

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Dalam penyusunan tugas akhir ini memerlukan penelitian terkait yang telah dilakukan sebelumnya sebagai referensi bahan uji serta perbandingan metode guna memperkuat hasil penelitian. Beberapa penelitian yang relevan dengan topik penelitian yang dilakukan saat ini dapat dilihat pada Tabel 2.1.

Pada Tabel 2.1 dituliskan (Tek dkk., 2010) melakukan penelitian tentang pendeteksi dan identifikasi apusan darah tipis penderita malaria. Jenis parasit yang digunakan pada penelitian ini terdapat 4 jenis, yaitu *P. Falciparum*, *P. Malariae*, *P. Ovale*, dan *P. Vivax* dengan metode pengambilan citra mikroskop pewarnaan giemsa. Fokus utama penelitian ini yaitu membangun sistem klasifikasi terbaik dengan membandingkan 3 metode klasifikasi sehingga dapat digunakan kembali pada citra apusan darah tipis yang berbeda. Untuk mengatasi tingkat keberagaman pada citra yang berbeda, sebelum masuk ke klasifikasi sistem akan memproses citra pada tahap *pre-processing* berupa perhitungan area granulometri, perbaikan pencahayaan, perbaikan skala, segmentasi latar belakang-latar depan, dan normalisasi warna. Setelah melewati tahap *pre-processing* citra akan masuk ke tahap ekstraksi ciri 83 fitur dan klasifikasi. Setelah melakukan penelitian, hasil yang didapatkan menunjukkan metode klasifikasi yang menghasilkan akurasi terbaik adalah *KNN* (*K Nearest Neighbour*) dengan presentase 93%. Dibandingkan dengan metode *FLD* (*Fisher Linear Discriminat*) sebesar 90,1% dan *BPNN* (*Back Propagation Neural Network*) sebesar 92,2%.

Penelitian dalam dunia medis untuk mendeteksi parasit malaria dalam sampel darah dengan pewarna giemsa kembali dilakukan oleh (Malihi dkk., 2013). Parasit malaria yang digunakan ada 4 jenis, yaitu *P. Falciparum*, *P. Malariae*, *P. Ovale*, dan *P. Vivax*. Konsep penelitian yang akan dijelaskan di bawah ini adalah melakukan perbandingan terhadap 5 metode klasifikasi yaitu Support Vector Machines (SVM), Nearest Mean (NM), K Nearest Neighbors (KNN), 1-NN, dan Fisher's linear discriminator untuk mengetahui akurasi terbaik. Langkah pertama yang dilakukan

pada penelitian ini adalah *pre-processing* dengan menggunakan median filter, morphology operator dan structural element untuk menghilangkan noise dan melakukan penyeragaman cahaya citra. Selanjutnya, melakukan masking pada sel darah merah karena sebagian besar parasit malaria terdapat pada sel darah merah dengan mengubah citra dalam bentuk biner dan melakukan deteksi tepi citra dengan menggunakan metode canny. Kemudian melakukan pemisahan sel darah putih, parasit dan trombosit dari elemen lain dengan menggunakan metode otsu yang diterapkan pada penggandaan dari tepi terdeteksi dan citra asli. Setelah itu melakukan ekstraksi dengan menggunakan 4 fitur berikut: gradient, flat texture, color histogram, dan area granulometry. Terakhir, sebanyak 363 citra parasit malaria dengan jenis *P. Falciparum*, *P. Vivax*, *P. Ovale*, dan *P. Malariae* dilakukan deteksi dengan 5 metode klasifikasi SVM, 1-NN, NM, KNN dan Fisher's linear discriminator. Hasil yang didapatkan adalah metode SVM memiliki tingkat akurasi sebesar 90%, metode KNN sebesar 91%, 1-NN sebesar 87%, NM sebesar 70%, dan Fisher sebesar 84%.

Selanjutnya (Wijaya & Kusumadewi, 2015) melakukan penelitian tentang penerapan Algoritma K-Means pada citra medis MRI ekstensi .png, .jpg, dan .bmp. Tujuan dilakukan penelitian ini untuk mengkompresi citra MRI dengan metode *clustering* agar memori yang digunakan untuk penyimpanan citra lebih sedikit. Metode *clustering* yang digunakan adalah K-Means *clustering* yang akan memisahkan objek utama dari latar belakang. Sebelum dilakukannya *clustering*, citra akan masuk tahap *pre-post* dengan membuat blok ukuran *vector* L-dimensi dari citra untuk digunakan pada teknik *clustering*. Setelah itu masuk pada tahap *clustering*, yang akan diteruskan ke tahap *main compression*. Pada tahap *compression*, piksel-piksel yang informasinya tidak dibutuhkan akan dipotong. Hasil yang didapatkan setelah melakukan tahapan di atas, citra MRI ekstensi .jpg membutuhkan waktu paling singkat untuk proses kompresi sedangkan citra MRI ekstensi .bmp membutuhkan waktu kompresi paling lama.

Dalam *machine learning*, deteksi parasit malaria dapat dilakukan dengan cara non konvensional seperti yang dilakukan oleh (Rahmanti dkk., 2016). Penelitian ini memiliki tujuan untuk membandingkan tingkat akurasi antara beberapa metode

yang digunakan untuk mengklasifikasikan *plasmodium falciparum* sehingga didapatkan nilai akurasi tertinggi. Tahapan yang dilakukan pada penelitian ini adalah *pre-processing*, ekstraksi fitur, dan klasifikasi. Preprocessing ini bertujuan untuk mendapatkan ROI (Region of Interest) dengan *cropping* secara manual dan mengubah ukuran gambar. Metode ekstraksi fitur menggunakan *Gray Level Co-occurrence Matrix* (GLCM) untuk mendapatkan nilai-nilai fitur tekstur seperti kontras, korelasi, energi, dan homogeneitas dengan orientasi sudut 0°, 45°, 90°, dan 135°. Terakhir tahap klasifikasi dengan menggunakan 3 metode klasifikasi yaitu KNN (*K-Nearest Neighbor*), *backpropagation*, dan LVQ (*Learning Vector Quantization*). Hasil perbandingan algoritma klasifikasi adalah KNN (K- Nearest Neighbor) memiliki nilai akurasi tertinggi dengan persentase sebesar 84.6667%.

