

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sel darah putih atau *leukocytes* berasal dari bahasa Yunani yaitu “*leukos*” yang berarti putih dan “*kytos*” yang berarti sel (Feher, 2012). Sel darah putih berperan penting sebagai sistem imun pada tubuh untuk melawan berbagai jenis patogen yang masuk ke dalam tubuh. Sel darah putih diproduksi di sumsum tulang, produksi sel darah putih dimulai dengan diferensiasi sel punca pluripoten menjadi berbagai jenis sel, waktu sirkulasi sel darah putih hanya beberapa jam saja yang jauh lebih singkat jika dibandingkan dengan sel darah merah (Louanjli et al., 2021). Sel darah putih terbagi menjadi dua jenis yang pertama *granular leukocytes* yang terdiri dari *eosinophil*, *neutrophil*, dan *basophil*, dinamai berdasarkan *granules* yang ada di dalam *cytoplasm*. Kedua *agranular leukocytes* yang terdiri dari *monocytes*, dan *lymphocytes*, dengan tidak adanya *granules*. Oleh karena itu *granulocytes* dapat diidentifikasi satu dengan yang lainnya berdasarkan morfologi dari nukleus, ukuran, dan warna dari granula (Louanjli et al., 2021). Dengan pemberian warna *histological* dapat membantu membedakan antara sel darah putih (Feher, 2012). Jumlah total dari sel darah putih merupakan salah satu hal yang penting dalam analisis darah. Akan tetapi jumlah sel darah putih saja tidaklah cukup informatif untuk dapat memahami informasi kesehatan secara utuh,

hal ini disebabkan karena peningkatan atau penurunan jumlah sel darah putih pada tubuh dapat dipengaruhi oleh faktor external seperti olahraga dan penyakit, maka perbedaan jenis leukosit atau *differential count* dapat membantu mengidentifikasi sel-sel tertentu yang jumlahnya berubah, serta dapat memberikan informasi penyebab terjadi perubahan tersebut (Louanjli et al., 2021).

Klasifikasi sel darah putih yang benar sangat penting untuk dapat dilakukan, akan tetapi melakukan identifikasi dan evaluasi melalui citra mikroskopik sangat rumit, memakan banyak waktu, dan terpengaruh oleh subjektif tenaga medis (Zhu et al., 2023), oleh sebab itu klasifikasi dengan sel darah putih secara otomatis dan akurat sangat bermanfaat. *Machine learning* dapat menjadi salah satu metode yang dapat digunakan untuk proses prediksi akan tetapi kemiripan antara sampel sel darah putih, ketidak seimbangan data sample yang dimiliki, kualitas, dan jumlah sampel yang dimiliki yang masih kurang sehingga membuat bidang *computer vision* pada bidang medis membawa sebuah tantangan baru (Chen et al., 2022).

Pada permasalahan jumlah data yang terjadi adalah masalah ketidak seimbangan data dimana salah satu kelas atau beberapa kelas memiliki jumlah data yang lebih banyak dibandingkan dengan kelas yang lainnya, atau yang dikenal dengan kelas mayoritas dan kelas minoritas. Data yang tidak seimbang dapat menyebabkan kesalahan klasifikasi akan cenderung terjadi pada kelas minoritas, yang membuat kelas minoritas menjadi lebih sulit untuk dapat diprediksi karena jumlahnya yang sedikit, dan menyebabkan bias pada data kelas mayoritas (Wijayanti et al., 2021). Pada dasarnya untuk mengatasi masalah ketidak seimbangan data ini terdapat dua pendekatan, yaitu pendekatan pada level algoritma

dan pendekatan pada level data. Pendekatan level data dilakukan dengan menyeimbangkan distribusi dari setiap kelas baik kelas mayoritas atau kelas minoritas, dengan menggunakan teknik pengambilan data seperti undersampling, overfitting atau gabungan antara kedua teknik tersebut. Sedangkan pendekatan dengan teknik level algoritma dapat dilakukan dengan memodifikasi atau mengoptimalkan algoritma tersebut (Wijayanti et al., 2021). Pendekatan pada level data dinilai lebih efektif untuk dapat menangani masalah ketidak seimbangan data, karena lebih fleksibel dan tidak terpengaruh dengan algoritma yang digunakan (He et al., 2008).

Penelitian terdahulu yang berkaitan dengan klasifikasi jenis sel darah putih sudah banyak dilakukan, dengan berbagai macam metode, baik dengan pendekatan tradisional atau dengan *deep learning* (Hegde et al., 2019). Pada pendekatan tradisional hal pertama yang harus dilakukan adalah melakukan ekstraksi fitur secara manual, sedangkan dengan menggunakan deep learning hal tersebut tidak perlu dilakukan secara manual, tetapi dengan menggunakan deep learning akan membutuhkan banyak parameter, membutuhkan banyak data training, sedangkan pada kasus ini dataset yang dimiliki tidak cukup besar (Tavakoli et al., 2021).

