

Klasifikasi Jenis dan Fase Parasit Malaria Plasmodium Falciparum dan Plasmodium Vivax Dalam Sel Darah Merah Menggunakan Support Vector Machine

Endi Permata¹, I Ketut Eddy Purnama², Mauridhi Hery Purnomo³

¹Jurusan Teknik Elektro, Universitas Sultan Ageng Tirtayasa, Jalan Raya Pakupatan KM.3 Serang

^{2,3}Institut Teknologi Sepuluh Nopember, Kampus ITS Sukolilo Surabaya

¹endi.permata10@mhs.ee.its.ac.id

²ketut@ee.its.ac.id

²hery@ee.its.ac.id

Abstrak — Malaria adalah penyakit yang disebabkan oleh protozoa yang disebut plasmodium, yang dalam salah satu tahap perkembang biakannya akan memasuki dan menghancurkan sel-sel darah merah. Ada empat spesies yang menyebabkan malaria pada manusia yaitu: plasmodium falciparum, plasmodium vivax, plasmodium ovale dan plasmodium malariae. Namun kasus malaria yang banyak ditemukan di Indonesia hanya spesies plasmodium falciparum dan plasmodium vivax. Masing-masing parasit memerlukan tindakan medis yang berbeda. Pemeriksaan dilaboratorium terkadang terpengaruh faktor subyektive petugas. Dari hasil percobaan tahap klasifikasi menggunakan metode support vector machine one against one didapatkan hasil akurasi falciparum thropozoit 95,55%, falciparum schizont 93,48% , falciparum gametocyte 91,11 % , vivax thropozoit 88,88%, vivax schizont 92,22% dan vivax gametocyte 85%. Pada tahap klasifikasi menggunakan metode support vector machine one against all didapatkan hasil akurasi falciparum thropozoit 95,55%, falciparum schizont 92,78% , falciparum gametocyte 91,67 % , vivax thropozoit 91,67%, vivax schizont 87,78% dan vivax gametocyte 84,44%. pada perhitungan akurasi multi kelas berdasarkan tiga kali pengujian didapatkan akurasi pengujian pertama svm one against one 76,67% dan svm one against all 65 % , pengujian kedua yaitu svm one against one 65% dan svm one against all 66,67% dan pengujian ketiga yaitu svm one against one 80% dan svm one against all 81,67%.

Kata Kunci : parasit malaria, support vector machine, ekstraksi fitur, klasifikasi jenis dan fase parasit malaria plasmodium falciparum dan plasmodium vivax

Abstract — Malaria is a disease caused by a protozoan called plasmodium, which in one stage of development will enter and destroy red blood cells. There are four species that cause malaria in humans: plasmodium falciparum, plasmodium vivax, plasmodium ovale and plasmodium malariae. However, malaria cases found in Indonesia are only of species plasmodium falciparum and plasmodium vivax. Each parasite requires a different medical treatment. Laboratory examinations are sometimes influenced by subyektive factors of officers. Classifying of one against one results in accuracy of 95.55% for falciparum thropozoit, of 93.48% for falciparum schizont, of 91.11% for falciparum gametocyte, of 88.88% for vivax thropozoit, of 92.22% for vivax schizont and 85% for vivax gametocyte. Classifying of one against all results in accuracy of 95.55% for falciparum thropozoit, of 92.78% for falciparum schizont, of 91.67% for falciparum gametocyte, of 91.67% for vivax thropozoit, of 87.78% for vivax schizont and 84.44% for vivax gametocyte. Based on calculation of accuracy multi class three times testing obtained first accuracy testing SVM one against one 76.67% and SVM one of against all 65%, Second testing SVM one against one 65% and SVM one of against all 66.67% and third testing SVM one against one 80% and SVM one of against all 81.67%.

Keywords : malaria parasite, support vector machines, feature extraction, classification type and phase of the malaria parasite plasmodium falciparum and plasmodium vivax.

