

**KLASIFIKASI SEL DARAH PUTIH MENGGUNAKAN
METODE SMOTE-SVM DENGAN EKSTRAKSI FITUR
GABUNGAN DAN SEGMENTASI CITRA MENGGUNAKAN
GAUSSIAN MIXTURE MODEL**



TESIS

**CELVINE ADI PUTRA
ENTERPRISE IT INFRASTRUCTURE
222420014**

**PROGRAM STUDI TEKNIK INFORMATIKA - S2
PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA
PALEMBANG
2024**

**KLASIFIKASI SEL DARAH PUTIH MENGGUNAKAN
METODE SMOTE-SVM DENGAN EKSTRAKSI FITUR
GABUNGAN DAN SEGMENTASI CITRA MENGGUNAKAN
GAUSSIAN MIXTURE MODEL**

**Tesis ini diajukan sebagai salah satu syarat
untuk memperoleh gelar**

MAGISTER KOMPUTER



**CELVINE ADI PUTRA
ENTERPRISE IT INFRASTRUCTURE
222420014**

**PROGRAM STUDI TEKNIK INFORMATIKA - S2
PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA
PALEMBANG**

2024

HALAMAN PENGESAHAN PEMBIMBING TESIS

Judul Tesis : KLASIFIKASI SEL DARAH PUTIH MENGGUNAKAN
METODE SMOTE-SVM DENGAN EKSTRAKSI FITUR
GABUNGAN DAN SEGMENTASI CITRA MENGGUNAKAN
GAUSSIAN MIXTURE MODEL

Oleh CELVINE ADI PUTRA NIM 222420014 Tesis ini telah disetujui dan
disahkan oleh Tim Penguji Program Studi Teknik Informatika – S2 konsentrasi
ENTERPRISE IT INFRASTRUCTURE, Program Pascasarjana Universitas Bina
Darma pada 22 Agustus 2024 dan telah dinyatakan LULUS.

Mengetahui,
Program Studi Teknik Informatika – S2
Universitas Bina Darma
Ketua,

Pembimbing,


Universitas Bina Darma
Magister Teknik Informatika

Dr. Usman Ependi, S.Kom., M.Kom.


Dr. Tata Sutabri, S.Kom., M.M.S.I

HALAMAN PENGESAHAN PENGUJI TESIS

Judul Tesis : KLASIFIKASI SEL DARAH PUTIH MENGGUNAKAN METODE SMOTE-SVM DENGAN EKSTRAKSI FITUR GABUNGAN DAN SEGMENTASI CITRA MENGGUNAKAN GAUSSIAN MIXTURE MODEL

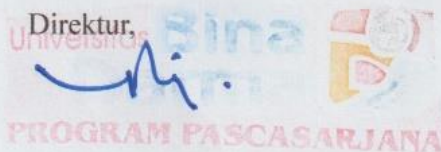
Oleh CELVINE ADI PUTRA NIM 222420014 Tesis ini telah disetujui dan disahkan oleh Tim Penguji Program Studi Teknik Informatika – S2 konsentrasi ENTERPRISE IT INFRASTRUCTURE, Program Pascasarjana Universitas Bina Darma pada 22 Agustus 2024 dan telah dinyatakan LULUS.

Palembang, 22 Agustus 2024

Mengetahui,

Program Pascasarjana

Universitas Bina Darma



Direktur,

Prof. Dr. Ir. Achmad Syarifudin, M.Sc.

Tim Penguji :

Penguji I,

Dr. Tata Sutabri, S.Kom., M.M.S.I.

Penguji II,

Prof. Dr. Edi Surya Negara, M.Kom.

Penguji III,

Dr. Yesi Novaria Kunang, S.T., M.Kom.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini

Nama : CELVINE ADI PUTRA
NIM : 222420014

Dengan ini menyatakan bahwa:

1. Karya tulis Saya Tesis ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik Magister di Universitas Bina Darma.
2. Karya tulis ini murni gagasan, rumusan dan penelitian Saya sendiri dengan arahan tim pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas kutipan dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar pustaka.
4. Karena yakin dengan keaslian karya tulis ini, Saya menyatakan bersedia Tesis yang Saya hasilkan di unggah ke internet.
5. Surat pernyataan ini Saya tulis dengan sungguh- sungguh dan apabila terdapat penyimpangan atau ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi dengan aturan yang berlaku di perguruan tinggi ini.