Pada pertengahan tahun (Adhinata dkk., 2016) melakukan penelitian terkait identifikasi parasit plasmodium SP pada apusan darah tipis. Penelitian ini bertujuan sebagai referensi pemberian obat malaria jika tidak ada tenaga medis yang menganalisa spesies parasit malaria berdasarkan citra mikroskopik apusan darah tipis. Terdapat 2 jenis parasit yang digunakan pada penelitian ini, yaitu *plasmodium falciparum* dan *plasmodium vivax*. Metode yang digunakan adalah *thresholding* dan erosi untuk segmentasi, *bounding box* untuk ekstraksi ciri, dan *rule based* untuk klasifikasi. Nilai dari ekstraksi ciri berdasarkan luas area sel (LA), perbandingan luas area (PLA), rasio diameter (RD), dan rasio kebundaran sel (RK) yang akan digunakan pada *rule based*. Setelah melewati metode yang telah disebutkan di atas, hasil yang didapatkan adalah sistem dapat mengidentifikasi parasit jenis *plasmodium falciparum* dan *plasmodium vivax* dengan akurasi sebesar 55%. Selain itu, sistem juga memiliki kelemahan yaitu sistem tidak dapat memisahkan sel yang saling menumpuk satu sama lain sehingga luas area sel, perbandingan luas area, rasio diameter, dan rasio kebundaran menghasilkan nilai yang berbeda dengan *rule* yang telah ditentukan.

(Harto & Fatichah, 2017) mengambil permasalahan tentang segmentasi dan pemisahan sel darah putih bersentuhan pada citra leukemia myeloid akut untuk penelitian thesisnya. Fokus utamanya yaitu melakukan segmentasi pada sel (*Acute Myeloid Leukemia*) AML karena keberhasilan identifikasi dan klasifikasi sangat

bergantung pada hasil segmentasi. Masalah yang lain terdapat juga pada sel yang bersentuhan, sehingga diperlukan pemisahan antar sel-sel tersebut guna meningkatkan akurasi perhitungan sel darah putih dalam proses diagnosis. Metode yang digunakan pada penelitian ini adalah *K-Means clustering* untuk segmentasi dan *Hierarchical Clustering Analysis* (HCA) untuk pemisahan sel yang bersentuhan. K-Means dipilih karena lebih cepat dan lebih efektif mendapatkan *region* dibandingkann dengan metode *thresholding* yang membutuhkan nilai *threshold* yang berbeda pada setiap citranya karena citra AML memiliki variasi informasi di setiap piksel. Dari 40 dataset citra AML, metode yang dipilih telah menghasilkan segmentasi sel darah putih dengan dan berhasil melakukan pemisahan sel bersentuhan menjadi sel-sel tunggal. Pada perhitungan jumlah sel, *error* yang dihasilkan cukup rendah yaitu sebesar 4,04%.

Penelitian dengan menggunakan machine learning kembali dilakukan oleh (Johnson & Yadav, 2018) untuk deteksi kesalahan dan klasifikasi untuk saluran transmisi DC Tegangan Tinggi menggunakan K- Nearest Neighbors. Data yang digunakan sebagai inputan pada penelitian ini adalah tegangan AC RMS penyearah ujung, tegangan saluran DC dan arus yang diukur pada kedua kutub. Sistem yang dibuat pada penelitian ini adalah sistem yang dapat mendeteksi gangguan arus DC sehingga pemeliharaan dapat dilakukan sedini mungkin untuk memastikan kontinuitas aliran daya. Dalam penerapan sistem pendeteksi kesalahan, model KNN yang digunakan ada 6 jenis yaitu *fine KNN*, *medium KNN*, *coarse KNN*, *cosine KNN*, *cubic KNN*, dan *weighted KNN*. Sebanyak 94 kasus (93 kasus rusak, diukur pada setiap 10 km, dan 1 kasus tidak ada kesalahan) telah digunakan untuk mendeteksi dan mengklasifikasikan masing-masing dari ketiga jenis kesalahan DC (tiang 1 ke ground, tiang 2 ke ground, dan tiang 3 ke ground). Kecuali KNN coarse (akurasi rata-rata 84,03%), 6 jenis teknik KNN yang digunakan memberikan hasil 100% akurat. Dengan demikian, jenis *fine KNN*, *medium*, *cosine*, *cubic* dan *weighted* memberikan hasil yang akurat dan dapat digunakan sebagai Pengklasifikasi untuk kesalahan pada saluran transmisi HVDC.

(Setiawan & Putra, 2018) memilih topik klasifikasi citra mammogram kanker payudara untuk penelitiannya. Tujuan dilakukannya penelitian ini untuk

memperlihatkan keabnormalan atau kelainan pada payudara dan dapat mengidentifikasi secara visual tanpa harus melalui tenaga ahli medis. Metode pertama yang digunakan pada penelitian ini adalah K-Means *clustering* untuk segmentasi. Selanjutnya, *Gray Level Co-Occurrence Matrix* (GLCM) untuk ekstraksi ciri yang menggunakan perhitungan tesktur meliputi kontras, energy, homogenitas dan korelasi dengan sudut 0° , 45° , 90° dan 135° . Terakhir, terdapat *Support Vector Machine* (SVM) sebagai metode klasifikasi. Penggunaan K-Means *clustering*, GLCM, dan SVM berhasil mengidentifikasi normal dan abnormal pada citra mammogram payudara dengan akurasi sebesar 80%.

Selanjutnya (Devi dkk., 2016) menggunakan metode K Nearest Neighbor (KNN) untuk memisahkan eritrosit terisolasi dan senyawa yang ada dalam citra mikroskopis dari apusan darah tipis. Gambar tersegmentasi perlu diproses karena beberapa pusat eritrosit tersegmentasi sebagai wilayah latar belakang dan artefak sebagai latar depan. Citra yang tersegmentasi selanjutnya mengalami pemrosesan morfologis seperti pengisian lubang untuk memulihkan bagian tengah eritrosit yang tersegmentasi sebagai wilayah latar belakang. Setelah sel tersegmentasi pulih, langkah selanjutnya adalah menghapus artefak. Karena area artefak jauh lebih kecil dari area eritrosit, area tersebut mudah dihilangkan dengan menggunakan area sebagai kriteria ambang batas. Dari citra biner, daerah rata-rata dari semua wilayah tersegmentasi sebagai eritrosit diperoleh. Setiap wilayah dengan area kurang dari 0,25% dari area rata-rata ditandai sebagai wilayah latar belakang. Langkah selanjutnya adalah mengekstrak fitur. Dari citra biner yang difilter, tiga fitur seperti luas, rasio kekompakan dan rasio aspek sel yang disegmentasi dari citra mikroskopis diekstraksi dan nilai dari 3 fitur ini menjadi inputan klasifikasi KNN. 20 citra yang mengandung sekitar 500 eritrosit digunakan untuk pelatihan KNN dan 10 citra lainnya yang mengandung 300 eritrosit digunakan untuk menguji kinerja klasifikasi kNN yang digunakan untuk memisahkan eritrosit. Dari hasil pengujian pemisahan eritrosit, dengan 30 citra parasit malaria yang mengandung ± 800 eritrosit memberikan akurasi rata-rata $\sim 0,942$.

