

Peningkatan Performa Klasifikasi Sel Darah Merah pada Pasien Talasemia Minor

Husnul Latifah*¹, Dyah Aruming Tyas², Agus Harjoko³

¹Program Studi Elektronika dan Instrumentasi, FMIPA, UGM, Yogyakarta, Indonesia

^{2,3}Departemen Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta, Indonesia

e-mail: *husnul.latifah@mail.ugm.ac.id, dyah.aruming.t@ugm.ac.id, aharjoko@ugm.ac.id

Abstrak

Talasemia merupakan kelainan darah turunan yang menyebabkan rusaknya rantai hemoglobin pada eritrosit penderita. Pada kasus talasemia minor, pasien hanya menjadi pembawa gen talasemia dan tidak bergejala. Hal ini menyebabkan sedikitnya penderita talasemia minor yang terdeteksi. Saat ini, ahli hematologi harus menghitung eritrosit abnormal secara manual berdasarkan bentuk, warna, dan tekstur sel. Untuk itu, banyak penelitian yang memanfaatkan citra eritrosit untuk melakukan mengklasifikasi keabnormalan pada citra eritrosit secara otomatis. Namun, jumlah data yang terbatas menyebabkan salah satu jenis keabnormalan yaitu sel pensil belum terklasifikasi pada penelitian sebelumnya. Tujuan dari penelitian ini adalah mengatasi kendala tersebut dengan melakukan klasifikasi secara bertingkat dimana pada tahapan klasifikasi pertama sel pensil dikelompokkan kepada sel yang mirip, yaitu sel eliptosis terlebih dahulu. Penelitian ini menggunakan klasifier Convolutional Neural Network (CNN) pada proses klasifikasi pertama dan Support Vector Machine (SVM) pada proses klasifikasi kedua. Hasil eksperimen menunjukkan klasifier CNN dengan arsitektur MiniVGGNet dan berhasil mengklasifikasi citra eritrosit ke dalam delapan kelas dengan nilai akurasi 96,05%, presisi 96,00%, sensitivitas 96,05%, dan F1 score 95,95%. Klasifier SVM Polinomial dengan kombinasi fitur geometris yang terdiri dari eccentricity, compactness, circularity, dan rasio sel berhasil mengklasifikasi sel pensil dengan nilai presisi 100,00%, sensitivitas 100,00%, dan F1 score 100,00%.

Kata kunci— Klasifikasi, Multilevel, Talasemia, Eritrosit, CNN, SVM, MiniVGGNet

Abstract

Thalassemia is a hereditary blood disorder that causes hemoglobin chain damage in a patient's erythrocyte. In the minor thalassemia case, the patient only became a carrier and had no symptoms. Subsequently, very few patients get diagnosed. At this moment, hematologists must count abnormal erythrocytes manually based on their shape, color, and texture. As such, many studies have utilized erythrocytes images to classify their abnormalities automatically. Due to limited data, a type of abnormality namely pencil cell remains unclassified in recent studies. The purpose of this study is to overcome these hurdles by carrying out multilevel classification, in which in the first classification pencil cells are grouped into similar cells, namely elliptocytes. This study used a Convolutional Neural Network (CNN) classifier in the first classification process and a Support Vector Machine (SVM) classifier in the second classification process. Results show the system with CNN classifier and MiniVGGNet architecture has succeeded in classifying erythrocyte images into eight classes with 96.05% accuracy, 96.00% precision, 96.05% sensitivity, and 95.95% F1 score value. The Polynomial SVM classifier with a geometric features combination consisting of eccentricity, compactness, circularity, and cell ratio succeeded in classifying pencil cells with precision 100.00%, 100.00% sensitivity, and 100.00% F1 score value.

Keywords— Classification, Multilevel, Thalassemia, Erythrocyte, CNN, SVM, MiniVGGNet

1. PENDAHULUAN

Talasemia merupakan kelainan genetik turunan yang menyerang sel darah merah penderitanya. Berdasarkan tingkat keparahannya, talasemia dibagi menjadi talasemia mayor dan talasemia intermedia yang membutuhkan tranfusi darah secara rutin serta talasemia minor. Penderita talasemia minor umumnya tidak memiliki gejala. Hal inilah yang menyebabkan skrining dini talasemia sangat krusial. Saat ini metode skrining yang dilakukan oleh para ahli hematologi dengan melibatkan pemeriksaan darah, masih menggunakan deteksi manual yang mengandalkan pemeriksaan mikroskopis sel darah merah, sebuah proses yang memakan waktu dan rentan terhadap kesalahan. Pada penelitian terbaru telah menggunakan *computer vision*, memanfaatkan data gambar mikroskopis untuk deteksi penyakit secara otomatis.

Pada penelitian tentang klasifikasi sel darah merah pada pasien talasemia, fitur bentuk sel menjadi fitur yang sering digunakan [1][2]. Pembagian kelas sel darah merah abnormal dengan data citra umumnya didasari oleh bentuk sel darah merah. Fitur bentuk ini dapat diwakili oleh parameter geometris sel darah merah [4][5][6]. Parameter geometris tersebut diantaranya adalah luas dan keliling sel darah merah serta axis mayor dan axis minor sel. Dari nilai luas dan keliling tersebut, dapat diekstrak nilai feature descriptor yang nantinya akan menjadi input klasifier citra.

Sel darah merah pasien diklasifikasikan ke dalam sembilan kelas dengan klasifier *Multi-Layer Perceptron* (MLP). Namun, penelitian ini belum berhasil mengklasifikasi citra sel pensil. Dimana seluruh anggota sel pensil diklasifikasikan menjadi sel eliptosis [3]. Sel pensil sendiri sangatlah krusial dalam membedakan talasemia minor dan kelainan darah lain karena digunakan sebagai salah satu pembeda pasien talasemia minor dan pasien Anemia Defisiensi Besi (ADB)[7].

Untuk mengatasi kesalahan klasifikasi, diusulkan metode klasifikasi bertingkat yang membagi sel darah merah menjadi kelas normal dan abnormal, kemudian kelas abnormal akan dispesifikasikan menjadi 3 kelas [8]. Terinspirasi oleh pendekatan ini, penulis menyarankan klasifikasi dua tahap. Pada tahap pertama, sel pensil digabungkan dengan sel eliptosis, dan CNN digunakan untuk membuat delapan kelas. Pada tahap kedua, SVM digunakan untuk membedakan sel eliptosis dan sel pensil karena keterbatasan kumpulan data sel pensil.