Demikian Surat Pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 22 Agustus 2024
Yang membuat Pernyataan,



CELVINE ADI PUTRA
222420014

ABSTRAK

Sel darah putih atau leukosit memainkan peran penting dalam sistem kekebalan tubuh. Struktur yang tidak beraturan dan beragam pada setiap jenis sel darah putih membuat identifikasi manual sulit dan rentan terhadap kesalahan karena subjektivitas dan kelelahan. Oleh karena itu, diperlukan metode klasifikasi yang cepat dan akurat. Penelitian ini mengusulkan penggunaan SMOTE dan SVMSMOTE untuk mengatasi ketidakseimbangan data, serta ekstraksi fitur gabungan antara *shape (size, convexity, circularity, dan solidity)* dan *convolutional autoencoder*, bersama dengan *Gaussian Mixture Model* untuk segmentasi *nucleus*. Hasil penelitian menunjukkan bahwa tanpa data balancing, metode ekstraksi fitur yang diusulkan sudah cukup baik, kecuali untuk sel eosinophils. Hasil terbaik diperoleh dengan SVM menggunakan kernel polynomial, dengan akurasi 92.4%, presisi 91.9%, recall 92.3%, F1-Score 92%, MCC 0.862, dan CEN 0.1376. Hasil terburuk dengan menggunakan kernel sigmoid. Pengujian yang hanya menggunakan ekstraksi fitur *shape* saja mendapatkan akurasi terbaik 86.8%, sedangkan untuk *convolutional autoencoder* mendapatkan akurasi 87.8%, sehingga dengan menggunakan *shape* dan *convolutional autoencoder* dapat meningkatkan hasil akurasi. Selain itu penggunaan SMOTE atau SVMSMOTE dapat meningkatkan recall pada sel eosinophils.

Kata Kunci : Convolutional AutoEncoder, Gaussian Mixture Model, SMOTE, SVMSMOTE, Support Vector Machine, Sel Darah Putih

ABSTRACT

White blood cells, or leukocytes, have a crucial role in the immune system. The irregular structure of white blood cells, along with each type having its unique structure, makes manual identification challenging. Manual identification processes can lead to errors due to the subjectivity of medical personnel and the fatigue caused by the time and effort required. Therefore, there is a need to develop a fast and accurate method for classifying different types of white blood cells. However, issues such as the quality and the number of samples available for every kind of cell remain problematic. This study proposes to use SMOTE and SVMSMOTE to address the issue of data imbalance and use a combination of shape (size, circularity, convexity, solidity) and convolutional autoencoder (CAE) as feature extraction, along with a Gaussian mixture model for nucleus segmentation. The study finds that, without using SMOTE or SVMSMOTE for data balancing, the proposed features are already sufficient to represent each cell type except eosinophils, achieving an accuracy of 92.4%, precision of 91.9%, recall of 92.3%, F1-Score of 92%, MCC of 0.862, and CEN of 0.1376 using polynomial kernel and the worst results were obtained with the sigmoid kernel. Moreover, combined feature extraction between shape and CAE worked better compared to using shape and CAE separately. At best, accuracy achieved was 86.8% using shape alone, while the results produced using CAE alone achieved as high as 87.8%. Furthermore, recall for eosinophil cells can be improved using SMOTE and SVMSMOTE.

Keywords : Convolutional AutoEncoder, Gaussian Mixture Model, SMOTE, SVMSMOTE, Support Vector Machine, White Blood Cell

MOTTO DAN HALAMAN PERSEMBAHAN

MOTTO

- ❖ *“Embrace change, even when it's unexpected and doesn't fit our plans, because God's plans are greater than ours.”*
- ❖ *“Keep your eyes on the stars”*

PERSEMBAHAN

- ❖ Tuhan Yang Maha Esa
- ❖ Kedua Orang Tua
- ❖ Keluarga
- ❖ Dosen Pembimbing
- ❖ Teman Seperjuangan Magister Teknik Informatika Universitas Bina Darma

KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada Tuhan yang Maha Esa, sehingga penulis dapat menyelesaikan laporan tesis yang berjudul **“Klasifikasi Sel Darah Putih dengan Menggunakan Metode SMOTE-SVM dengan Ekstraksi Fitur Gabungan dan Segmentasi Citra Menggunakan Gaussian Mixture Model”**. Sebagai salah satu syarat untuk memperoleh gelar Magister Komputer pada Universitas Bina Darma Palembang.