Penelitian terkait yang digunakan sebagai referensi pada judul penelitian ini adalah penelitian dari (Vallabhaneni & Rajesh, 2018) yang berfokus pada

pembuatan sistem deteksi tumor otak pada citra yang memiliki *noise*. Teknik *Edge Adaptive Total Variation Denoising* (EATVD) menjadi metode yang digunakan untuk menghilangkan *noise* pada citra. Selama proses *denoising* sinyal, atau dalam istilah lain memperlancar fluktuasi yang tidak diinginkan hadir dalam sinyal, algoritma yang digunakan secara konvensional juga menghaluskan variasi aktual yang ada dalam sinyal. Sementara itu, hasil dari EATVD akan di *input* kedalam *mean shift* untuk proses segmentasi. Selanjutnya ekstraksi ciri GLCM membantu menghitung nilai-nilai fitur yang akan dikirimkan ke *multi class SVM* yang akan mengenali bagian tumor tersebut. Hasil yang diperoleh dari penggunaan metode di atas adalah citra otak yang memiliki *noise* berhasil dideteksi dengan hasil citra tumor tersegmentasi tanpa *noise* dari citra otak sebelumnya.

Bedasarkan beberapa jurnal yang telah diulas di atas, dapat ditarik kesimpulan bahwa sistem yang telah dirancang pada penelitian terdahulu banyak menggunakan metode texture *Gray Level Co-occurrence Matrix* (GLCM), sebagian menggunakan gabungan ekstraksi fitur histogram, area *granulometry*, *gradient* dan *flat texture*, selain itu menggunakan ekstraksi fitur luas area sel yang digabungkan dengan rasio diameter maupun rasio kebundaran, atau dengan rasio kekompakan dan rasio aspek sel. Seluruh ekstraksi fitur yang digunakan nantinya akan menjadi inputan sistem klasifikasi yang akan menjadi pembeda satu jenis objek dengan jenis lainnya sehingga mendapatkan persentase akurasi yang tinggi. Pada sistem klasifikasi yang menggunakan fitur GLCM menghasilkan akurasi lebih dari atau sama dengan (\geq) 80%, sedangkan gabungan 3 - 4 jenis fitur ekstraksi yang berbeda menghasilkan akurasi minimal sebesar 55% dan akurasi maksimal sebesar 93%.

Tabel 2. 1 Referensi penelitian terkait.

Peneliti (Tahun)	Citra Penelitian	Judul Penelitian	Klasifikasi atau Segmentasi	Ekstraksi Fitur	Hasil
Tek et al (2010)	Malaria P.Falciparum P. Malariae P. Ovale P. Vivax	<i>Parasite detection and identification for automated thin blood film malaria diagnosis</i>	- KNN -FLD -BPNN	Total 83 fitur, terdiri dari: - Histogram - Local area granulometry - A shape measurements vector	- 93% -90,1% -92,2%
Malihi et al (2013)	Malaria P.Falciparum P. Malariae P. Ovale P. Vivax	<i>Malaria Parasite Detection in Giemsa – Stained Blood Cell Images</i>	- K-NN, -1-NN - NM -SVM - Fisher linear discriminator	- Gradient - Flat texture - Color histogram - Area granulometry	- 91% - 87% - 70% - 90% - 84%
Wijaya et al (2015)	MRI otak ekstensi .png, .jpg dan .bmp	<i>Penerapan Algoritma K- Means Pada Kompresi Adaptif Citra Medis MRI</i>	- K-Means clustering	Tidak ada ekstraksi	Kompresi paling singkat .jpg

Peneliti (Tahun)	Citra Penelitian	Judul Penelitian	Klasifikasi atau Segmentasi	Ekstraksi Fitur	Hasil
Rahmanti et al (2016)	Malaria P.Falciparum	<i>A Comparison of Plasmodium Falciparum Identification from Digitalization Microscopic Thick Blood Film</i>	- KNN - Back propagation - LVQ	- Texture	- 84,6667% - 82.6668 % - 79.3334%
Adhinata et al (2016)	Malaria P.Falciparum P. Vivax	<i>Identification of Parasite Pasmodium SP. on Thin Blood Smears With Rule-Based Method</i>	- Rule Based	Bounding Box, terdiri dari: - Luas area - Perbandingan luas area - Rasio diameter - Rasio kebundaran sel.	Akurasi 55%
Harto et al (2017)	Citra Leukemia Myeloid	<i>Segmentasi Dan Pemisahan Sel Darah Putih Bersentuhan Menggunakan K-Means Dan Hierarchical Clustering Analysis Pada</i>	- K-Means - Hierarchical Clustering Analysis (HCA).	Tidak ada ekstraksi	Error 4.04%

Peneliti (Tahun)	Citra Penelitian	Judul Penelitian	Klasifikasi atau Segmentasi	Ekstraksi Fitur	Hasil
		<i>Citra Leukemia Myeloid Akut</i>			
Johnson et al (2018)	HVDC transmission system	<i>Fault Detection And Classification Technique For HVDC Transmission Lines Using KNN</i>	6 jenis KNN -Fine -Medium -Coarse -Cosine -Cubic -Weighted	Tidak ada ekstraksi	Akurasi semua tipe 100% kecuali coarse KNN
Setiawan et al (2018)	Kanker Payudara	<i>Klasifikasi Citra Mammogram Menggunakan Metode K-Means, GLCM, dan Support Vector Machine (SVM)</i>	- K-Means - SVM	- Texture	Akurasi 80%
Devi et al (2018)	Eritrosit Parasit Malaria	<i>kNN Classification based Erythrocyte Separation in</i>	-KNN	- luas - rasio kekompakan - rasio aspek sel	Akurasi 0.942 Error