(Lin et al., 2023) pada tahun 2023 melakukan pengujian *pada dataset new-thyroid1, Ecoli2, Wisconsin, dan lung Cancer, dataset ini merupakan dataset yang imbalance*. Penelitian ini menggunakan SVM dan beberapa teknik data *balancing*, diperoleh hasil terbaik dengan menggunakan SMOTE untuk dataset *new-thyroid1* dengan akurasi 96.5%, pada *dataset Ecoli2* juga mendapatkan hasil terbaik dengan menggunakan SMOTE dengan akurasi 88.8%, sedangkan untuk *dataset wisconsin*

dan *lung cancer* hasil terbaik diperoleh dengan menggunakan MMTD-ELM dengan akurasi 90.5 untuk dataset wisconsin, dan 99.8% untuk dataset lung cancer.

(Rustam et al., 2022) pada tahun 2022 mengajukan *texture feature extraction* dan RGB pada WBCs, ekstraksi dengan menggunakan *texture* menghasilkan 67.500 *feature*, dan ekstraksi dengan *texture* menghasilkan 22.500 *feature*, sehingga total dari kedua *feature* sebanyak 90.000 *feature*, dari 90.000 *feature* ini akan seleksi lagi dengan menggunakan Chi2 sehingga hasil akhir mendapatkan 6000 *feature*. Sehingga dengan 6000 *feature* ini akan diklasifikasi dengan SVM + SMOTE mendapatkan akurasi 81%, presisi 81%, recall 81%, dan f1-score 81%.

(Tavakoli et al., 2021) pada tahun 2021 mengajukan penelitian dengan menggunakan *shape feature extraction* (Solidity, convexity, circularity) dan color (RGB, HSV, LAB, YCrCb), prose klasifikasi dengan menggunakan algoritma SVM mendapatkan akurasi 94.65%

(Devella et al., 2021) pada tahun 2021 mengajukan penggunaan *saliency* sebagai metode untuk melakukan segmentasi sel darah putih, dan dengan menggunakan *Speeded-Up Robust Features (SURF)* sebagai *feature extraction*, mendapatkan akurasi sebesar 78.60% dengan menggunakan SVM.

(Yohannes et al., 2021) pada tahun 2021 mengajukan penggunaan *saliency* sebagai metode untuk melakukan segmentasi sel darah putih, dan dengan menggunakan *scale invariant feature transform (SIFT)* sebagai *feature extraction*, mendapatkan akurasi sebesar 77.08% dengan menggunakan SVM.

(Riaz et al., 2020) pada tahun 2020 menggunakan *gaussian mixture model* untuk segmentasi citra medis. Segmentasi digunakan pada tiga *distinct imaging modalities*, *MRI*, *dermoscopy*, dan *chromoendoscopy*. Hasil dari penelitian ini menunjukkan bahwa metode yang diusulkan memberikan hasil kualitatif dan kuantitatif yang lebih baik dibandingkan dengan metode segmentasi citra medis saat ini.

(P. Wang et al., 2021) pada tahun 2019 menggunakan *color*, *texture*, menggunakan GLCM dan dengan parameter *values of area*, *solidity*, *eccentricity*, dan *perimeter*, dengan total gambar 200 gambar dimana setiap jenis sel sebanyak 40 gambar. mendapatkan akurasi sebesar 88.5% dengan menggunakan SVM.

(Paco Ramos et al., 2019) mengajukan penggunaan *convolutional autoencoder (CAE)* sebagai *feature extraction* pada kasus klasifikasi daun tanaman, hasil dari CAE akan diklasifikasi dengan menggunakan SVM mendapatkan akurasi sebesar 94.74%.

Berdasarkan latar belakang yang telah dijelaskan, maka untuk penelitian yang akan dilakukan kali ini memiliki judul **Klasifikasi Sel Darah Putih Menggunakan Metode SMOTE-SVM dengan Ekstraksi Fitur Gabungan dan Segmentasi Citra Menggunakan Gaussian Mixture Model.**

1.2 Identifikasi Masalah

Berdasarkan latar belakang dari masalah yang telah diuraikan pada bagian sebelumnya maka dapat disimpulkan beberapa hal yang akan menjadi landasan dari penelitian ini akan dilakukan, yang pertama bagaimana caranya memanfaatkan

dataset yang tersedia untuk memperoleh hasil yang semaksimal mungkin. Memanfaatkan algoritma *gaussian mixture model* untuk dapat melakukan segmentasi citra sel darah putih.