2. METODE PENELITIAN

Pada bagian ini, difokuskan pada penguraian metode yang akan digunakan dalam penelitian. Tujuan utamanya adalah mengembangkan sistem yang mampu mengklasifikasikan gambar sel darah merah. Sistem ini dimaksudkan untuk mengidentifikasi kelainan pada sel darah merah dan menjadi parameter penting dalam proses skrining talasemia minor.

2.1 Pengumpulan Data Penelitian

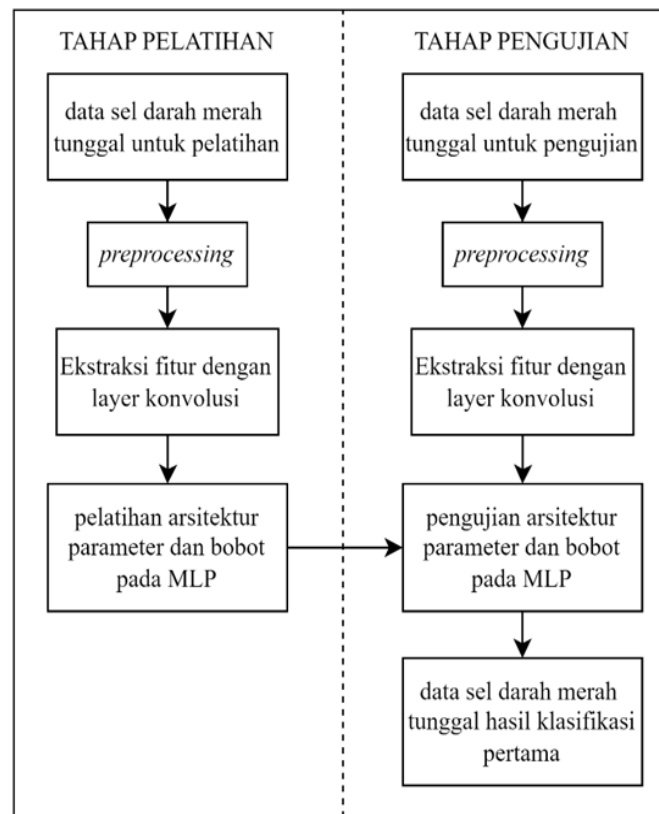
Penelitian ini menggunakan dataset yang bersumber dari *erythrocyte (red blood cell) dataset in thalassemia case* [9] yang berasal dari Laboratorium Patologi Fakultas Kesehatan Masyarakat Kedokteran dan Keperawatan Universitas Gadjah Mada. Dataset terdiri dari 7.108 data sel darah tunggal yang dikategorikan ke dalam 9 kelas berdasarkan jenis sel. Kategori tersebut antara lain 354 sel akantosit, 1.121 sel *elliptocyte*, 222 sel hipokromik, 1.426 sel normal, 24 sel pensil, 562 sel sferis, 382 sel stomatosit, 851 sel target, dan 2.076 sel *teardrop*.

2.2 Perancangan Sistem

Pada penelitian ini akan dilakukan 2 tahapan klasifikasi, dimana pada klasifikasi pertama kelas pensil akan dimasukkan ke dalam kelas *elliptocyte*. Sehingga tujuan dari klasifikasi pertama adalah membagi data menjadi 8 kelas, yaitu *acantocyte*, *elliptocyte*, *hypochromic*, normal, *sperocyte*, *stomatocyte*, target, dan *teardrop*. Selanjutnya pada klasifikasi kedua, kelas *elliptocyte* akan dibagi menjadi kelas *elliptocyte* dan pensil. Selain itu, akan dibuat sistem pembanding yang langsung membagi sel darah merah menjadi 9 kelas dengan CNN.

2. 2.1 Perancangan sistem klasifikasi pertama dengan CNN

Pada Gambar 1 mengilustrasikan langkah-langkah yang terlibat dalam proses ini, dimana CNN menggunakan lapisan konvolusi untuk mengekstrak fitur gambar. Klasifikasi kemudian dilakukan dengan metode MLP (Multi-Layer Perceptron), mengkategorikan data gambar ke dalam delapan kelas. Kelas-kelas tersebut antara lain kelas sel akantosit, sel eliptosit, sel hipokromik, sel normal, sel sferis, sel stomatosit, sel target, dan sel *teardrop*. Kelas sel pensil dikategorikan ke dalam kelas sel eliptosit dalam tahapan ini.

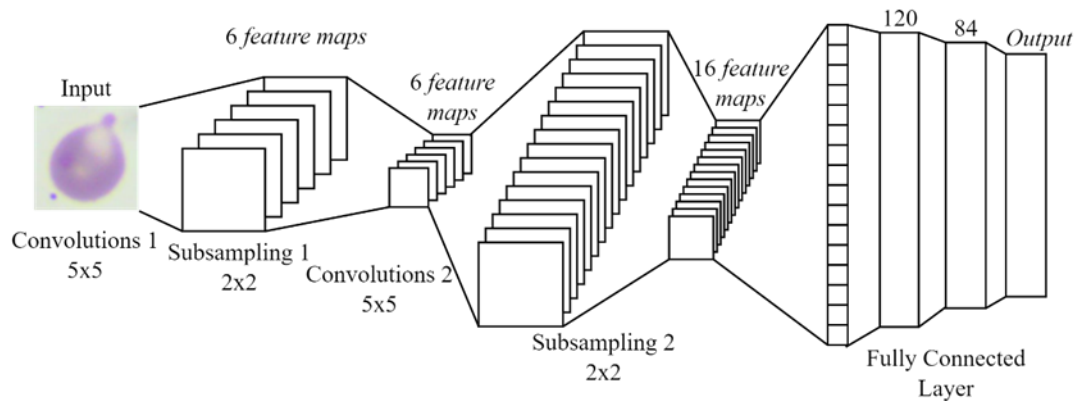


Gambar 1 Diagram Alir Proses Klasifikasi Pertama dengan CNN

Penulis memilih metode Convolutional Neural Network (CNN) karena kesederhanaannya dalam mengklasifikasikan sel darah merah dasar berbentuk elips, menghindari kerumitan pada pencarian *hand-crafted feature* untuk setiap kelas. Dalam proses klasifikasi awal, dengan dataset yang telah tersegregasi, tahap pra-pemrosesan melibatkan perubahan ukuran dan zero padding untuk mengatasi berbagai rasio gambar. Zero padding menambahkan lapisan hitam untuk menstandarisasi dimensi, diikuti dengan mengubah ukuran menjadi 32x32 piksel. Khusus untuk data citra sel pensil, akan dilakukan augmentasi. Pada proses ini, citra sel pensil akan dirotasi sebesar 90°, 180°, 270° dan direfleksikan (rotasi 180° pada sumbu horizontal).

