

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kanker adalah penyakit yang dimulai ketika sebuah sel di dalam tubuh mulai tumbuh secara tidak terkontrol. Ada banyak jenis kanker, dimana salah satunya adalah leukemia. Menurut *American Cancer Society* leukemia adalah sebuah penyakit kanker darah yang dimulai dari sel yang berkembang menjadi berbagai tipe sel darah, dan sebagian besar adalah sel darah putih[1]. Ada beberapa jenis leukemia, dibagi menurut kondisinya yaitu akut dan kronis, dan dimana leukemia tersebut terjadi, yaitu sel *myeloid* dan sel *lymphoid*. Penyakit leukemia yang akan diklasifikasi pada penelitian ini adalah penyakit leukemia jenis Acute Lymphocytic Leukemia (ALL) atau biasa juga disebut Acute Lymphoblastic Leukemia. Sesuai namanya, penyakit leukemia ini bersifat akut, yang berarti memiliki perkembangan yang pesat, dan berkembang dari bentuk awal dari *lymphocytes* yang merupakan salah satu jenis sel darah putih. Jumlah sel darah putih abnormal yang banyak dan jumlah sel darah lain yang tidak cukup mengakibatkan terserangnya sistem kekebalan tubuh, sehingga menyebabkan penderita merasa lemas, demam tinggi, mudah untuk berdarah dan lainnya. Penyakit ini memiliki resiko untuk berkembang paling tinggi pada umur dibawah 5 tahun dan setelah umur 50 tahun, dengan umur diantara tersebut, resiko terjangkit menurun. Jumlah anak-anak yang terjangkit penyakit ini di Indonesia masih belum banyak diketahui [2]. Dengan jumlah 79.5 juta anak yang hidup di Indonesia pada tahun 2018, diperkirakan terdapat 3434 kasus penyakit ALL yang baru pada anak.

Berdasarkan penelitian oleh Jose Elwyslan Mauricio de Oliveira dan Daniel Oliveira Dantas[3] cara diagnosa yang dilakukan sekarang masih menggunakan cara manual, salah satunya adalah tes darah dimana sample darah

akan diambil dan dilihat melalui mikroskop. Setelah itu ahli onkologi akan melakukan analisis hasil dari mikroskop tersebut dan menentukan apakah pasien yang diperiksa darahnya memiliki kanker atau tidak. Cara ini tentu saja memakan waktu lama dan tingkat keakuratan sangat bergantung dari keahlian orang tersebut. Tidak hanya itu, kesalahan diagnosis juga bisa terjadi ketika ahli tersebut terlalu lelah dalam melakukan diagnosis. Maka dari itu perlu sebuah teknologi yang dapat membantu dalam pengklasifikasian tersebut, yaitu sebuah *deep learning* yang sudah dilatih untuk mengklasifikasikan sel kanker tersebut.

Penelitian yang menggunakan *deep learning* untuk mengklasifikasikan kanker leukemia dengan jenis ALL telah banyak dilakukan sebelumnya. Beberapa penelitian tersebut seperti yang dilakukan oleh Amreen Batool dan Yung-Cheol Byun pada penelitian mereka yang menggunakan EfficientNetB3 yang sudah diubah menggunakan basis *Depthwise Separable Convolutions* agar lebih ringan[4], dan penelitian yang dilakukan oleh Niranjana Sampathila, dkk. dimana mereka membuat model CNN mereka sendiri yang dikhususkan untuk melakukan klasifikasi terhadap penyakit ALL[5]. Akan tetapi, penelitian-penelitian tersebut memiliki kekurangan dimana mereka tidak punya cara untuk memberitahukan kepada penggunanya, alasan model tersebut menilai bahwa suatu gambar adalah kelas ALL atau HEM.

Pada dasarnya model berbasis neural networks seperti CNN, adalah model yang dimana hasil prediksinya seperti kotak hitam. Dimana pengguna model tersebut hanya mengetahui seberapa bagus performa model tersebut dalam mempelajari dan diuji dengan dataset yang disediakan. Akan tetapi tidak mengetahui proses yang terjadi selama pelatihan dan pengujian tersebut, fitur mana saja yang diambil yang dipakai dalam proses klasifikasi dikarenakan model yang sudah sangat kompleks. Pada dunia medis, penanganan pasien harus sangat hati-hati agar tidak salah mengambil keputusan. Maka dari itu transparansi dari segala alat yang dipakai sangat diperlukan agar ahli bisa mengetahui dan mempercayai hasil dari alat tersebut. Kemampuan untuk bisa menjelaskan hasil prediksi juga termasuk salah satu aspek dari transparansi yang dibutuhkan dalam

dunia medis. Hal ini diperkuat juga dengan beberapa *stakeholders* menginginkan algoritma model yang digunakan bisa transparan dan bisa diimplementasikan dengan alur kerja sekarang[6].