1.3 Batasan Masalah

Tujuan dari adanya batasan pada penelitian yang akan dilakukan agar penelitian ini dapat fokus pada permasalahan yang ada, maka ditetapkan batasan dari penelitian ini sebagai berikut:

1. Dataset yang akan digunakan adalah dataset publik dengan nama Raabin-WBC yang dipublikasikan pada tahun 2021 (Kouzehkanan et al., 2021), yang berisi 5 jenis sel darah putih, yaitu *basofil*, *neutrophils*, *monocytes*, *lymphocyte*, dan *eosinophil*.
2. Proses segmentasi gambar dilakukan dengan menggunakan *gaussian mixture model (GMM)* dan dengan menggunakan metode yang berdasarkan pada konsep cara manusia melihat sebuah objek.
3. Menggunakan algoritma *support vector machine* dan dengan kernel *radial basis function (RBF)*, linear, polynomial, dan sigmoid sedangkan untuk mengatasi masalah *imbalance* menggunakan *borderline-SMOTE SVM* atau *SVMSMOTE* dan *SMOTE*.
4. *Feature extraction* menggunakan *size*, *solidity*, *convexity*, *circularity* yang akan digabungkan dengan *convolutional autoencoder*.
5. Dataset yang akan digunakan sebanyak 14.251 data dengan ukuran citra *575 X 575 pixels*
6. Menggunakan *Confusion matrix* untuk evaluasi performa model

7. Menggunakan K-Fold *cross-validation* dengan K= 5
8. Dataset menggunakan format file *.jpg
9. Satu citra hanya mengandung satu jenis sel darah putih.
10. Implementasi akan menggunakan bahasa pemrograman python.

1.4 Rumusan Masalah

Rumusan permasalahan pada penelitian ini sebagai berikut :

1. Bagaimana hasil evaluasi model SVM yang dilatih dengan hasil ekstraksi fitur *size*, *solidity*, *convexity*, dan *circularity* dan *convolutional autoencoder* dalam melakukan indentifikasi jenis sel darah putih.
2. Bagaimana hasil penggunaan algoritma data *balancing* SMOTE dan SVM SMOTE berdasarkan ekstraksi fitur *convolutional autoencoder* dan *size*, *solidity*, *convexity*, dan *circularity* dalam melakukan indentifikasi jenis sel darah putih yang imbalance

1.5 Tujuan Penelitian

Penelitian ini memiliki beberapa tujuan pada bidang *machine learning* dan dunia kedokteran yang akan memanfaatkan kecerdasan buatan, yang diharapkan penelitian ini dapat ikut berkontribusi. Adapun beberapa tujuan tersebut sebagai berikut

1. Mengetahui efektivitas dari penggunaan *convolutional autoencoder* serta fitur *size*, *convexity*, *circularity*, dan *solidity*, dalam merepresentasikan setiap jenis sel darah putih.

2. Mengevaluasi hasil model SVM antara tanpa data balancing, SMOTE, dan SVMSMOTE dalam mengidentifikasi jenis sel darah putih yang *imbalance*.
3. Mengetahui efektivitas dari penggunaan *gaussian mixture model* dengan titik tengah, untuk melakukan segmentasi *nucleus* sel darah putih.
4. Pengembangan penelitian pada bidang kedokteran dan biologi dengan memanfaatkan *machine learning*.

1.6 Manfaat Penelitian

Adapun manfaat dalam penelitian ini sebagai berikut :

1. Meningkatkan akurasi klasifikasi dan objektivitas hasil klasifikasi sel darah putih.
2. Meningkatkan kecepatan klasifikasi dengan mengurangi waktu dan tenaga yang dibutuhkan oleh tenaga medis.
3. Penulis mengetahui cara menggunakan dan mengembangkan *gaussian mixture model* untuk dapat digunakan untuk segmentasi citra sel darah putih.
4. Memberikan pengetahuan tentang keefektifan dari metode yang diusulkan dalam penelitian.
5. Berkontribusi pada bidang medis dengan memanfaatkan *machine learning*.

1.7 Sistematika Penulisan

Adapun sistematika penulisan pada proposal tesis ini susunannya dibagi menjadi beberapa bab yaitu :

BAB I PENDAHULUAN

Pada bab ini berisi latar belakang masalah, identifikasi masalah, batasan masalah, rumusan masalah, tujuan penelitian, manfaat penelitian serta sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini akan membahas mengenai dasar-dasar teori, penelitian terdahulu, serta hipotesis penelitian.

BAB III METODOLOGI PENELITIAN

Pada bab ini akan membahas mengenai metode pengumpulan data serta metode penelitian yang digunakan, dan tahapan dan kegiatan pada penelitian yang akan dilakukan.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini akan menjelaskan hasil dari pengujian yang telah dilakukan, mengenai klasifikasi sel darah putih dengan menggunakan metode yang diusulkan.

BAB V PENUTUP

Pada bagian bab akhir ini akan membahas mengenai kesimpulan dan saran yang dapat berikan untuk penelitian selanjutnya, berdasarkan apa yang telah diperoleh selama penelitian ini berlangsung.