Studi ini menggunakan dua arsitektur CNN, Lenet-5 dan Mini VGGNet, dengan 50 dan 100 iterasi, memanfaatkan *hyperparameter* seperti *learning rate* 0,001, 128 *batch*, dan *Adam Optimizer*. Arsitektur Lenet-5 seperti pada Gambar 2 terdiri dari dua lapisan konvolusi diikuti oleh lapisan subsampling dan diakhiri dengan *fully connected layer*. Studi ini secara metodis membandingkan efektivitas pendekatan-pendekatan ini.

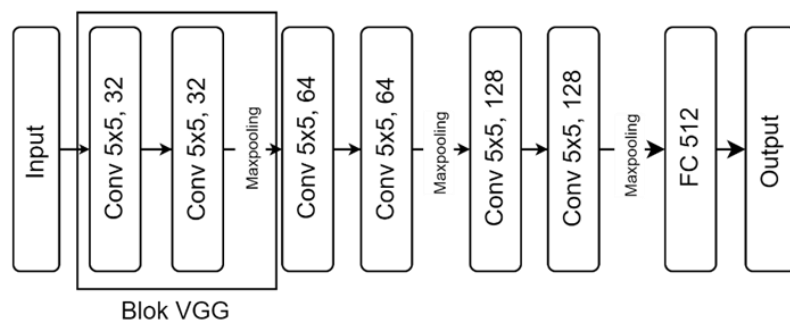
Pendekatan komprehensif ini mengoptimalkan klasifikasi sel darah merah berbentuk elips, menekankan keseimbangan antara kesederhanaan dan efektivitas dalam aplikasi CNN, pra-pemrosesan, dan pemilihan arsitektur.



Gambar 2 Model CNN Lenet 5

Arsitektur Lenet-5 memiliki dua lapisan konvolusi dengan 6 dan 16 filter, masing-masing menggunakan kernel 5x5 dan fungsi aktivasi relu untuk ekstraksi fitur gambar. Selanjutnya, dilakukan proses pooling mengurangi ukuran gambar menggunakan *average pooling* dengan kernel 2x2. *Feature maps* yang dihasilkan berfungsi sebagai masukan untuk *fully connected layer* dalam *multi-layer perceptron*. Sebelum memasuki *fully connected layer*, dilakukan *flattening* untuk mengubah tensor menjadi bentuk 1 dimensi. Fungsi *dense* mewakili *multi-layer perceptron*, dengan dua *hidden layer* yang memiliki 120 dan 84 neuron [10], kedua layer ini menggunakan fungsi aktivasi relu. Lapisan terakhir dari *fully connected layer* menggunakan fungsi aktivasi softmax untuk pembagian kelas dengan nilai probabilitas input citra.

Arsitektur MiniVGGNet terdiri dari 3 VGG blocks diikuti oleh *fully connected layer* [11]. Setiap VGG blocks disini mengintegrasikan dua lapisan konvolusi dan satu lapisan subsampling, sehingga berkontribusi terhadap ekstraksi fitur yang efektif dalam penelitian ini. Pada Gambar 3 dapat dilihat susunan arsitektur MiniVGGNet.



Gambar 3 Model MiniVGGNet

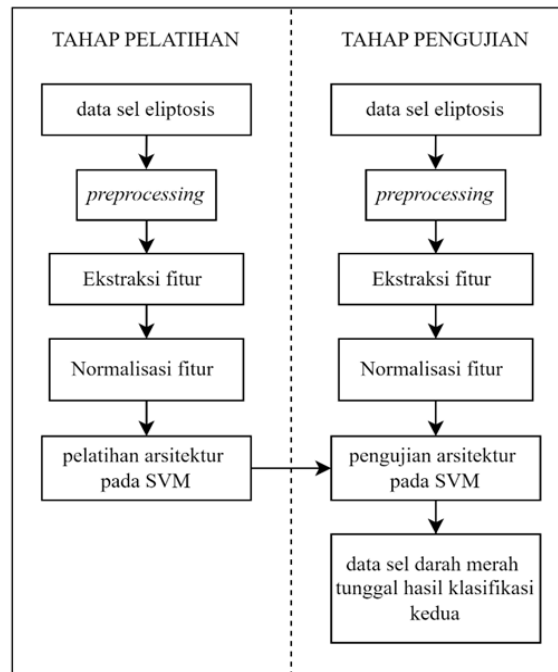
Arsitektur terdiri dari tiga blok VGG. Tiap-tiap blok memiliki 2 layer konvolusi dengan kernel 5x5 yang diikuti dengan layer *max pooling* (2x2). Hal ini menghasilkan 32, 64, dan 128 *features map* untuk blok pertama, kedua, dan ketiga.

Setelah layer *pooling* terakhir, nilai *features map* yang didapatkan berfungsi sebagai masukan untuk *fully connected layer* yang merupakan klasifier *multi-layer perceptron*. Sebelum itu, dilakukan proses *flattening* dengan mengubah tensor menjadi bentuk 1 dimensi tanpa mengubah ukurannya. Untuk *fully connected layer*, digunakan fungsi aktivasi relu dan 512 neuron. Kombinasi ini memiliki akurasi yang lebih unggul dibandingkan arsitektur VGG konvensional dengan dua *fully connected layer* [12].

Lapisan terakhir menggunakan fungsi aktivasi softmax untuk pembagian kelas berdasarkan nilai probabilitas. Selanjutnya seluruh anggota kelas elliptosis menjalani proses klasifikasi kedua.

2. 2.2 Perancangan sistem klasifikasi kedua dengan SVM

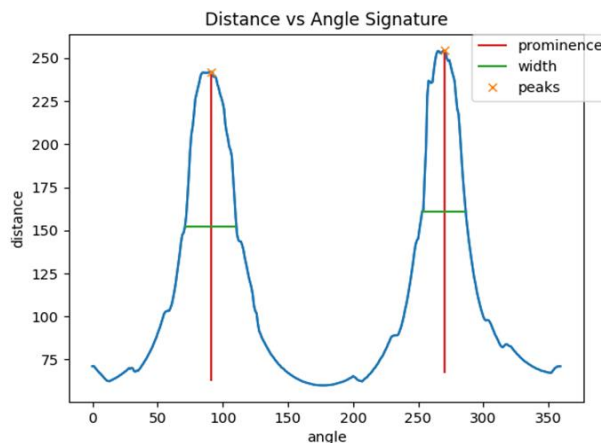
Setelah klasifikasi awal, sel eliptosis dipilih untuk proses klasifikasi kedua. Diagram alir merinci klasifikasi kedua ini dapat dilihat pada Gambar 4. Kumpulan data sel darah merah tunggal disegmentasi dan menjalani pra-pemrosesan, yang melibatkan perubahan ukuran dan *zero padding* untuk standardisasi.



