kesalahan[1][10]. Darah yang diamati dapat didiagnosis untuk berbagai penyakit oleh ahli medis. Diagnosis berbasis manusiawi memiliki presisi yang tidak standar karena pada dasarnya tergantung pada keterampilan ahli hematologi, juga tidak dapat diandalkan dalam sudut pandang statistik[11]. Oleh karena itu, pendekatan *artificial intelligence* untuk proses deteksi ALL pada citra mikroskopis sel darah akan menjadi hal yang sangat dibutuhkan. *Deep learning* merupakan pengembangan dari *artificial intelligence* pada *machine learning* yang algoritmanya meniru cara kerja otak manusia dan terdiri dari banyak *layer* pemrosesan [12].

Penelitian mengenai klasifikasi menggunakan metode *deep learning* telah banyak dilakukan karena dinilai memiliki performa kinerja yang lebih baik dibandingkan dengan metode *machine learning*. Seperti penelitian yang dilakukan pada tahun 2019, Ahmed, *et al* [13] melakukan penelitian untuk mendiagnosis semua sub tipe leukemia dari citra mikroskopis menggunakan *Convolutional Neural Network* (CNN) dan metode *machine learning*. Pada penelitian ini dilakukan enam skenario uji coba yaitu dua eksperimen menggunakan model CNN dengan nilai *epoch* 25 dan 100, sedangkan empat eksperimen lainnya menggunakan metode *machine learning* yaitu Support Vector Machine (SVM), Naïve Bayes (NB), k-NN ( $k = 3$ ), dan Decision Tree (DT) dengan menggunakan *5-fold cross validation* untuk semua eksperimen. Diakhir, dilakukan perbandingan semua eksperimen yang menggunakan metode CNN dengan metode *machine learning* dan menghasilkan bahwa metode CNN memiliki performa yang lebih baik daripada metode *machine learning* [13].

Namun, dalam mengimplementasikan metode *deep learning* membutuhkan jumlah pemakaian memory yang cukup besar untuk menyimpan *weight* yang akan digunakan dalam proses *training* atau disebut dengan *trainable* parameter [14]. Oleh karena itu, untuk mengatasi permasalahan tersebut digunakan metode *deep learning* dengan kombinasi aljabar *hypercomplex*. Seperti

penelitian yang telah dilakukan pada tahun 2022, Vieira, *et al* [15] melakukan penelitian untuk mengidentifikasi penyakit ALL pada citra mikroskopis menggunakan *Hypercomplex-valued* CNN. Model yang dikembangkan merupakan kombinasi arsitektur CNN dengan aljabar *hypercomplex* sehingga disebut dengan *hypercomplex-valued* CNN. Aljabar *hypercomplex* diterapkan pada *convolutional layer* untuk mengurangi jumlah parameter yang digunakan. Skenario uji coba dilakukan dengan membandingkan arsitektur CNN asli atau disebut dengan *real-CNN* dengan *hypercomplex-valued* CNN yang terdiri dari 8 *hypercomplex-valued* CNN dengan RGB, 8 *hypercomplex-valued* CNN dengan HSV, dan 1 arsitektur *real-CNN* dengan RGB dan HSV sehingga total percobaan adalah 18. Pembagian data *training* dan *testing* dilakukan menggunakan teknik split dengan proporsi data 50:50. Dari 18 percobaan didapatkan rata-rata hasil akurasi *hypercomplex* adalah sebesar 96.6%. Akurasi tertinggi diperoleh dari percobaan aljabar *hypercomplex* co-quaternions HSV dan tessarines HSV [15]. Kombinasi CNN dengan *hypercomplex* dinilai dapat mengurangi jumlah *trainable parameter* yang digunakan, namun tetap menghasilkan nilai akurasi yang optimal atau serupa dengan *real-CNN*.

Pada proses *training* penyesuaian *hyperparameter* selalu menjadi hal yang penting karena berdampak terhadap kinerja model yang dihasilkan, salah satunya yaitu pemilihan nilai *hyperparameter learning rate*. *Learning rate* dapat menentukan seberapa laju pembelajaran saat mengupdate nilai *weight* untuk dapat dengan cepat menuju ke *minimum loss*[16]. Namun, apakah besarnya nilai *learning rate* yang digunakan sangat berpengaruh terhadap tingkat akurasi yang dihasilkan. Mengenai hal tersebut, dalam penelitian ini dilakukan uji-coba *learning rate* untuk mengetahui seberapa pengaruh besarnya nilai *learning rate* yang digunakan terhadap tingkat akurasi yang dihasilkan. Salah satu penelitian mengenai uji-coba *learning rate* telah dilakukan pada tahun 2022, Lasniari dkk melakukan penelitian untuk mengetahui pengaruh *hyperparameter* CNN arsitektur Resnet-50 pada klasifikasi citra daging sapi dan daging babi. *Hyperparameter* yang diuji-coba adalah *split* data *training* dan *testing* 70:30, 80:20, 90:10. *Batch-size* 16, 32, 64. *Learning rate* 0.1, 0.01, 0.001. *Epoch* 50, 75, 100. Penelitian tersebut menghasilkan nilai akurasi tertinggi 100% yang didapatkan pada uji-coba *split* data 90:10, *batch-size* 32, *epoch* 100, dan *learning rate* 0.001.

Pada penelitian ini dilakukan perbandingan hasil dari dua jenis metode klasifikasi yaitu klasifikasi menggunakan CNN atau disebut dengan *Real-CNN* dan klasifikasi menggunakan kombinasi CNN dengan aljabar *hypercomplex* atau disebut dengan *hypercomplex-valued* CNN.