I. PENDAHULUAN

Malaria merupakan penyakit yang disebabkan oleh parasit protozoa dari genus plasmodium yang menginfeksi sel darah merah penderita. Parasit tersebut masuk ke dalam tubuh manusia melalui gigitan nyamuk anopheles betina. Spesies plasmodium yang menginfeksi manusia yaitu Plasmodium falcifarum, Plasmodium vivax, Plasmodium ovale, dan Plasmodium malariae. Plasmodium vivax menyebabkan malaria tertiana, Plasmodium malaria merupakan penyebab malaria kuartana. Plasmodium ovale menyebabkan malaria ovale, sedangkan Plasmodium falciparum menyebabkan malaria tropika. Spesies terakhir ini

paling berbahaya karena malaria yang ditimbulkan dapat menjadi berat. Hal ini disebabkan dalam waktu singkat dapat menyerang eritrosit dalam jumlah besar, sehingga menimbulkan berbagai komplikasi di dalam organ-organ tubuh. Namun kasus malaria yang banyak ditemukan di Indonesia hanya spesies plasmodium falciparum dan plasmodium vivax.

Penelitian pada bidang laboratorium klinis mengenai analisis morfologi sel darah merah yang terinfeksi malaria masih tergolong sedikit di indonesia. Analisis yang dilakukan oleh paramedis dan pihak laboratorium pada saat ini masih konvensional. Di laboratorium sel darah merah yang terinfeksi malaria diteliti secara

seksama dengan mikroskop. Diagnosis malaria secara pasti dapat dilakukan apabila ditemukan parasit malaria dalam darah penderita. Hasil yang diperoleh dalam setiap layang pandang kemudian di analisis, sehingga diperoleh informasi parasit malaria dalam darah penderita terinfeksi *plasmodium falciparum* atau *plasmodium vivax*. Metode konvensional ini akan menimbulkan perbedaan diagnosis jika dilakukan oleh ahli yang berbeda.

Diagnosis malaria secara pasti dapat dilakukan apabila ditemukan parasit malaria dalam darah penderita. *Plasmodium falciparum* dan *Plasmodium vivax* dalam darah memiliki beberapa fase penting diantaranya adalah *thropozit*, *schizont* dan *gametozit*. Beberapa grup riset telah melakukan pengkajian terhadap citra preparat darah. Penelitian yang dilakukan adalah melakukan segmentasi menggunakan metode *k-mean clustering* dan kemudian melakukan ekstraksi ciri terhadap citra data yang akan diuji. Ekstraksi ciri yang digunakan sebagai masukan pada sistem yang akan dibangun pada penelitian ini ada dua kelompok ciri, yaitu ciri warna dan ciri histogram. Langkah terakhir adalah melakukan uji identifikasi dan mengklasifikasi parasit *plasmodium falciparum* kedalam empat kelas dengan menggunakan metode jaringan saraf tiruan (JST) *Learning Vektor Quantization* [5].

Beberapa penelitian yang dilakukan melakukan segmentasi dengan pemisahan sel darah merah diantaranya menggunakan representasi *Inclusion-Tree* dan melakukan dua klasifikasi untuk mengidentifikasi sel darah merah yang terinfeksi oleh *plasmodium* menggunakan *binary classifier* dan menentukan fase *plasmodium* menggunakan *multiclass classifier* [2]. Proses segmentasi dan dilanjutkan dengan klasifikasi menggunakan pendekatan morfologi dan kesamaan warna histogram [3]. Penelitian lainnya menggunakan *region based Active Contour* dan dilanjutkan dengan klasifikasi menggunakan *Support Vektor Machine* [1].

II. Tinjauan Pustaka

Penelitian ini bertujuan membangun sistem klasifikasi jenis dan fase parasit malaria *plasmodium falciparum* dan *plasmodium vivax*. fitur yang digunakan adalah *standard deviation*, *mean*, *skewness*, *entropy*, *kurtosis* dan *grayscale* dari histogram warna, histogram *grayscale* dan histogram tingkat saturasi. selanjutnya dilakukan proses klasifikasi. metoda yang digunakan untuk pengenalan pola dalam penelitian ini adalah *support vector machine (svm)*. Data citra parasit malaria yang digunakan dalam penelitian ini diperoleh *public health image library (phil)* dari *phil.cdc.gov*.