Gambar 4 Diagram Alir Proses Klasifikasi Kedua

Selanjutnya, fitur-fitur gambar diekstraksi, termasuk karakteristik geometris seperti luas, keliling, sumbu mayor dan minor, serta DAS (*Distance Angle Signature*) yang merepresentasikan bentuk sel. *Contouring* digunakan untuk fitur area dan perimeter, menggabungkan tepi untuk membentuk suatu objek (kontur) dan memperoleh keliling dari *outline* piksel dan luas dari semua piksel di dalam *outline*.

Fitur *DAS*, axis mayor, dan axis minor diekstraksi dengan mencari titik pusat citra lalu membaginya menjadi 36 bagian (berjarak 10°). Selanjutnya ditarik garis *distance* (d) antara titik pusat citra dan kontur tepinya. Garis *distance* (d) terpanjang yang membentuk sudut 180° dilambangkan menjadi axis mayor. Garis *distance* (d) yang tegak lurus dengan axis mayor menjadi axis minor.



Gambar 5 Contoh Grafik DAS Sel Pensil

Dari Gambar 5 dapat diekstraksi sembilan fitur *Ekstrak* yang mewakili grafik *Distance versus Angle Signature* dengan persamaan (1) selisih tinggi bukit (Δp), persamaan (2) selisih lebar bukit (Δw), persamaan (3) rerata tinggi bukit (\bar{p}), persamaan (4) rerata lebar bukit (\bar{w}), persamaan (5) simpangan baku tinggi bukit (STD_p), persamaan (6) simpangan baku lebar bukit (STD_w), persamaan (7) rerata puncak bukit (\overline{pks}), persamaan (8) selisih puncak bukit (Δpks), dan persamaan (9) rerata seluruh distance (\bar{d}). [3]

$$\Delta p = \frac{1}{n-1} \sum_{k=1}^n |p_k + p_{k+1...n}| \quad (1)$$

$$\Delta w = \frac{1}{n-1} \sum_{k=1}^n |w_k + w_{k+1...n}| \quad (2)$$

$$\bar{p} = \frac{1}{n} \sum_{k=1}^n p_k \quad (3)$$

$$\bar{w} = \frac{1}{n} \sum_{k=1}^n w_k \quad (4)$$

$$STD_p = \frac{1}{n-1} \sum_{k=1}^n |p_k - \bar{p}|^2 \quad (5)$$

$$STD_w = \frac{1}{n-1} \sum_{k=1}^n |w_k - \bar{w}|^2 \quad (6)$$

$$\Delta pks = \frac{1}{n-1} \sum_{k=1}^n |pks_k + pks_{k+1...n}| \quad (7)$$

$$\overline{pks} = \frac{1}{n} \sum_{k=1}^n pks_k \quad (8)$$

$$\bar{d} = \frac{1}{360} \sum_{k=1}^{360} d_k \quad (9)$$

Dari luas dan keliling citra, diekstraksi nilai *feature descriptor* yang menjadi fitur dengan persamaan (1) *eccentricity*, persamaan (2) *compactness*, persamaan (3) *circularity*, dan persamaan (4) *ratio* [11], yang berfungsi sebagai karakteristik bentuk sel darah merah.

$$circularity = \frac{4\pi A}{p^2} \quad (1)$$

$$compactness = \frac{p^2}{A} \quad (2)$$

$$ratio = \frac{axis\ mayor}{axis\ minor} \quad (3)$$

$$eccentricity = \sqrt{1 - \left(\frac{1}{ratio}\right)^2} \quad (4)$$

Secara keseluruhan fitur-fitur yang digunakan dapat dilihat pada tabel 1 berikut.

Tabel 1 Fitur SVM

Jenis Fitur	Keterangan	Fitur	Jumlah
<i>Geo</i>	Parameter Geometis Sel	<i>eccentricity, compactness, circularity, dan ratio</i>	4
<i>Ekstrak</i>	Morfologi DAS Sel	selisih tinggi bukit (Δp), selisih lebar bukit (Δw), rerata tinggi bukit (\bar{p}), rerata lebar bukit (\bar{w}), simpangan baku tinggi bukit (STD_p), simpangan baku lebar bukit (STD_w), rerata puncak bukit (\overline{pks}), selisih puncak bukit (Δpks), dan rerata seluruh distance (\bar{d}).	9
<i>Sudut</i>	Jarak dari titik pusat ke tepi sel		36

Fitur-fitur tersebut kemudian dinormalisasi menggunakan normalisasi min-max sebelum dimasukkan ke dalam classifier Support Vector Machine (SVM). Proses klasifikasi dengan SVM melibatkan perkalian *hyperplane* untuk setiap input, dengan hyperparameter yang ditentukan meliputi variasi *cost* (c), *degree* (d) untuk kernel polinomial, dan γ untuk kernel *Radial Basis Function* (RBF).

2.3 Analisis dan Pengujian Sistem

Setelah mengimplementasikan sistem, analisis dan pengujian dilakukan untuk mengevaluasi fungsionalitasnya untuk kebutuhan pengguna dan kemampuannya dalam memproses data. Mengevaluasi kinerja sistem klasifikasi sangatlah penting, dan *Confusion Matrix* adalah metode yang digunakan untuk tujuan ini. Matriks tersebut membandingkan hasil klasifikasi sistem dengan hasil yang diharapkan. Dalam pengukuran kinerja berbasis *Confusion Matrix*, terdapat empat istilah: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). TN mewakili data negatif yang terdeteksi dengan benar, FP adalah data negatif yang terdeteksi sebagai positif, TP adalah data positif yang terdeteksi dengan benar, dan FN adalah data positif yang terdeteksi sebagai negatif.