Penulis menyadari bahwa laporan yang dibuat masih belum sempurna, yang terjadi karena keterbatasan pengetahuan yang dimiliki oleh penulis, sehingga segala saran yang dapat membangun akan dapat diterima dengan senang hati.

Penulis juga banyak mengucapkan terima kasih kepada semua pihak yang telah memberikan bimbingan, arahan, serta kontribusi pemikiran mereka dalam penelitian ini. Juga kepada mereka yang memberikan dukungan kepada penulis yang menjadi energi positif bagi penulis untuk dapat terus melanjutkan penelitian ini sampai dengan selesai. Oleh karena itu penulis ingin mengucapkan terima kasih sebesar-besarnya kepada:

1. Tuhan yang Maha Esa yang telah menolong penulis sejauh ini untuk dapat terus semangat sehingga penulis dapat menyelesaikan tesis ini.
2. Seluruh anggota keluarga, khususnya kedua orang tua penulis yang telah membantu penulis hingga dapat sampai titik ini.

3. Bapak Prof. Dr. Ir. Achmad Syarifudin, M.Sc. selaku direktur program pascasarjana Universitas Bina Darma.
4. Bapak Dr. Usman Ependi, S.Kom., M.Kom selaku ketua program studi pascasarjana.
5. Bapak Dr. Tata Sutabri, S.Kom., M.M.S.I selaku dosen pembimbing yang dapat memberikan bimbingan, arahan, dan masukan selama proses pengerjaan tesis.
6. Bapak Prof. Dr. Edi Surya Negara, M.Kom, S.T., M.Kom. selaku dosen penguji yang memberikan masukan untuk penelitian ini.
7. Ibu Dr. Yesi Novaria Kunang, S.T., M.Kom. selaku dosen penguji yang telah memberikan masukan untuk penelitian ini.
8. Dosen Universitas Bina Darma yang telah membagi ilmu dan pengalamannya selama perkuliahan.
9. Seluruh teman-teman MTI angkatan 2022, yang telah bersama sampai dengan selesai.
10. Seluruh karyawan pada Universitas Bina Darma dan pihak-pihak lain yang telah mendukung dan membantu dalam penyusunan tesis ini yang tidak dapat penulis sebutkan satu persatu.

Palembang, 22 Agustus 2024
Penulis,

Celvine Adi Putra
222420014

DAFTAR ISI

COVER TESIS	i
HALAMAN DEPAN	ii
HALAMAN PENGESAHAN PEMBIMBING TESIS	iii
HALAMAN PENGESAHAN PENGUJI TESIS	iv
SURAT PERNYATAAN	iv
ABSTRAK	vi
ABSTRACT	vii
MOTTO DAN HALAMAN PERSEMBAHAN	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah.....	5
1.3 Batasan Masalah	6
1.4 Rumusan Masalah.....	7
1.5 Tujuan Penelitian	7
1.6 Manfaat Penelitian	8
1.7 Sistematika Penulisan	9
BAB II TINJAU PUSTAKA	10
2.1 Sel Darah Putih	10
2.2 Segmentasi Citra Digital	13
2.3 Klasifikasi	13
2.4 Clustering	14

2.5	Ekstraksi Fitur	15
2.6	Gaussian Mixture Model.....	16
2.7	Hue Saturation Value (HSV)	18
2.8	Convolutional Autoencoder	19
2.9	Synthetic Minority Over-Sampling Technique.....	21
2.10	Support Vector Machine	23
2.11	K-Fold Cross-Validation.....	27
2.12	Confusion Matrix	28
2.13	Accuracy	29
2.14	Precision.....	30
2.15	Recall	30
2.16	F1-Score.....	30
2.17	Matthews Correlation Coefficient (MCC).....	31
2.18	Confusion Entropy (CEN)	31
2.19	Kerangka Berpikir.....	33
2.20	Penelitian Sebelumnya.....	39
BAB III METODOLOGI PENELITIAN		48
3.1.	Metode Penelitian	48
3.1.1	Identifikasi masalah	49
3.1.2	Studi literatur	49
3.1.3	Pengumpulan data.....	49
3.1.4	Perancangan sistem.....	51
3.1.5	Implementasi.....	53
3.1.6	Pengujian	53
3.1.7	Pembuatan laporan.....	54
BAB IV HASIL DAN PEMBAHASAN		55
4.1.	Pendahuluan.....	55
4.2.	Proses Segmentasi Citra.....	55
4.3.	Proses Ekstraksi Fitur	59
4.4.	Hasil Pengujian	61
4.4.1.	Kernel Linear	62