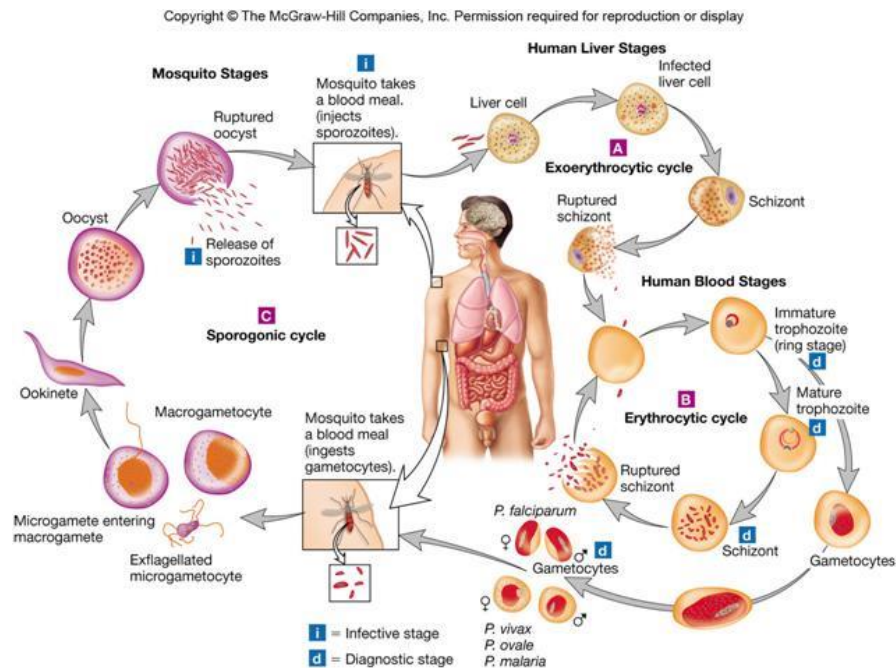
Peneliti (Tahun)	Citra Penelitian	Judul Penelitian	Klasifikasi atau Segmentasi	Ekstraksi Fitur	Hasil
		<i>Microscopic Images of Thin Blood Smear</i>			0.057
Vallabhane ni et al (2018)	Tumor Otak	<i>Brain tumour detection using mean shift clustering and GLCM features with edge adaptive total variation denoising technique</i>	- Mean shift clustering - SVM	- Texture	Akurasi 100%

2.2 Dasar Teori

2.2.1 Parasit Malaria

Malaria merupakan penyakit endemik yang menyerang sel darah merah melalui gigitan nyamuk *anopheles* betina yang telah terjangkit parasit jenis *plasmodium*. Siklus hidup *plasmodium* dimulai pada saat nyamuk menghisap darah *carrier* (penderita malaria) yang mengandung sel gamet jantan (*mikrogametosit*) dan sel gamet betina (*makrogametosit*) secara bersamaan. Kedua sel kelamin tersebut selanjutnya akan membentuk zigot lalu *ookinet* dan berubah menjadi *ookista*, terakhir *ookista* pecah berubah menjadi *sporozoit* yang nantinya akan berada di kelenjar ludah nyamuk *anopheles*.

Ketika nyamuk menggigit manusia dan melakukan transfer ludah ke dalam tubuh, maka nyamuk akan memindahkan parasitnya yang berbentuk *sprozoit* pada manusia. Kurang lebih dalam waktu 30 menit *sprozoit* akan sampai ke sel hati dan berubah menjadi *schizon* hati muda, tua kemudian matang. Selanjutnya *schizon* hati matang akan melepaskan *merozoit* untuk masuk ke dalam sirkulasi darah dan menyebar ke seluruh tubuh. Siklus hidup parasit jenis *plasmodium* dapat dilihat pada Gambar 2.1.



Gambar 2. 1 Siklus hidup parasit plasmodium pada tubuh manusia
 Sumber: mheducation.com/The McGraw and Hill Companies

Spesies *plasmodium* yang termasuk parasit malaria adalah *plasmodium falciparum*, *plasmodium ovale*, *plasmodium knowlesi*, *plasmodium malariae*, dan *plasmodium vivax*. Penyebaran parasit malaria yang dibawa oleh *anopheles* dipengaruhi oleh intensitas curah hujan, serta sumber parasit *plasmodium* atau penderita positif malaria. Selain itu, lingkungan rumah yang berdekatan dengan perairan kotor seperti rawa dan sungai yang tercemar akan memberikan situasi berbahaya karena *anopheles* semakin mudah menularkan parasit nya.

Pemeriksaan malaria umumnya dilakukan ketika penderita memiliki gejala klinis seperti mual, muntah, pusing, dan demam berkala. Penderita akan diambil sampel darah tepi yang biasanya diambil dari ujung jari tangan atau jempol kaki. Selanjutnya membuat sediaan darah tipis untuk melihat morfologi parasit untuk menentukan spesiesnya dan melihat ketebalan untuk menentukan kepadatannya (Hakim, 2011). Tetapi pemeriksaan mikroskopis tidak dapat digunakan sebagai pondasi utama dalam diagnosis terutama pada pasien yang telah diberikan obat anti malaria atau *profilaksis*, karena obat ini secara parsial dapat menyebabkan berkurangnya jumlah parasit sehingga berada di bawah ambang pemeriksaan mikroskopis. Inilah yang menyebabkan ditemukannya sedikit parasit yang menggambarkan parasitemia yang rendah padahal pasien sedang menderita malaria fase berat.