2.3 Rancangan Pengujian

Sistem klasifikasi sel darah merah mengalami pengujian untuk menilai akurasi, sensitivitas, dan tingkat ketelitian sistem pada setiap tahap klasifikasi. Data uji dari delapan kelas diberikan ke sistem CNN, diikuti dengan analisis kinerja. Kelas *elliptocyte* kemudian diekstraksi dan menjalani tahap klasifikasi kedua dengan SVM, dilanjutkan dengan perbandingan kinerja antara sistem yang mengklasifikasikan langsung ke dalam sembilan kelas dan yang menggunakan proses klasifikasi dua tahap. Uji kinerja sistem secara keseluruhan mengevaluasi nilai akurasi (1), presisi (2), sensitivitas (*recall*) (3), dan Skor F1 (4) [13], menggunakan 25% data gambar sel darah merah dari kumpulan data untuk pengujian.

$$Akurasi = \frac{\sum_{i=1}^n \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{n} \times 100\% \quad (1)$$

$$Presisi = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FP_i)} \times 100\% \quad (2)$$

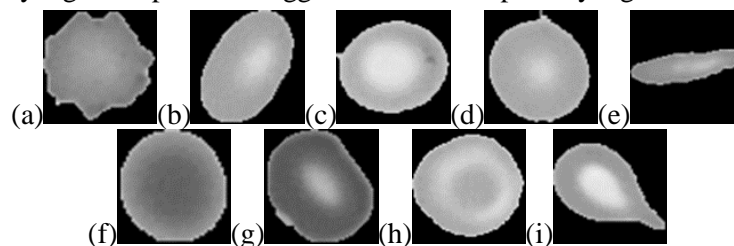
$$Recall = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FN_i)} \times 100\% \quad (3)$$

$$F1 \text{ Score} = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Dataset Pengujian

Penelitian disini melibatkan 1.822 gambar sel darah merah yang dikategorikan ke dalam sembilan kelas. Hasil yang ditampilkan menggunakan data sel pensil yang telah diaugmentasi.



Gambar 6 sel darah merah (a) akantosit, (b) *elliptocyte*, (c) hipokromik, (d) normal, (e) pensil, (f) sferis, (g) stomatosit, (h) target, (i) *teardrop*.

3.2 Tahapan Klasifikasi Pertama dengan CNN untuk 8 Kelas Sel

Pada tahap awal klasifikasi, kumpulan data citra sel darah merah dikategorikan ke dalam delapan kelas: kelas *acantocyte*, *elliptocyte*, *hypochromic*, normal, *sperocyte*, *stomatocyte*, target, dan *teardrop*. Selanjutnya data dari kelas *elliptocyte* digunakan pada tahap klasifikasi selanjutnya. Empat percobaan dilakukan, memvariasikan kombinasi arsitektur CNN dan jumlah iterasi. Secara khusus, Percobaan 1 menggunakan Arsitektur CNN Lenet-5 dengan 50 iterasi, Percobaan 2 menggunakan Arsitektur CNN Lenet-5 dengan 100 iterasi, sedangkan Percobaan 3 dan 4 mengadopsi Arsitektur MiniVGGNet dengan masing-masing 50 dan 100 iterasi.

3. 2.1 CNN Lenet-5

Tabel 2 Hasil Percobaan menggunakan Lenet-5

Kelas	Percobaan 1			Percobaan 2			Data uji
	Presisi (%)	Recall (%)	Skor F1 (%)	Presisi (%)	Recall (%)	Skor F1 (%)	
Elliptocyte	95,82	98,01	96,90	94,02	98,58	96,24	351
Hypochromic	74,51	67,86	71,03	75,51	66,07	70,48	56
Normal	92,86	94,68	93,76	93,87	94,40	94,13	357
Sperocyte	97,90	99,29	98,59	100,00	99,29	99,64	141
Stomatocyte	92,68	79,17	85,39	88,75	73,96	80,68	96
Target	94,88	95,77	95,33	94,14	98,12	96,09	213
Teardrop	99,22	98,65	98,94	99,61	98,27	98,93	519
macro avg	92,81	91,40	92,01	92,69	90,94	91,68	1822
weighted avg	95,38	95,44	95,37	95,29	95,39	95,28	
Akurasi total	95,44			95,39			

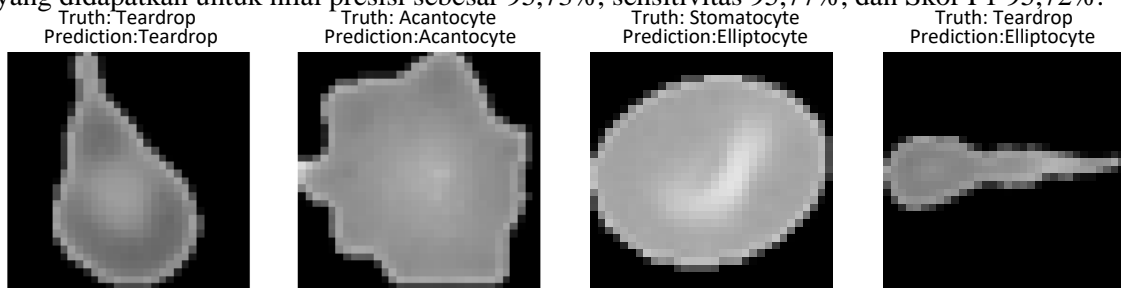
Untuk percobaan dengan Model Lenet-5 terlihat perbandingan Percobaan 1 (50 iterasi) dan Percobaan 2 (100 iterasi) pada Tabel 2, Percobaan 1 memiliki performa yang lebih baik daripada Percobaan 2 dengan akurasi sebesar 95,44%. Performa *weighted average* atau rata-rata sesuai bobot kelas yang didapatkan adalah nilai presisi 95,38%, sensitivitas 95,44%, dan Skor F1 95,37%. Sedangkan Percobaan 2 memiliki akurasi sebesar 95,39% dengan performa *weighted average* untuk nilai presisi sebesar 95,29%, sensitivitas 95,39%, dan Skor F1 95,28%.

3. 2.2 CNN MiniVGGNet

Tabel 3 Hasil Percobaan menggunakan MiniVGGNet

Kelas	Percobaan 3			Percobaan 4			Data uji
	Presisi (%)	Recall (%)	Skor F1 (%)	Presisi (%)	Recall (%)	Skor F1 (%)	
Elliptocyte	94,26	98,29	96,23	96,05	96,87	96,45	351
Hypochromic	79,63	76,79	78,18	90,70	69,64	78,79	56
Normal	94,99	95,52	95,25	93,30	97,48	95,34	357
Sperocyte	96,53	98,58	97,54	99,28	97,87	98,57	141
Stomatocyte	88,10	77,08	82,22	90,48	79,17	84,44	96
Target	95,77	95,77	95,77	95,37	96,71	96,04	213
Teardrop	99,61	98,84	99,23	98,29	99,42	98,85	519
macro avg	93,47	92,19	92,77	95,29	91,86	93,35	1822
weighted avg	95,73	95,77	95,72	96,00	96,05	95,95	
Akurasi total	95,77			96,05			

Pada percobaan dengan Model MiniVGGNet terlihat di perbandingan Percobaan 3 (50 iterasi) dan Percobaan 4 (100 iterasi) pada Tabel 3, Percobaan 4 memiliki performa yang lebih baik dengan akurasi sebesar 96,05%. Performa *weighted average* atau rata-rata sesuai bobot kelas yang didapatkan adalah nilai presisi 96,00%, sensitivitas (*recall*) 96,05%, dan Skor F1 95,95%. Sedangkan Percobaan 3 memiliki akurasi sebesar 95,77% dengan performa *weighted average* yang didapatkan untuk nilai presisi sebesar 95,73%, sensitivitas 95,77%, dan Skor F1 95,72%.