Terdapat dua tahapan utama yang dilaksanakan pada penelitian ini. Tahap pertama adalah ekstraksi fitur dengan tujuan untuk mendapatkan fitur-fitur yang akan digunakan untuk melakukan klasifikasi. Fitur yang akan digunakan adalah *standard deviation*, *mean*, *skewness*, *entropy*, *kurtosis* dan *grayscale* dari histogram warna, histogram *grayscale* dan histogram tingkat saturasi. Tahap selanjutnya dilakukan proses klasifikasi. Metoda yang digunakan untuk pengenalan pola dalam penelitian ini adalah *support vector machine (SVM)*.

A. Histogram warna (Color Histogram)

Untuk membangun histogram warna dari citra parasit malaria *plasmodium falciparum* dan *plasmodium vivax*, terlebih dahulu citra tersebut dinormalisasi. Citra dipetakan pada *pixel* dengan ukuran 50 x 50. Tujuan dari normalisasi adalah mengurangi resolusi citra yang berguna saat proses pengenalan citra dan juga meningkatkan akurasi pengenalan. Persamaan yang digunakan untuk normalisasi warna tersebut adalah sebagai berikut :

$$NormalizedRed = \frac{Red}{\sqrt{Red^2 + Green^2 + Blue^2}} \quad (1)$$

$$NormalizedGreen = \frac{Green}{\sqrt{Red^2 + Green^2 + Blue^2}} \quad (2)$$

$$NormalizedBlue = \frac{Blue}{\sqrt{Red^2 + Green^2 + Blue^2}} \quad (3)$$

Dari hasil normalisasi ini digunakan red, green dan blue untuk kemudian dibangun histogramnya.

B. Histogram Tingkat Keabuan (grayscale histogram)

Nilai citra parasit malaria merupakan model warna RGB. Untuk mendapatkan nilai tingkat keabuan dari citra yang terdiri dari komponen warna RGB dilakukan dengan menggunakan persamaan berikut :

$$Grayscale = 0.42R + 0.32G + 0.28B \quad (4)$$

Hasil dari proses *grayscale* ini akan berada pada tingkat keabuan sebesar sebesar 8 bit. Distribusi nilai-nilai dari setiap piksel citra *grayscale* dimasukkan ke dalam histogram.

c. Histogram Tingkat Saturasi (Saturation Level Histogram)

Histogram tingkat saturasi digunakan untuk mendapatkan nilai-nilai intensitas warna berdasarkan kejenuhannya (saturasi). Komponen warna berdasarkan kejenuhannya diperoleh dari citra eritrosit model warna RGB melalui perhitungan dengan persamaan sebagai berikut :

$$Saturasi = 1 - \frac{3}{Red + Green + Blue} \min(R, G, B) \quad (5)$$

Dimana nilai saturasi pada setiap piksel ini digunakan untuk membangun histogram distribusinya.

Dari nilai histogram warna, grayscale dan tingkat saturasi hasil perhitungan sudah dapat dijadikan sebagai vektor *input*, namun untuk mengurangi masalah komputasi yang besar maka nilai-nilai tersebut diwakili oleh nilai *mean*, standar deviasi, *kurtosis* dan *skewness* dari distribusinya histogram tersebut, dimana dapat dihitung dengan persamaan :

$$Mean = \mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (6)$$

$$\text{Standar deviasi} = s = \left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right)^{\frac{1}{2}} \quad (7)$$

$$\text{Kurtosis} = k = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right)^2} \quad (8)$$

$$\text{Skewness} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right)^{\frac{3}{2}}} \quad (9)$$

Untuk pencarian nilai *entropy* yang digunakan adalah matriks *co-occurrence*. Matrik *co-occurrence* ini digunakan untuk menggambarkan frekuensi munculnya pasangan dua atau beberapa *pixel* dengan intensitas tertentu dalam citra. Hasil yang di dapat akan ditemukan nilai keteracakan (*entropy*) dari distribusi intensitas tersebut, dimana persamaannya adalah :

$$\text{Entropy} = - \sum_{i_1} \sum_{i_2} p(i_1, i_2) \log p(i_1, i_2) \quad (10)$$

dengan $p(i_1, i_2)$ adalah matriks *co-occurrence* dari citra parasit malaria.