Pada penelitian ini, peneliti akan menggunakan model EfficientNetV2B3 dikarenakan EfficientNetV2B3 adalah versi yang lebih baru dari keluarga EfficientNet terutama EfficientNetB3 dan juga dengan tujuan untuk mendapatkan performa yang lebih baik dari penelitian terdahulu, dimana penelitian tersebut mendapatkan nilai akurasi, recall dan f1-score sekitar 98%. Walaupun nilai tersebut sudah tinggi, akan tetapi masih memiliki ruang untuk peningkatan lebih lagi. Nilai yang semakin tinggi tersebut juga bisa membantu dalam pengambilan daerah untuk prediksi yang lebih tepat. Juga akan digunakan sebuah *Explainable AI* yaitu LIME yang akan menjelaskan dari pengklasifikasian tersebut dan diharapkan dapat membantu dalam menjelaskan hasil diagnosis

Untuk datasetnya, penulis akan menggunakan dataset publik yang bisa didapat dari Kaggle dengan isi datasetnya bernama C\_NMC 2019. Dataset ini berisi 10661 gambar yang terdiri dari 7272 gambar sel ALL dan 3389 gambar dari sel yang sehat. Dataset ini dipilih dikarenakan dataset ini dipakai dalam banyak penelitian sebelumnya, dengan contoh 2 penelitian yang sudah penulis sebutkan sebelumnya, dan dataeset ini sudah dilakukan *preprocessing* dimana setiap gambar yang ada adalah potongan satu gambar sel yang didapat dari gambar mikroskopik yang dilakukan pada saat melakukan pengecekan jumlah sel darah. Gambar-gambar tersebut sudah dilakukan segmentasi dan sudah dikelompokan berdasarkan jenisnya oleh ahli onkologi. Tidak hanya itu, jumlahnya yang banyak juga diharapkan dapat membuat model tersebut bisa mengklasifikasikan antara sel sehat dan yang tidak sehat dengan akurat.

Dikarenakan beberapa penelitian sebelumnya yang menggunakan dataset C-NMC 2019 sudah bisa mencapai hasil akurasi antara 95% sampai 98%, maka penelitian ini akan lebih memfokuskan terhadap hasil penjelasan dari LIME tersebut. Dikarenakan tidak adanya penelitian sebelumnya yang menggunakan

LIME untuk menjelaskan hasil prediksi dari model yang dipakai. Penulis akan melakukan evaluasi terhadap bagian-bagian yang diambil oleh modul LIME tersebut dan memeriksa apakah bagian yang diambil adalah bagian yang merupakan karakteristik dari sel *lymphoblast* dan *lymphocyte* berdasarkan review paper yang menjelaskan tentang pengujian blood smear[7]. Akan dilihat juga apakah modul LIME tersebut bisa mengeluarkan hasil yang sama untuk setiap iterasinya dan apakah hasil tersebut bisa dipercaya.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang yang sudah dipaparkan, identifikasi masalah terdiri dari beberapa poin berikut :

- 1.2.1 Apakah penggunaan model dengan versi lebih baru yaitu EfficientNetV2B3 dapat meningkatkan performa klasifikasi penyakit ALL?
- 1.2.2 Apakah daerah-daerah dari hasil penjelasan LIME sesuai dengan panduan yang digunakan oleh para ahli untuk mengklasifikasikan ALL?

## 1.3 Batasan Penelitian

Batasan dari penelitian ini sebagai berikut :

- 1.3.1 Dataset yang digunakan adalah dataset publik
- 1.3.2 Model yang digunakan adalah model pre-trained yang disediakan oleh Tensorflow dan Keras
- 1.3.3 Penelitian ini tidak melakukan validasi ke ahli onkologi untuk memeriksa keakuratan hasil dari *explainable AI*, akan tetapi diperiksa dengan panduan dari paper yang menjelaskan pengujian blood smear[6].

## 1.4 Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan untuk membandingkan performa EfficientNetV2B3 yang lebih baru dengan EfficientNetB3, dan juga untuk melihat apakah penggunaan *Explainable AI* terutama LIME dapat membantu

menjelaskan kepada pengguna dalam pengklasifikasian sel leukemia dengan jenis ALL.

### 1.5 Manfaat Penelitian

Manfaat yang dapat diambil dari penelitian ini adalah :

- Dapat digunakan sebagai sumber penelitian selanjutnya tentang klasifikasi sel leukemia ALL yang menggunakan *pre-trained model* yang lebih baru
- Dapat digunakan sebagai sumber penelitian selanjutnya tentang klasifikasi sel leukemia ALL yang menggunakan *explainable AI*.
- Dapat membantu dokter untuk mempercepat dan memperjelas hasil pengujian penyakit ALL menggunakan gambar mikroskopik.