Gambar 7 Hasil Prediksi Percobaan 4 (MiniVGGNet 100 iterasi)

3. 3 Klasifikasi dengan CNN Pembandingan untuk 9 Kelas Sel

Pada bagian ini dilakukan eksperimen komparatif yang melibatkan segmentasi dataset citra sel darah merah menjadi sembilan kelas. Khususnya, kelas sel *pencil*, yang sebelumnya dikelompokkan dengan kelas sel *elliptocyte*, kini ditetapkan sebagai kelas berbeda..

3. 3.1 CNN Lenet-5

Tabel 4 Hasil Percobaan menggunakan Lenet-5

Kelas	Percobaan 1			Percobaan 2			Data uji
	Presisi (%)	Recall (%)	Skor F1 (%)	Presisi (%)	Recall (%)	Skor F1 (%)	
Elliptocyte	95,45	97,03	96,24	96,07	96,70	96,38	303
Hypochromic	79,59	69,64	74,29	73,68	75,00	74,34	56
Normal	95,21	94,68	94,94	96,01	94,40	95,20	357
<i>Pencil</i>	90,20	95,83	92,93	82,76	100,00	90,57	48
Sperocyte	97,90	99,29	98,59	99,29	98,58	98,93	141
Stomatocyte	83,33	78,13	80,65	81,72	79,17	80,42	96
Target	92,73	95,77	94,23	91,82	94,84	93,30	213
Teardrop	98,84	98,84	98,84	99,61	97,50	98,54	519
macro avg	91,69	91,01	91,29	90,97	92,78	91,78	1822
weighted avg	94,81	94,90	94,83	95,11	95,01	95,03	
Akurasi total	94,90			95,01			

Untuk percobaan pembandingan dengan Model Lenet-5 terlihat perbandingan Percobaan 1 (50 iterasi) dan Percobaan 2 (100 iterasi) pada Tabel 4, Percobaan 2 memiliki performa yang lebih baik daripada Percobaan 1 dengan akurasi sebesar 95,01%, Performa kelas *Pencil* yang didapatkan adalah nilai presisi 82,76%, sensitivitas (*recall*) 100,00%, dan Skor F1 90,57%. Sedangkan Percobaan 1 memiliki akurasi sebesar 94,90% dengan performa kelas *Pencil* yang didapatkan untuk nilai presisi 90,20%, sensitivitas (*recall*) 95,83%, dan Skor F1 92,93%.

3. 3.2 CNN MiniVGGNet

Tabel 5 Hasil Percobaan menggunakan MiniVGGNet

Kelas	Percobaan 3			Percobaan 4			Data uji
	Presisi (%)	Recall (%)	Skor F1 (%)	Presisi (%)	Recall (%)	Skor F1 (%)	
Elliptocyte	97,25	93,40	95,29	94,86	97,36	96,09	303
Hypochromic	72,86	91,07	80,95	84,78	69,64	76,47	56
Normal	95,04	96,64	95,83	95,03	96,36	95,69	357
<i>Pencil</i>	91,84	93,75	92,78	90,57	100,00	95,05	48
Sperocyte	97,89	98,58	98,23	98,57	97,87	98,22	141
Stomatocyte	87,91	83,33	85,56	95,29	84,38	89,50	96
Target	96,14	93,43	94,76	94,59	98,59	96,55	213
Teardrop	97,91	99,23	98,56	99,22	98,46	98,84	519
macro avg	92,85	93,51	93,04	94,39	93,13	93,61	1822
weighted avg	95,62	95,44	95,48	96,08	96,10	96,03	
Akurasi total	95,44			96,10			

Pada percobaan dengan Model MiniVGGNet terlihat di perbandingan Percobaan 3 (50 iterasi) dan Percobaan 4 (100 iterasi) pada Tabel 5, Percobaan 4 memiliki performa yang lebih baik dengan akurasi sebesar 96,00%. Performa kelas *Pencil* yang didapatkan adalah nilai presisi 90,57%, sensitivitas (*recall*) 100,00%, dan Skor F1 95,05%. Sedangkan Percobaan 3 memiliki akurasi sebesar 95,44% dengan performa kelas *Pencil* yang didapatkan untuk nilai presisi sebesar 91,84%, sensitivitas (*recall*) 93,75%, dan Skor F1 92,78%. Untuk kelas *Pencil*, Percobaan 4 memiliki performa yang lebih baik.

3. 4 Tahapan Klasifikasi Kedua dengan SVM

Pada klasifikasi tahap kedua, data uji yang awalnya diberi label *elliptocyte* (Elips) diklasifikasi menjadi kelas sel Elips dan sel *pencil* (Pensil). Data uji ini akan diklasifikasi menggunakan Support Vector Machine (SVM) dengan tiga kombinasi fitur. Kombinasi pertama, bernama *Geo*, melibatkan fitur geometris seperti *eccentricity*, *compactness*, *circularity*, dan rasio sel. Kombinasi kedua, *Geo+Sudut*, menggabungkan *Geo* dengan fitur *Distance versus Angle Signature* (DAS) murni. Kombinasi *Geo+Peaks*, menggabungkan *Geo* dengan fitur yang diekstrak dari DAS, sehingga menghasilkan sembilan fitur. Hasil pengujian dapat dilihat pada tabel 6, 7, dan 8 berikut.

3. 4.1 SVM Linear

Tabel 6 Hasil SVM Linear

No	Jenis Fitur	Cost	Akurasi (%)	Sensitivitas (%)		Presisi (%)		Skor F1 (%)	
				Pensil	Elips	Pensil	Elips	Pensil	Elips
1	<i>Geo</i>	10	94,15	63,83	98,72	88,24	94,77	74,07	96,70
	<i>Geo + Sudut</i>	1	94,71	59,57	100,00	100,00	94,26	74,67	97,05
	<i>Geo + Peaks</i>	10	91,92	72,34	94,87	68,00	95,79	70,10	95,33
2	<i>Geo</i>	10	94,57	62,50	99,38	93,75	94,64	75,00	96,95
	<i>Geo + Sudut</i>	100	94,84	60,42	100,00	100,00	94,40	75,32	97,12
	<i>Geo + Peaks</i>	10	92,93	72,92	95,94	72,92	95,94	72,92	95,94
3	<i>Geo</i>	10	94,54	62,50	99,37	93,75	94,61	75,00	96,93
	<i>Geo + Sudut</i>	100	94,81	60,42	100,00	100,00	94,36	75,32	97,10
	<i>Geo + Peaks</i>	10	92,62	72,92	95,60	71,43	95,90	72,16	95,75
4	<i>Geo</i>	10	94,35	62,50	99,35	93,75	94,41	75,00	96,82
	<i>Geo + Sudut</i>	100	94,63	60,42	100,00	100,00	94,15	75,32	96,99
	<i>Geo + Peaks</i>	10	94,63	72,92	98,04	85,37	95,85	78,65	96,93