Pada penelitian kali ini hanya berfokus pada 3 spesies parasit *plasmodium*, yaitu; *plasmodium falciparum*, *plasmodium malarie*, dan *plasmodium vivax*. Karakteristik yang dapat menunjukkan spesies sehingga dapat diidentifikasi dapat dilihat pada Gambar 2.2 dengan karakteristik adalah sebagai berikut:

a. *Plasmodium Falciparum*

Merupakan spesies yang paling berbahaya karena memiliki sifat resistansi terhadap beberapa jenis antibiotik, potensial menimbulkan *hiperparasitemia*, dan komplikasi. Morfologi *trophozite P.Falciparum* berupa ring/cincin kecil dengan diameter kurang dari 1/3 diameter eritrosit normal. Bentuk cincin ini halus, mempunyai sitosplasma berwarna biru, kromatin inti merah dan *P.Falciparum* adalah satu-satunya spesies yang mempunyai 2 kromatin inti



















(*double chromatin*). Bentuk *skizon* jarang ada dalam sirkulasi darah tepi, dan jika ditemukan *skizon* dalam darah tepi hal ini menunjukkan penderita telah mengalami malaria berat. Bentuk gametosit sangat khas yaitu elips (*crescent*) berpigmen warna hitam dengan sitoplasma kuning. Pada sediaan darah tipis mudah diidentifikasi adanya presipitasi Hb (hemoglobin) berupa bintik merah kasar dalam sitoplasma eritrosit terinfeksi yang disebut bintik *maurer*.

b. Plasmodium Malariae

Termasuk parasit yang mempunyai trophozoit yang serupa dengan *P. Vivax* tetapi lebih kecil dan sitoplasmanya lebih kompak/lebih biru. Trophozoit matur mempunyai granula coklat tua sampai hitam dan kadang mengumpul sampai membentuk pita. Skizon *P. Malariae* mempunyai 8-10 merozoit yang tersusun seperti kelopak bunga/rosete. Bentuk gametosit sangat mirip dengan *P. Vivax* tetapi lebih kecil.

c. Plasmodium Vivax

Parasit ini biasanya menginfeksi eritrosit muda (*retikulosit*) yang diameternya lebih besar dibanding eritrosit normal dan sering terdapat bintik halus *schuffner* sebagai hasil denaturasi Hb (hemoglobin). Bentuk *trophozoit* mula-mula mirip *P. Falciparum* dengan diameter kira-kira 1/3 diameter eritrosit terinfeksi. Seiringnya dengan maturnya, *trophozoit vivax* berubah menjadi bentuk *amoeboid*. Skizon terdiri dari 12-24 merozoit dengan bentuk oval dan pigmen kuning tengguli. Gametosit berbentuk oval hampir memenuhi seluruh eritrosit kromatin inti eksentris, pigmen kuning tengguli.

	Trophozoites		Schizonts		Gametocytes	
	Old	Young	Mature	Immature	Female	Male
<i>P. malariae</i>						
<i>P. vivax</i>						
<i>P. falciparum</i>						

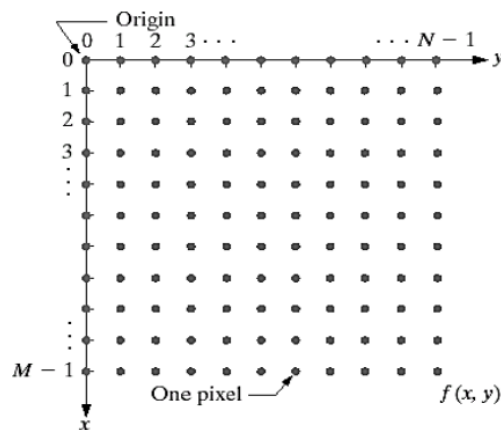
Gambar 2. 2 Perbandingan gambaran Sediaan darah tepi pada parasit plasmodium falciparum, plasmodium vivax dan plasmodium malariae

Sumber: yayanakhyar.wordpress.com/ Yayan Akhyar Israr

2.2.2 Citra

Citra merupakan salah satu aspek penting yang digunakan dalam kehidupan sehari-hari. Pada beberapa kesempatan, citra akan digunakan sebagai sebuah ilustrasi (penjelasan) maupun perbandingan objek satu dengan yang lain, interpretasi, petunjuk, pemantapan, penggambaran, hiburan, serta masih banyak lagi. Namun, ketika penggunaan citra dimanfaatkan pada level yang lebih tinggi lagi, citra akan digunakan sebagai salah satu representasi yang dapat diolah dan dimengerti oleh komputer. Konsep tersebut dinamakan pengolahan citra.

Citra yang digunakan dalam pengolahan citra adalah citra digital 2 dimensi dimana x dan y adalah koordinat bidang sedangkan amplitudo dari f pada pasangan koordinat apapun adalah intensitas atau tingkat keabu-abuan dari citra pada titik tersebut (Adhinata dkk., 2016). Matriks dari citra digital biasanya di representasikan dengan nilai M sebagai baris dan nilai N sebagai kolom, matriks $M \times N$ dapat diilustrasikan seperti gambar di bawah ini.



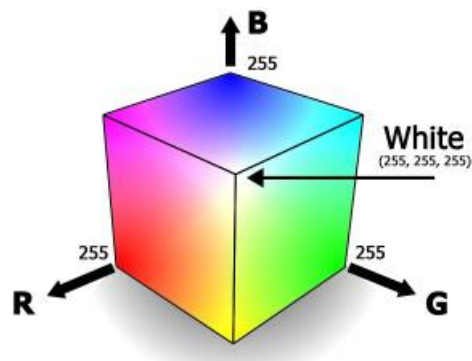
Gambar 2. 2 Representasi citra piksel dengan matriks $M \times N$
Sumber: (Adhinata et al., 2016)

Citra digital terdiri dari berbagai jenis dengan berpedoman pada rentang warna citra seperti yang dijeaskan di bawah ini.

a. Citra RGB

Citra warna atau biasa yang disebut dengan RGB merupakan citra yang di representasikan dari warna dasar Red, Green, dan Blue. Dalam MATLAB,

komponen merah, hijau, biru dari gambar RGB berada dalam tiga matriks intensitas terpisah, masing-masing memiliki dimensi baris dan kolom yang sama dengan gambar RGB asli (Thompson & Loren, 2002). Dengan rentang nilai minimal adalah 0 dan rentang nilai maksimal adalah 255. Intensitas piksel dari setiap matriks akan bergabung untuk membentuk campuran warna $256 \times 256 \times 256$, apabila dijumlahkan sebanyak 16.777.216 jenis warna.