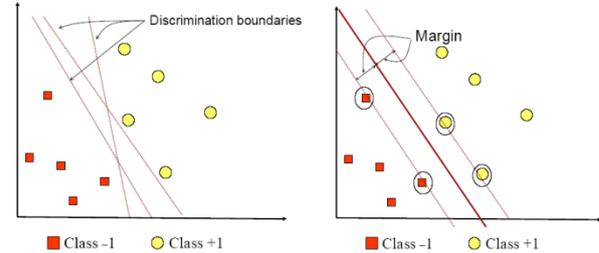
C. Support Vector Machine (SVM)

Support Vector Machine (SVM) dikembangkan oleh Boser, Guyon, dan Vapnik, pertama kali dipresentasikan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. Konsep dasar *SVM* sebenarnya merupakan kombinasi harmonis dari teori-teori komputasi yang telah ada puluhan tahun sebelumnya, seperti *margin hyperplane* (Duda & Hart tahun 1973, Cover tahun 1965, Vapnik 1964, dan sebagainya), *kernel* diperkenalkan oleh Aronszajn tahun 1950, demikian juga dengan konsep-konsep pendukung yang lain. Akan tetapi hingga tahun 1992, belum pernah ada upaya merangkaikan komponen-komponen tersebut.

Berbeda dengan strategi *neural network* yang berusaha mencari *hyperplane* pemisah antar kelas, *SVM* berusaha menemukan *hyperplane* yang terbaik pada *input space*. Prinsip dasar *SVM* adalah *linear classifier*, dan selanjutnya dikembangkan agar dapat bekerja pada problem *non-linear* dengan memasukkan konsep *kernel trick* pada ruang kerja berdimensi tinggi. Perkembangan ini memberikan rangsangan minat penelitian di bidang *pattern recognition* untuk investigasi potensi kemampuan *SVM* secara teoritis maupun dari segi aplikasi. Saat ini *SVM* telah berhasil diaplikasikan dalam masalah dunia nyata (*real-world problems*), dan secara umum memberikan solusi yang lebih baik dibandingkan metode konvensional seperti misalnya *artificial neural network* [4].

Konsep *SVM* dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada *input space*. *Hyperplane* dalam ruang vektor berdimensi d adalah *affine subspace* berdimensi $d-1$ yang membagi ruang vektor tersebut ke dalam dua bagian, yang masing-masing berkorespondensi pada kelas yang berbeda[4]. Gambar 1 memperlihatkan beberapa *pattern* yang merupakan anggota dari dua buah kelas :

+1 dan -1. *Pattern* yang tergabung pada kelas -1 disimbolkan dengan warna merah (kotak), sedangkan *pattern* pada kelas +1, disimbolkan dengan warna kuning (lingkaran). Masalah klasifikasi dapat diterjemahkan dengan usaha menemukan garis (*hyperplane*) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) ditunjukkan pada gambar 1 (a).



Gambar 1. *Hyperplane* yang memisahkan dua kelas (-1 dan +1)

Hyperplane pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* tersebut dengan *pattern* terdekat dari masing-masing kelas. *Pattern* yang paling dekat ini disebut sebagai *support vector*. Garis solid pada gambar 1 (b) menunjukkan *hyperplane* yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua kelas, sedangkan titik merah dan kuning yang berada dalam lingkaran hitam adalah *support vector*. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pembelajaran pada *SVM*. Data yang tersedia dinotasikan sebagai $\vec{x}_i \in R^d$, sedangkan label masing-masing dinotasikan $\vec{y}_i = \{+1, -1\}$ untuk $i=1, 2, 3 \dots l$. Dengan l adalah banyaknya data. Diasumsikan kedua kelas -1 dan +1 dapat terpisah secara sempurna oleh *hyperplane* berdimensi d , yang didefinisikan :

$$\vec{w} \cdot \vec{x} + b = 0 \quad (11)$$

Pattern \vec{w} yang termasuk kelas -1 (sampel negatif) dapat dirumuskan sebagai *pattern* yang memenuhi pertidaksamaan

$$\vec{w} \cdot \vec{x}_i + b \leq -1 \quad (12)$$

Sedangkan *pattern* \vec{w} yang termasuk kelas +1 (sampel positif)

$$\vec{w} \cdot \vec{x}_i + b \geq +1 \quad (13)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu $1/\|\vec{w}\|$. Hal ini dapat dirumuskan sebagai *Quadratic Programming* (QP) *problem*, yaitu mencari titik minimal persamaan (14), dengan memperhatikan *constraint* persamaan (15).