Pada tabel 6 di atas dapat dilihat bahwa performa terbaik dari keseluruhan data didapatkan dengan fitur *Geo+Peaks* pada nomor percobaan 4 dengan nilai sensitivitas sebesar 72,92%, presisi 85,37%, dan Skor F1 78,65% ($cost=10$). Pengklasifikasi SVM Linear masih memiliki nilai sensitivitas yang lebih rendah daripada nilai kepresisiannya. Nilai kepresisian yang tinggi artinya ada sedikit sel *elliptocyte* yang terdeteksi sebagai sel pensil (*False Positive*). Namun, masih ada sel pensil yang terdeteksi sebagai sel *elliptocyte* (*False Negative*) sehingga nilai sensitivitasnya sangat rendah. Hal ini menyebabkan nilai F1 score yang rendah.

3. 4.2 SVM Polinomial

Tabel 7 Hasil SVM Polinomial

No	Jenis Fitur	Parameter		Akurasi (%)	Sensitivitas (%)		Presisi (%)		Skor F1 (%)	
		cost	degree		Pensil	Elips	Pensil	Elips	Pensil	Elips
1	<i>Geo</i>	10	2	99,72	100,00	99,68	97,92	100,00	98,95	99,84
	<i>Geo + Sudut</i>	0,1	2	94,99	61,70	100,00	100,00	94,55	76,32	97,20
	<i>Geo + Peaks</i>	100	2	98,89	100,00	98,72	92,16	100,00	95,92	99,35
2	<i>Geo</i>	10	2	99,73	100,00	99,69	97,96	100,00	98,97	99,84
	<i>Geo + Sudut</i>	1	1	94,84	60,42	100,00	100,00	94,40	75,32	97,12
	<i>Geo + Peaks</i>	10	3	99,18	100,00	99,06	94,12	100,00	96,97	99,53
3	<i>Geo</i>	10	2	99,73	100,00	99,69	97,96	100,00	98,97	99,84
	<i>Geo + Sudut</i>	0,1	2	94,81	60,42	100,00	100,00	94,36	75,32	97,12
	<i>Geo + Peaks</i>	10	2	99,18	100,00	99,06	94,12	100,00	96,97	99,53
4	<i>Geo</i>	100	2	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	<i>Geo + Sudut</i>	1	1	94,81	60,42	100,00	100,00	94,15	75,32	96,99
	<i>Geo + Peaks</i>	10	2	99,44	100,00	99,35	96,00	100,00	97,96	99,67

Pada tabel 7 di atas dapat dilihat bahwa performa terbaik dari keseluruhan data didapatkan dengan fitur *Geo* pada nomor percobaan 4 dengan nilai sensitivitas sebesar 100,00%, presisi 100,00%, dan Skor F1 100,00% ($cost=100$; $degree=2$).

Pengklasifikasi SVM Polinomial memiliki performa yang cukup baik dengan rata-rata nilai Skor F1 untuk sel elips pada semua percobaan mencapai 98,84%. Sistem juga berhasil mendeteksi sel pensil dengan sensitivitas dan kepresisian sebesar 100,00%. Nilai sensitivitas yang tinggi artinya tidak ada sel pensil yang terdeteksi sebagai sel elips dan tidak ada sel elips yang terdeteksi sebagai sel pensil.

3. 4.3 SVM RBF

Tabel 8 Hasil SVM RBF

No	Jenis Fitur	Parameter		Akurasi (%)	Sensitivitas (%)		Presisi (%)		Skor F1 (%)	
		cost	gamma		Pensil	Elips	Pensil	Elips	Pensil	Elips
1	<i>Geo</i>	100	1	99,72	100,00	99,68	97,92	100,00	98,95	99,84
	<i>Geo + Sudut</i>	1	1	95,26	65,96	99,68	96,88	95,11	78,48	97,34
	<i>Geo + Peaks</i>	100	0,1	98,33	100,00	98,08	88,68	100,00	94,00	99,03
2	<i>Geo</i>	100	1	99,73	100,00	99,69	97,96	100,00	98,97	99,84
	<i>Geo + Sudut</i>	1	1	95,11	64,58	99,69	96,88	94,94	77,50	97,26
	<i>Geo + Peaks</i>	100	0,1	99,18	100,00	99,06	94,12	100,00	96,97	99,53
3	<i>Geo</i>	100	1	99,73	100,00	99,69	97,96	100,00	98,97	99,84
	<i>Geo + Sudut</i>	1	1	95,08	64,58	99,69	96,88	94,91	77,50	97,24
	<i>Geo + Peaks</i>	100	0,1	99,18	100,00	99,06	94,12	100,00	96,97	99,53
4	<i>Geo</i>	100	1	99,72	100,00	99,67	97,96	100,00	98,97	99,84
	<i>Geo + Sudut</i>	1	1	94,92	64,58	99,67	96,88	94,72	77,50	97,13
	<i>Geo + Peaks</i>	100	0,1	99,15	100,00	99,02	94,12	100,00	96,97	99,51

Pada tabel 8 di atas dapat dilihat bahwa performa terbaik dari keseluruhan data didapatkan dengan fitur *Geo* pada nomor percobaan 2,3, dan 4 dengan nilai sensitivitas sebesar 100,00%, presisi 97,96%, dan Skor F1 98,97% ($cost=100$; $gamma=1$). Pengklasifikasi SVM RBF memiliki performa yang cukup baik dengan rata-rata nilai Skor F1 untuk sel elips pada semua percobaan mencapai 98,83%. Sistem juga berhasil mendeteksi sel pensil dengan sensitivitas sebesar 100,00%. Nilai sensitivitas yang tinggi artinya tidak ada sel pensil yang terdeteksi sebagai sel elips (*False Negative*). Namun masih ada sel elips yang terdeteksi sebagai sel pensil.

Hasil percobaan dengan kombinasi CNN MiniVGGNet dan SVM Polinomial performa yang sangat baik dengan nilai presisi 100,00%, nilai sensitivitas 100,00%, dan F1 score 100,00% baik untuk kelas *Pencil* (Pensil) maupun kelas *Elliptocyte* (Elips). Performa ini juga melampaui performa sistem pembandingan yang hanya melewati 1 tahapan klasifikasi CNN, dengan performa terbaik untuk kelas *Pencil* (Pensil) dengan nilai presisi sebesar 91,84%, sensitivitas (*recall*) 93,75%. Penelitian ini juga menunjukkan peningkatan performa yang cukup signifikan terutama dalam mengklasifikasikan kelas *Pencil* (Pensil) jika dibandingkan dengan penelitian sebelumnya [3] yang memiliki nilai sensitivitas kelas *Pencil* (Pensil) sebesar 0,00%.