Gambar 2. 3 Model ruang citra RGB
Sumber: pemrogramanmatlab.com

b. Citra *grayscale*

Citra *greyscale* merupakan citra yang memiliki rentang nilai 0-255, nilai 0 didefinisikan sebagai warna hitam dan nilai 255 didefinisikan sebagai warna putih. Derajat keabuan citra *greyscale* akan mengrepresentasikan citra dengan dengan intensitas warna yang berbeda namun masih dalam rentang nilai 0-255. Konversi *greyscale* dari citra RGB dapat dilakukan dengan cara membentuk jumlah bobot dari komponen R, G, dan B, dengan algoritma:



Gambar 2. 4 Citra greyscale

c. Citra Biner

Sebuah citra biner adalah citra yang memiliki jenis intensitas yang spesial. Citra biner hanya terdiri dari 2 level rentang warna, nilai 0 mewakili hitam sedangkan nilai 1 mewakili putih. Pada pengolahan citra, citra biner digunakan sebagai pemisah antara objek dengan latar belakangnya. Dengan cara menyatakan objek yang diinginkan dengan nilai 1, sedangkan objek lain yang harus dihilangkan (latar belakang) dinyatakan dengan 0 seperti teknik segmentasi.



Gambar 2. 5 Citra biner

2.2.3 Image Processing

Pengolahan citra atau *image processing* merupakan teknik yang mengolah citra sedemikian rupa sehingga citra dapat diinterpretasi informasi setiap piksel pada komputer. Pada dasarnya, citra yang diakuisi kedalam komputer masih memiliki tingkat *noise* yang tinggi, sehingga komputer tidak dapat memahami informasi apa yang terkandung dalam citra tersebut. Untuk

meminimalisir bahkan menghilangkan *noise*, citra tersebut harus memasuki beberapa tahapan pengolahan citra berikut ini, perbaikan kualitas citra (*image enhancement*), pemugaran citra (*image restoration*), pemampatan citra (*image compression*), segmentasi citra (*image segmentation*), analisis citra (*image analysis*), dan ekonstruksi citra (*image reconstruction*).

Pengolahan citra ke visi komputer dapat dipecah menjadi proses tingkat rendah, menengah dan tinggi. Tingkat rendah masih menggunakan teknik-teknik sederhana seperti *pre-processing* yang digunakan sebagai pengurangan derau, mengatur kontras, dan mengatur ketajaman citra. Sedangkan tingkat medium, telah melibatkan teknik yang cukup rumit seperti segmentasi dan klasifikasi citra. Untuk tingkat tinggi, sistem sudah menjadikan objek-objek yang sudah dikenali melakukan fungsi kognitif yang diasosiasikan dengan vision.

Salah satu penggunaan yang dibutuhkan pada pengolahan citra adalah proses segmentasi, dikarenakan pada tahap ini citra akan di olah sedemikian rupa sehingga berfokus pada informasi yang dibutuhkan.

2.2.4 MATLAB

Matrix Laboratory atau yang biasa dikenal dengan MATLAB dengan logo seperti pada Gambar 2.5 merupakan salah satu *software* dengan bahasa pemrograman tingkat tinggi yang dikembangkan oleh The MathWorks.Inc. Pemrograman yang terdapat pada MATLAB tidak sama dengan bahasa pemrograman yang ada sebelumnya seperti C, C++, java, phyton, HTML, maupun PHP. Penggunaan software MATLAB lebih dikhususkan pada pemrograman berbasis komputasi numerik yang akan melakukan eksekusi perhitungan matematika, simulasi dan pemodelan, visualisasi dan analisis data, pembuatan grafik perhitungan, dan pengembangan aplikasi yang memanfaatkan GUI (*Graphical User Interface*). Pengoperasian MATLAB mendukung perangkat dengan sistem operasi berbasis Windows, Linux dan Mac OS.



Gambar 2. 6 Logo MATLAB
Sumber: cleanpng.com

Software MATLAB didukung dengan berbagai fitur yang biasanya dikenal dengan *toolbox* yang di dalamnya terdapat *tutorial toolbox*, pengertian *toolbox*, contoh program, fungsi program, hingga hasil dari program tersebut. *Toolbox* MATLAB sendiri terdiri puluhan fitur seperti *toolbox image processing*, *toolbox control system*, *toolbox optimization*, dan masih banyak lagi. Dalam pengolahan citra, *image processing toolbox* pada MATLAB terdiri dari:

a. Deskripsi dari produk *image processing toolbox*

Menyediakan referensi algoritma standar dan alur kerja secara lengkap untuk pemrosesan gambar, analisis, visualisasi, dan pengembangan algoritma.

b. Tutorial

Disini akan dijeskan *step by step* penggunaan *toolbox* dari tingkat dasar hingga tinggi contohnya, cara membaca *image* pada MATLAB hingga menemukan vegetasi dalam gambar multispektral. Keuntungan lain dari MATLAB adalah *software* ini menyediakan *help window* yang dapat diakses secara *offline* sehingga pengguna MATLAB dapat mengakses informasi yang tersedia pada menu *help* kapan saja dan dimana saja

2.2.5 *K-Means Clustering*

K-Means merupakan salah satu metode pembelajaran mesin (*machine learning*) yang bertujuan untuk melakukan proses segmentasi dengan cara

pengelompokan iteratif. Seperti namanya, K-Means *clustering* akan mempartisi data kedalam ruang-ruang *cluster* berdasarkan spesifikasi maupun sifatnya, sehingga data yang memiliki persamaan spesifikasi dan sifat akan diletakkan pada satu *cluster* atau kelompok yang sama, sedangkan untuk spesifikasi dan sifat yang berbeda akan ditempatkan kedalam *cluster* atau kelompok yang lainnya sejumlah K-*cluster* yang telah ditetapkan sejak awal. Penggunaan metode ini bertujuan untuk mengoptimasikan hasil dari proses *clustering*, sehingga dalam satu *cluster* akan memiliki variasi yang sedikit namun memperbanyak variasi antar *cluster*.

Ketika sebuah data direpresentasikan dengan variabel $X = \{x_i | i = 1, \dots, N\}$ maka algoritma K-Means akan mempartisi X dalam k buah *cluster*, sehingga titik data x_i hanya akan berada pada satu k partisi. Setiap *cluster* tentu memiliki labelnya sendiri yang ditandai dengan ID *cluster*. Data yang memiliki ID *cluster* yang sama akan ditempatkan pada satu *cluster* yang sama, tetapi untuk data dengan ID *cluster* yang berbeda dari sebelumnya, akan ditempatkan dalam *cluster* yang berbeda. ID *cluster* pada algoritma K-Means biasanya dikelompokkan berdasarkan konsep kedekatan atau kemiripan, tetapi untuk kuantitas yang digunakan untuk mengukur cluster satu dengan yang lain adalah menggunakan ketidakmiripan (*dissmiliarity*). Artinya, data-data dengan ketidakmiripan (jarak) yang kecil/dekat maka lebih besar kemungkinannya untuk bergabung dalam satu *cluster*. Ukuran ketidakmiripan dapat dihitung dengan menggunakan jarak Euclidean (Prasetyo, 2014).