$$\min_w \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \quad (14)$$

dengan

$$y_i(\bar{x}_i \cdot \bar{w} + b) - 1 \geq 0, \forall i \tag{15}$$

Masalah ini dapat dipecahkan dengan berbagai teknik komputasi, di antaranya *Lagrange Multiplier*.

$$L(\bar{w}, b, \alpha) = \frac{1}{2} \|\bar{w}\|^2 - \sum_{i=1}^l \alpha_i (y_i ((\bar{x}_i \cdot \bar{w} + b) - 1))$$

dengan $i = 1, 2, \dots, l$. (16)

α_i adalah *Lagrange multipliers*, yang bernilai nol atau positif ($\alpha_i \geq 0$). Nilai optimal dari persamaan (6) dapat dihitung dengan meminimalkan L terhadap \bar{w} dan b , dan memaksimalkan L terhadap α_i . Dengan memperhatikan sifat bahwa pada titik optimal gradien L = 0, persamaan (16) dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung saja α_i , sebagaimana persamaan (17) di bawah.

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \bar{x}_i \bar{x}_j \tag{17}$$

dengan

$$\alpha_i \geq 0 (i = 1, 2, \dots, l) \quad \sum_{i=1}^l \alpha_i y_i \tag{18}$$

Dari hasil dari perhitungan ini diperoleh α_i yang kebanyakan bernilai positif. Data yang berkorelasi dengan α_i yang positif inilah yang disebut sebagai *support vector* (Nugroho, 2003).

Metode SVM *one against one* adalah salah satu metode untuk mengimplementasi SVM untuk *multiclass* dengan menggunakan pendekatan yang kedua. Model klasifikasi biner yang dibangun menggunakan metode ini dapat dihitung dengan mengikuti persamaan (19) :

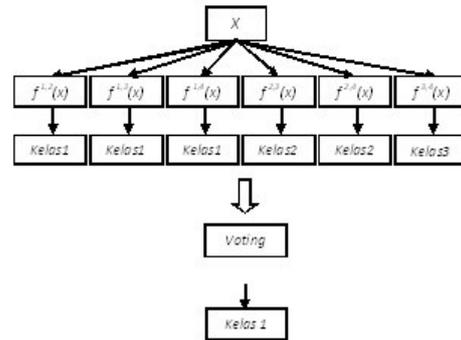
$$\frac{k(k-1)}{2} \tag{19}$$

dengan k adalah banyaknya kelas. Pada tahap pelatihan, setiap model klasifikasi dilatih menggunakan data latih dari dua kelas. Sedangkan pada tahap pengujian terdapat beberapa cara untuk melakukan pengujian setelah semua $k(k-1)/2$ model klasifikasi telah selesai dibangun. Salah satunya cara yang dapat digunakan adalah dengan menggunakan metode *voting* (Hsu, 2002). Contoh penggunaan metode SVM *one against one* dapat ditunjukkan pada tabel 1 dan gambar 2.

TABEL 1
CONTOH METODE ONE AGAINST ONE

$y_i = 1$	$y_i = -1$	Hipotesis
Kelas 1	Kelas 2	$f^{1,2}(x) = (w^{1,2})x + b^{1,2}$
Kelas 1	Kelas 3	$f^{1,3}(x) = (w^{1,3})x + b^{1,3}$
Kelas 1	Kelas 4	$f^{1,4}(x) = (w^{1,4})x + b^{1,4}$
Kelas 2	Kelas 3	$f^{2,3}(x) = (w^{2,3})x + b^{2,3}$

Kelas 2	Kelas 4	$f^{2,4}(x) = (w^{2,4})x + b^{2,4}$
Kelas 3	Kelas 4	$f^{3,4}(x) = (w^{3,4})x + b^{3,4}$



Gambar 2 Metode klasifikasi SVM *one against one*

Dari gambar 2 jika data x_i dimasukkan ke dalam fungsi yang didapatkan dari tahap pelatihan pada persamaan 20 :

$$f(x) = (w^{ij})^T \phi(x) + b \tag{20}$$

dan hasil yang didapatkan x adalah kelas termasuk kelas i , maka kelas i mendapatkan satu suara (*vote*). Dan selanjutnya data x_i diujikan ke semua model klasifikasi yang didapatkan dari tahap pelatihan. Dan pada akhirnya kelas dari data x ditentukan dari jumlah perolehan suara terbanyak. Apabila terdapat dua buah kelas yang memiliki jumlah suara yang sama, maka kelas dengan indeks yang lebih kecil dinyatakan sebagai kelas dari data yang diujikan.