Selain itu, dilakukan juga uji-coba scenario *learning rate* dengan nilai 0.1, 0.01, dan 0.001. Tujuan dari penelitian ini adalah untuk mengetahui seberapa pengaruh nilai *learning rate* yang digunakan terhadap tingkat akurasi yang dihasilkan serta mengetahui bagaimana pengaruh *hypercomplex-valued* jika diterapkan pada arsitektur CNN terhadap tingkat akurasi yang dihasilkan serta jumlah parameter yang digunakan.

## **1.2. Rumusan Masalah**

### **1.2.1. Permasalahan**

Bagaimana mengimplementasikan metode *deep learning* dengan penggunaan *trainable* parameter yang lebih sedikit serta bagaimana pengaruh nilai *learning rate* terhadap tingkat akurasi yang dihasilkan dalam proses klasifikasi penyakit ALL.

### **1.2.2. Solusi Permasalahan**

Solusi yang ditawarkan adalah dengan menggunakan *Hypercomplex-valued* CNN yang dilatih menggunakan dataset ALL untuk menghasilkan nilai akurasi yang optimal dengan jumlah *trainable* parameter yang digunakan lebih sedikit. Serta dilakukan uji-coba scenario terhadap nilai *learning rate* yang digunakan untuk mengetahui seberapa pengaruh nilai *learning rate* terhadap tingkat akurasi yang dihasilkan.

### **1.2.3. Pertanyaan Penelitian**

Pertanyaan penelitian digunakan untuk mengetahui apa yang akan dilakukan dalam penelitian ini sehingga dapat memperoleh hasil yang sesuai yaitu:

1. Bagaimana pengaruh nilai *learning rate* yang digunakan terhadap tingkat akurasi yang dihasilkan dalam proses klasifikasi menggunakan metode *Real-CNN*?
2. Bagaimana pengaruh nilai *learning rate* yang digunakan terhadap tingkat akurasi klasifikasi pada metode *Hypercomplex-valued* CNN?
3. Bagaimana pengaruh *Hypercomplex-valued* terhadap tingkat akurasi dan jumlah *trainable* parameter yang digunakan jika diterapkan dalam proses klasifikasi menggunakan CNN?

## **1.2. Batasan Masalah**

Batasan masalah digunakan untuk mengetahui batasan dari ruang lingkup penelitian yaitu:

1. Klasifikasi citra *Acute Lymphoblastic Leukemia* (ALL) dilakukan dengan 2 kelas target yaitu positif ALL dan negatif ALL.

2. Dataset yang digunakan berjumlah 260 data dan diperoleh dari Departemen Teknologi Informasi - Universitas degli Studi di Milano [17].
3. Uji coba dilakukan menggunakan *real*-CNN dan 4 jenis aljabar *hypercomplex* yaitu 2 aljabar *hypercomplex commutative* dan 2 aljabar *hypercomplex non-commutative*.
4. Jumlah *epoch* yang digunakan yaitu 100 [15].
5. Jumlah *dropout* yang digunakan yaitu 0.5 [15].
6. Jumlah *batch-size* yang digunakan yaitu 32 [15].
7. *Optimizer* yang digunakan ADAM [15].
8. Nilai *learning rate* yang digunakan untuk uji-coba yaitu 0.1, 0.01, 0.001 [18].

## 1.4. Tujuan dan Manfaat

### 1.4.1. Tujuan

Tujuan penelitian digunakan untuk menjawab pertanyaan peneliti yaitu:

1. Mengetahui pengaruh nilai *learning rate* terhadap tingkat akurasi yang dihasilkan pada proses klasifikasi menggunakan *Real*-CNN.
2. Mengetahui pengaruh nilai *learning rate* terhadap tingkat akurasi yang dihasilkan pada proses klasifikasi menggunakan *Hypercomplex-valued* CNN.
3. Mengetahui pengaruh *Hypercomplex-valued* terhadap tingkat akurasi dan jumlah *trainable* parameter yang digunakan jika diterapkan dalam proses klasifikasi citra *Acute Lymphoblastic Leukemia* (ALL) menggunakan CNN.

### 1.4.2. Manfaat

Dengan menerapkan model klasifikasi ini, dapat digunakan untuk membantu mempercepat proses diagnosis pasien ALL berdasarkan citra mikroskopis ALL dan dapat dijadikan sebagai bahan rujukan untuk dikembangkan selanjutnya.

## 1.5. Sistematika Tugas Akhir

Sistematika penulisan Tugas Akhir ini adalah sebagai berikut:

### ● BAB I PENDAHULUAN

Pada bab ini memaparkan mengenai latar belakang dari permasalahan, solusi yang diusulkan untuk menangani masalah tersebut, tujuan dan manfaat dari solusi yang diberikan, dan sistematika penulisan dalam tugas akhir ini.

- **BAB II KAJIAN PUSTAKA**

Bab ini memaparkan secara teoritis mengenai metode *Deep Learning Convolutional Neural Network* (CNN), aljabar *Hypercomplex-valued*, Arsitektur *Hypercomplex-valued Convolutional Neural Network* (HvCNN), metode perhitungan akurasi, serta penelitian terkait sebelumnya.

- **BAB III METODE PENELITIAN**

Bab ini menjelaskan mengenai *dataset* yang digunakan, arsitektur yang akan diterapkan, perhitungan manual, dan skenario *training* dan *testing* model.

- **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.