Indikator utama yang harus berada pada algoritma K-Means adalah nilai k. Nilai k yang digunakan biasanya didapatkan dari perhitungan berapa banyak *cluster* yang berada pada data X dengan melihat ketidakmiripan *cluster* dengan *cluster* lainnya, banyaknya *cluster* yang dibutuhkan, atau jenis *cluster* dicari dengan melakukan percobaan *trial* dan *error* dengan beberapa nilai k. Ketidakmiripan dapat dilihat berdasarkan perhitungan setiap *centroid* dari setiap *cluster*, ketidakmiripan yang paling kecil sebagai yang akan diikuti sebagai relokasi data pada *cluster* di sebuah iterasi. Relokasi sebuah data dalam

cluster yang diikuti dapat dinyatakan dengan nilai keanggotaan α yang bernilai 0 atau 1. Nilai 0 jika tidak menjadi anggota sebuah *cluster* dan 1 jika menjadi anggota sebuah *cluster*. Karena K-Means mengelompokkan secara tegas data hanya pada satu *cluster*, maka dari nilai α sebuah data pada semua *cluster*, hanya satu yang bernilai 1, sedangkan lainnya 0 seperti dinyatakan oleh persamaan berikut:

$$\alpha_{ij} = \begin{cases} 1 & \text{arg min } \{d(x_i, c_j)\} \\ 0 & \dots\dots\dots \end{cases} \quad (2.1)$$

$d(x_i, c_j)$ menyatakan ketidakmiripan (jarak) dari data ke-*i* ke *cluster* c_j

Algoritma K-Means *clustering* dapat dilakukan dengan langkah-langkah di bawah ini:

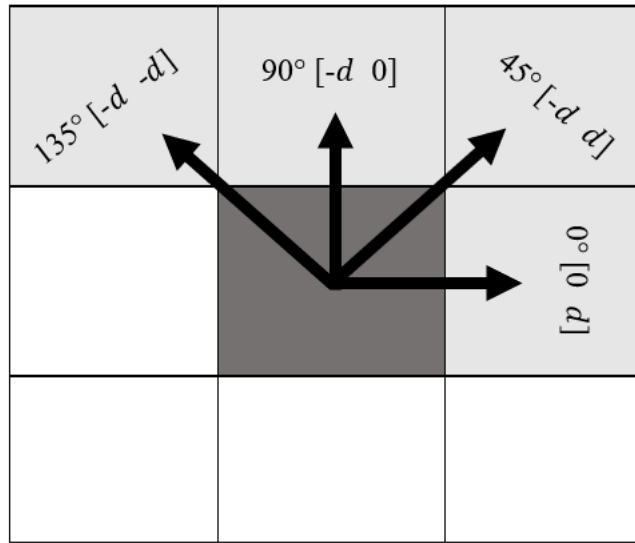
1. Mengkonversi citra dari RGB ke ruang warna L*a*b* untuk mengidentifikasi berbagai warna pada citra tersebut. Ruang warna L*a*b* diturunkan dari nilai tristimulus CIE XYZ. Ruang L*a*b* terdiri dari lapisan luminositas “L *”, lapisan kromatisitas “a *” yang menunjukkan di mana warna berada di sepanjang sumbu merah-hijau, dan lapisan kromatisitas “b *” yang menunjukkan di mana warna berada di sepanjang sumbu biru-kuning.
2. Mengkonversi citra L*a*b* ke citra *single* dengan cara mengubah ukuran piksel citra sebelum masuk ketahapan segmentasi, hal ini dilakukan untuk menyamaratakan nilai piksel pada seluruh citra.
3. Menentukan nilai k sebagai jumlah *cluster* yang diinginkan pada proses *clustering* dengan menggunakan informasi warna dalam ruang warna “a*b*”.
4. Menggunakan algoritma ‘*imsegkmeans*’ yang nantinya pada algoritma ini akan mengalokasikan semua data ke *centroid* terdekat dengan *cluster* k yang telah ditetapkan

Berdasarkan beberapa jurnal yang telah diulas, metode K-Means banyak digunakan sebagai sistem segmentasi citra terutama citra medis seperti pada

penelitian (Wijaya & Kusumadewi, 2015) untuk mengetahui perbandingan ekstensi .png, .jpg, dan .bmp pada citra medis MRI yang menghasilkan kompresi segmentasi paling cepat. Selanjutnya (Harto & Fatichah, 2017) mengkombinasikan metode K-Means dan *Hierarchical Clustering Analysis* (HCA) untuk pemisahan sel darah putih bersentuhan pada citra leukemia myeloid akut. Kembali metode K-Means digunakan untuk segmentasi citra medis mammogram pada sel kanker payudara oleh (Setiawan & Putra, 2018). Segmentasi yang dilakukan pada penelitian ini bertujuan untuk memisahkan sel kanker dari unsur elemen lainnya. (Nanoti dkk., 2016) menggunakan metode K-Means untuk memisahkan sel parasit dengan sel normal dan latar belakang terhadap citra parasit malaria jenis *plasmodium falciparum*, *plasmodium malariae*, *plasmodium ovale*, dan *plasmodium vivax*.

2.2.6 Gray Level Co-Occurance Matrix (GLCM)

Algoritma GLCM merupakan suatu metode yang digunakan untuk mengekstraksi fitur sebuah citra berbasis *statistical* dan berdasarkan nilai-nilai piksel matriks yang akan membentuk suatu sudut pola. Sedangkan kookurensi dapat diartikan sebagai kejadian bersama, berarti banyaknya kejadian pada satu level piksel yang bertetangga dengan nilai piksel yang lainnya berdasarkan jarak (d) dan orientasi suatu sudut (Θ). Jarak direpresentasikan sebagai piksel sedangkan orientasi direpresentasikan dalam derajat. Orientasi terbentuk dari empat arah sudut dengan interval 45° , yaitu 0° , 45° , 90° dan 135° , dan jarak antar piksel diinterpretasikan dari piksel ke- d dan kelipatannya secara berulang. Nilai antar piksel akan saling berhubungan dengan piksel tetangganya berdasarkan arah sudut (0° , 45° , 90° dan 135°) yang dihitung berdasarkan jarak antar piksel (d). Pemilihan jarak antar piksel dapat disesuaikan dengan citra yang akan diekstrak.