4. KESIMPULAN

Sistem klasifikasi bertingkat yang menggunakan pengklasifikasi CNN MiniVGGNet dengan 100 iterasi pada tahapan klasifikasi pertama dan SVM Polinomial pada tahapan klasifikasi kedua memiliki performa yang lebih baik daripada sistem yang hanya melewati satu tahapan klasifikasi CNN untuk mengklasifikasi kelas *Pencil* (Pensil). Pada tahapan klasifikasi pertama, sistem mendapatkan nilai akurasi 96,05% dengan rerata nilai kepresisian 96,00%, sensitivitas 96,05%, dan Skor F1 95,95%. Keluaran klasifikasi pertama ini direklasifikasi sehingga mendapatkan nilai presisi 100,00%, nilai sensitivitas 100,00%, dan nilai F1 score 100,00% untuk kelas *Pencil* (Pensil). Performa ini menunjukkan peningkatan sensitivitas sebesar 100,00% daripada penelitian sebelumnya dalam mengklasifikasikan kelas *Pencil* (Pensil) walaupun jumlah dataset sangatlah terbatas. Kombinasi fitur *Geo* yang terdiri dari *eccentricity*, *compactness*, *circularity*, dan *ratio* memiliki performa terbaik dalam mengklasifikasikan kelas *Pencil* (Pensil) dari kelas *Elliptocyte* (Elips) dengan pengklasifikasi SVM Polinomial dibandingkan dengan sistem serupa yang ditambahkan fitur Distance versus Angle Signature (DAS).

5. SARAN

Terdapat beberapa saran untuk pengembangan dan penelitian lebih lanjut adalah sebagai berikut. Perluas metode klasifikasi bertingkat ke jenis sel lain dengan kinerja rendah dalam klasifikasi satu tahap. Misalnya, mengelompokkan sel berdasarkan bentuk dasar seperti bulat, bertekstur, dll terlebih dahulu. Meningkatkan penggunaan dataset sel darah merah, khususnya saat menggunakan pembelajaran mesin dalam proses klasifikasi, untuk mengoptimalkan proses pelatihan sistem. Pertimbangkan untuk menerapkan hasil proses pelatihan sistem ke data mentah yang tidak tersegmentasi untuk mengevaluasi kinerja sebenarnya.

DAFTAR PUSTAKA

- [1] Tyas, D. A., Hartati, S., Harjoko, A., & Ratnaningsih, T. (2019). Erythrocyte Classification using Multi Layer Perceptron, Naïve Bayes Classifier, RBF Network and SVM. *International Journal of Engineering and Advanced Technology*, 9(2), 2024–2028. <https://doi.org/10.35940/ijeat.b3231.129219>
- [2] Tyas, D. A., Ratnaningsih, T., Harjoko, A., & Hartati, S. (2017). The Classification of Abnormal Red Blood Cell on The Minor Thalassemia Case Using Artificial Neural Network and Convolutional Neural Network. *Proceedings of the International Conference on Video and Image Processing*, 228–233. <https://doi.org/10.1145/3177404.3177438>
- [3] Tyas, D. A., Hartati, S., Harjoko, A., & Ratnaningsih, T. (2020). Morphological, Texture, and Color Feature Analysis for Erythrocyte Classification in Thalassemia Cases. *IEEE Access*, 8, 69849–69860. <https://doi.org/10.1109/ACCESS.2020.2983155>
- [4] Chy, T. S., & Anisur Rahaman, M. (2019). A comparative analysis by KNN, SVM ELM classification to detect sickle cell anemia. 1st International Conference on Robotics, Electrical and Signal Processing Techniques, ICREST 2019, 455–459. <https://doi.org/10.1109/ICREST.2019.8644410>
- [5] Hartati, S., Harjoko, A., Rosnelly, R., Chandradewi, I., & Faizah. (2019). Performance of SVM and ANFIS for classification of malaria parasite and its life-cycle-stages in blood smear. In *Communications in Computer and Information Science* (Vol. 937, Issue C, pp. 110–121). https://doi.org/10.1007/978-981-13-3441-2_9
- [6] Sen, B., Ganesh, A., Bhan, A., & Dixit, S. (2021). Deep Learning based diagnosis of sickle cell anemia in human RBC. *Proceedings of 2021 2nd International Conference on Intelligent Engineering and Management, ICIEM 2021*, Iciev, 526–529. <https://doi.org/10.1109/ICIEM51511.2021.9445293>
- [7] Harrington, A. M., Ascp, M. T., Ward, P. C. J., & Kroft, S. H. (2008). Iron Deficiency Anemia, β -Thalassemia Minor, and Anemia of Chronic Disease A Morphologic Reappraisal. 466–471. <https://doi.org/10.1309/LY7YLUPE7551JYBG>
- [8] Safca, N., Popescu, D., Ichim, L., Elkhatib, H., & Chenaru, O. (2018). Image processing techniques to identify red blood cells. *2018 22nd International Conference on System Theory, Control and Computing, ICSTCC 2018 - Proceedings*, 93–98. <https://doi.org/10.1109/ICSTCC.2018.8540708>
- [9] Tyas, D. A., Ratnaningsih, T., Harjoko, A., & Hartati, S. (2022). Erythrocyte (red blood cell) dataset in thalassemia case. *Data in Brief*, 41, 107886. <https://doi.org/10.1016/j.dib.2022.107886>
- [10] Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- [11] Liu, B., Li, F., Wang, X., Zhang, B., & Yan, J. (2023). *Ternary Weight Networks*. 1–5. <https://doi.org/10.1109/icassp49357.2023.10094626>
- [12] Campos-Leal, J. A., Yee-Rendon, A., & Vega-Lopez, I. F. (2022). Simplifying VGG-16 for Plant Species Identification. *IEEE Latin America Transactions*, 20(11), 2330–2338. <https://doi.org/10.1109/TLA.2022.9904757>
- [13] Liyantoko, A. N., Candradewi, I., & Harjoko, A. (2019). Klasifikasi Sel Darah Putih dan Sel Limfoblas Menggunakan Metode Multilayer Perceptron Backpropagation. *IJEIS (Indonesian Journal of Electronics and Instrumentation Systems)*, 9(2), 173. <https://doi.org/10.22146/ijeis.49943>