4.4.2.	Kernel RBF	68
4.4.3.	Kernel Sigmoid	75
4.4.4.	Kernel Polynomial	81
4.5.	Perbandingan Hasil Pengujian	88
4.5.1	Tanpa data Balancing.....	88
4.5.2	SMOTE.....	89
4.5.3	SVMSMOTE	90
4.5.4	Perbandingan dengan Penelitian Terdahulu	91
BAB V	PENUTUP.....	93
5.1.	Kesimpulan	93
5.1.	Saran	94
DAFTAR PUSTAKA		96
DAFTAR RIWAYAT HIDUP		102
LAMPIRAN.....		103

DAFTAR TABEL

Tabel 2.1 Binary Confusion Matrix.....	29
Tabel 2.2 Multi Class Confusion Matrix	29
Tabel 2.3 Penelitian Sebelumnya.....	40
Tabel 4.1 Persentase Hasil Segmentasi Train.rar.....	59
Tabel 4.2 Persentase Hasil Segmentasi Test.rar.....	59
Tabel 4.3 Model Summary Convolutional Autoencoder	60
Tabel 4.4 Jumlah Feature.....	61
Tabel 4.5 Hasil Kernel Linear (Normal).....	62
Tabel 4.6 Evaluasi Model Kernel Linear (Normal).....	62
Tabel 4.7 Hasil Kernel Linear SMOTE	63
Tabel 4.8 Evaluasi Model Kernel Linear (SMOTE).....	64
Tabel 4.9 Hasil Kernel Linear SVM SMOTE	65
Tabel 4.10 Evaluasi Model Kernel Linear (SVM SMOTE).....	66
Tabel 4.11 Hasil Kernel RBF (Normal).....	68
Tabel 4.12 Evaluasi Model Kernel RBF (Normal).....	68
Tabel 4.13 Hasil Kernel RBF (SMOTE)	70
Tabel 4.14 Evaluasi Model Kernel RBF (SMOTE)	70
Tabel 4.15 Hasil Kernel RBF (SVM SMOTE).....	71
Tabel 4.16 Evaluasi Model Kernel RBF (SVM SMOTE).....	72
Tabel 4.17 Hasil Kernel Sigmoid (Normal).....	75
Tabel 4.18 Evaluasi Model Kernel Sigmoid (Normal).....	75
Tabel 4.19 Hasil Kernel Sigmoid (SMOTE)	77
Tabel 4.20 Hasil Evaluasi Kernel Sigmoid (SMOTE)	77
Tabel 4.21 Hasil Kernel Sigmoid (SVM SMOTE).....	78
Tabel 4.22 Hasil Evaluasi Kernel Sigmoid (SVM SMOTE).....	79
Tabel 4.23 Hasil Kernel Polynomial (Normal).....	81

Tabel 4.24 Hasil Evaluasi Kernel Polynomial (Normal).....	81
Tabel 4.25 Hasil Kernel Polynomial (SMOTE)	83
Tabel 4.26 Hasil Evaluasi Kernel Polynomial (SMOTE).....	83
Tabel 4.27 Hasil Kernel Polynomial (SVMSMOTE).....	84
Tabel 4.28 Hasil Evaluasi Kernel Polynomial (SVMSMOTE).....	85
Tabel 4.29 Perbandingan Hasil Pengujian Tanpa Data Balancing	88
Tabel 4.30 Perbandingan Pengujian Ekstraksi Fitur Tanpa Data Balancing	88
Tabel 4.31 Perbandingan Hasil Pengujian dengan SMOTE.....	89
Tabel 4.32 Perbandingan Pengujian Ekstraksi Fitur SMOTE	89
Tabel 4.33 Perbandingan Hasil Pengujian dengan SVMSMOTE	90
Tabel 4.34 Perbandingan Pengujian Ekstraksi Fitur SVMSMOTE	90