Gambar 2. 7 Sudut GLCM pada jarak antar piksel d

Di dalam GLCM terdapat nilai fitur *contrast*, *corellation*, *energy*, dan *homogeneity* untuk menghitung tekstur yang akan digunakan. Untuk penjelasan lebih detail dari ke empat fitur ini dapat dilihat di bawah ini:

a. *Contrast*

Menghitung kontras intensitas antara piksel dan tetangganya pada seluruh citra. Untuk sebuah citra dengan intensitas konstan, nilai kontasnya adalah 0. Kontras juga dikenal sebagai varian dan inersia.

b. *Correlation*

Mengukur ketidakmiripan piksel terhadap tetangganya pada keseluruhan citra. Nilai besar apabila ketidakmiripan tersusun acak, dan bernilai kecil apabila ketidakmiripan tersusun seragam. Korelasi adalah NaN untuk citra konstan.

c. *Energy*

Jumlah elemen kuadrat dalam GLCM yang telah di normalisasi. Untuk sebuah citra dengan energi konstan, nilai energinya adalah 1. Properti Energi juga dikenal sebagai keseragaman, keseragaman energi, dan momen kedua sudut.

d. *Homogeneity*

Menunjukkan kehomogenan citra yang derajat keabuannya sejenis. Citra yang homogen akan memiliki homogeneity yang besar. Untuk sebuah diagonal GLCM, nilai diagonalnya adalah 0.

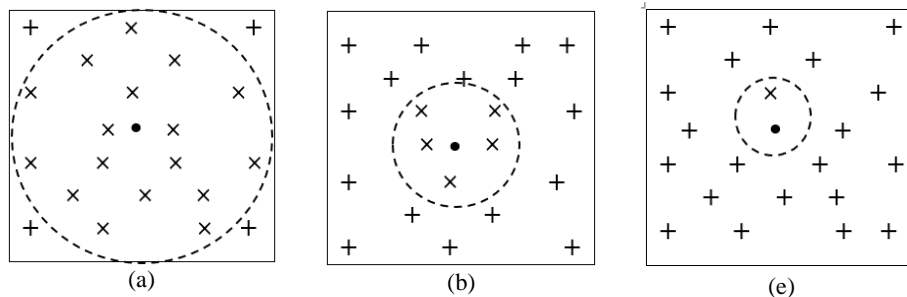
Metode GLCM banyak digunakan pada penelitian terdahulu untuk menghitung nilai fitur pada sebuah data yang berupa citra. Seperti pada penelitian (Rahmanti dkk., 2016) yang menggunakan GLCM untuk mencari nilai tekstur pada citra parasit jenis *plasmodium falciparum*. Lalu ada (Setiawan & Putra, 2018) menerapkan metode GLCM pada citra mammogram kanker payudara setelah sel kanker dipisahkan dari elemen lainnya. (Vallabhaneni & Rajesh, 2018) mengaplikasikan metode GLCM untuk ekstraksi fitur pada citra tumor otak. Kemudian penelitian ekstraksi fitur dengan menggunakan GLCM juga digunakan oleh (Öztürk & Akdemir, 2018) yang mengaplikasikannya pada citra jaringan kanker.

2.2.7 K-Nearest Neighbor (K-NN)

Metode KNN menerapkan konsep klasifikasi berdasarkan kemiripan suatu data dengan data yang lain. Kemiripan data uji dengan suatu data tersebut didapatkan dari jarak terdekat data uji dengan data yang lainnya. Apabila di sekeliling data uji merupakan sekelompok data A maka data uji tersebut akan diidentifikasi sebagai data A. Pedoman *neighbor* (tetangga) data uji diambil dari data latih yang memiliki ciri-ciri yang sama. Hal ini disebabkan karena semakin dekat lokasi data uji terhadap data latih, maka dapat dikatakan bahwa data latih tersebut yang lebih dipandang mirip oleh data uji.

Algoritma KNN sangat bergantung pada nilai K, nilai K disini merupakan jumlah tetangga terdekat yang akan digunakan dalam penentuan prediksi label kelas pada data uji. Apabila nilai K terlalu kecil maka dapat mengakibatkan hasil uji sangat sensitif terhadap noise di sekitarnya. Namun, jika nilai K terlalu besar maka tetangga terdekat yang masuk dalam satu kategori jadi terlalu banyak, yang sebenarnya tidak memiliki ciri yang sama. Untuk memprediksi nilai K yang paling tepat, dapat dilakukan dengan

menggunakan teknik validasi silang. Untuk nilai *default* pada K adalah 0 (Prasetyo, 2014). Adapun ilustrasi penentuan nilai K adalah sebagai berikut:



Gambar 2. 8 Nilai K terlalu besar (b) Nilai K ideal (c) Nilai K terlalu kecil

K-Nearest Neighbor merupakan salah satu metode pada *machine learning* yang digunakan untuk klasifikasi. Metode ini banyak digunakan pada penelitian sebelumnya seperti berikut: (Tek dkk., 2010) menggunakan KNN untuk mendeteksi dan mengidentifikasi pada 4 jenis citra parasit malaria yaitu *plasmodium falciparum*, *plasmodium malariae*, *plasmodium ovale*, dan *plasmodium vivax*. Klasifikasi menggunakan KNN terhadap 4 jenis parasit malaria kembali dilakukan oleh (Malihi dkk., 2013). Kemudian (Rahmanti dkk., 2016) mengaplikasikan metode KNN untuk mengidentifikasi parasit malaria jenis *P. Falciparum*. Selanjutnya KNN digunakan sebagai pendeteksi dan klasifikasi kesalahan pada saluran transmisi HVDC oleh (Johnson & Yadav, 2018). KNN yang digunakan pada penelitian ini menggunakan 6 jenis KNN yang tersedia di MATLAB yaitu *fine KNN*, *medium KNN*, *coarse KNN*, *cosine KNN*, *cubic KNN*, dan *weighted KNN* untuk membandingkan dan mengetahui nilai akurasi tertinggi. Ada pula (Devi dkk., 2016) yang menggunakan metode KNN untuk mengidentifikasi sel parasit malaria sehingga sel parasit dapat dipisahkan dari latar belakang dan elemen lain.