DAFTAR GAMBAR

Gambar 2.1 Neutrophils	11
Gambar 2.2 Eosinophils	11
Gambar 2.3 Basophils.....	12
Gambar 2.4 Lymphocytes.....	12
Gambar 2.5 Monocytes.....	12
Gambar 2.6 Model Jaringan Autoencoder	20
Gambar 2.7 SMOTE.....	22
Gambar 2.8 Support Vector Machine	25
Gambar 2.9 K-Fold Cross Validation	28
Gambar 2.10 Kerangka Berpikir	33
Gambar 2.11 Proses Segmentasi dan Titik Tengah.....	34
Gambar 2.12 Contoh Konversi Warna dari RGB ke HSV	35
Gambar 2.13 Contoh Visualisasi Klasterisasi.....	35
Gambar 2.14 Contoh Visualisasi Titik Tengah Basophil.....	36
Gambar 2.15 Contoh Visualisasi Titik Tengah Eosinophil.....	36
Gambar 2.16 Contoh Visualisasi Titik Tengah Lymphocyte	36
Gambar 2.17 Contoh Visualisasi Titik Tengah Monocyte	36
Gambar 2.18 Contoh Visualisasi Titik Tengah Neutrophil.....	37
Gambar 3.1 Metodologi Penelitian.....	48
Gambar 3.2 Jumlah Data Train.rar Sebelum dan Setelah Seleksi	50
Gambar 3.3 Jumlah Data TestA.rar Sebelum dan Setelah Seleksi	50
Gambar 3.4 Perancangan Sistem	52
Gambar 4.1 Citra Konversi RGB ke HSV Sel Basophil	56
Gambar 4.2 Citra Konversi RGB ke HSV Sel Eosinophil	56
Gambar 4.3 Citra Konversi RGB ke HSV Sel Lymphocyte.....	56
Gambar 4.4 Citra Konversi RGB ke HSV Sel Monocyte	56

Gambar 4.5 Citra Konversi RGB ke HSV Sel Neutrophil	56
Gambar 4.6 Contoh Hasil Clustering	57
Gambar 4.7 Contoh Pemberian Titik Tengah	58
Gambar 4.8 Contoh Hasil Sel Segmentasi.....	58
Gambar 4.9 Contoh Hasil Sel Gagal Segmentasi	58
Gambar 4.10 Model Loss Convolutional AutoEncoder	60
Gambar 4.11 Perbandingan Antar Kelas Kernel Linear (Normal)	63
Gambar 4.12 Perbandingan Antar Kelas Kernel Linear (SMOTE).....	65
Gambar 4.13 Perbandingan Antar Kelas Kernel Linear (SVMSMOTE)	66
Gambar 4.14 Perbandingan Hasil Akhir Kernel Linear	67
Gambar 4.15 Perbandingan Antar Kelas Kernel RBF (Normal)	69
Gambar 4.16 Perbandingan Antar Kelas Kernel RBF (SMOTE).....	71
Gambar 4.17 Perbandingan Antar Kelas Kernel RBF (SVMSMOTE)	73
Gambar 4.18 Perbandingan hasil Akhir Kernel RBF	74
Gambar 4.19 Perbandingan Antar Kelas Kernel Sigmoid (Normal)	76
Gambar 4.20 Perbandingan Antar Kelas Kernel Sigmoid (SMOTE).....	78
Gambar 4.21 Perbandingan Antar Kelas Kernel Sigmoid (SVMSMOTE)	80
Gambar 4.22 Perbandingan hasil Akhir Kernel Sigmoid	80
Gambar 4.23 Perbandingan Antar Kelas Kernel Polynomial (Normal)	82
Gambar 4.24 Perbandingan Antar Kelas Kernel Polynomial (SMOTE).....	84
Gambar 4.25 Perbandingan Antar Kelas Kernel Polynomial (SVMSMOTE)	86
Gambar 4.26 Perbandingan hasil Akhir Kernel Polynomial	87

DAFTAR LAMPIRAN

Lampiran 1. SK Pembimbing.....	103
Lampiran 2. Hasil Pengujian Lainnya.....	104
Lampiran 3. Artikel Jurnal.....	155
Lampiran 4. Lembar Konsultasi Tesis, Hasil dan Proposal.....	165
Lampiran 5. Lembar Perbaikan Tesis.....	169