Metode ini akan membangun sejumlah k SVM biner, dimana k adalah banyaknya kelas (Hsu, et al., 2002). SVM ke- i dilatih dengan seluruh *sample* pada kelas ke- i dengan label kelas positif dan seluruh *sample* lainnya dengan label kelas negatif. Jika diberikan l data pelatihan $(x_i, y_i), \dots, (x_l, y_l)$, dengan $x_i \in \mathbb{R}^n, i = 1, \dots, l$ adalah kelas dari x_i , maka SVM ke- i akan menyelesaikan permasalahan berikut :

$$\begin{aligned} \min_{w^i, b^i, \xi^i} & \frac{1}{2} (w^i)^T w^i + C \sum_{j=1}^l \xi_j^i (w^i)^T \\ & (w^i)^T \phi(x_j) + b^i \geq 1 - \xi_j^i, \text{ if } y_j = i \\ & (w^i)^T \phi(x_j) + b^i \leq -1 + \xi_j^i, \text{ if } y_j \neq i \\ & \xi_j^i \geq 0, j = 1, \dots, l \end{aligned} \tag{21}$$

dengan data pelatihan x_i dipetakan ke ruang dimensi yang lebih tinggi menggunakan fungsi Φ dan C sebagai parameter pinalti.

Meminimisasi $\frac{1}{2} (w^i)^T w^i$ berarti memaksimalkan

$\frac{2}{|w|^2}$ atau margin antara dua kelompok data. Ketika

data tidak terpisah secara linier, maka terdapat pinalti sebesar $C \sum_j \xi_j^i$ yang dapat mengurangi jumlah *error* pelatihan. Ide dari SVM adalah menyeimbangkan regulasi $\frac{1}{2} (w^i)^T w^i$ dan *error* pelatihan.

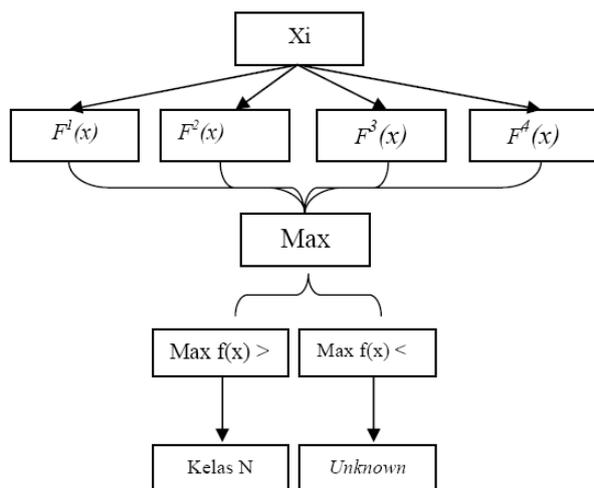
Setelah menyelesaikan permasalahan pada minimisasi, maka terdapat sejumlah k fungsi keputusan.

$$f^1(x) = (w^1)x + b^1, \dots, f^k(x) = (w^k)x + b^k \quad (22)$$

Kelas data x akan ditentukan berdasarkan nilai fungsi keputusan yang tertinggi. Untuk pencarian solusi minimisasi pada persamaan 22 menggunakan *quadratic programming*.

TABEL 2
CONTOH METODE ONE AGAINST ALL

$y_i = 1$	$y_i = -1$	Hipotesis
Kelas 1	Bukan kelas 1	$f^1(x) = (w^1)x + b^1$
Kelas 2	Bukan kelas 2	$f^2(x) = (w^2)x + b^2$
Kelas 3	Bukan kelas 3	$f^3(x) = (w^3)x + b^3$
Kelas 4	Bukan kelas 4	$f^4(x) = (w^4)x + b^4$



Gambar 3 Metode klasifikasi SVM *one against all*

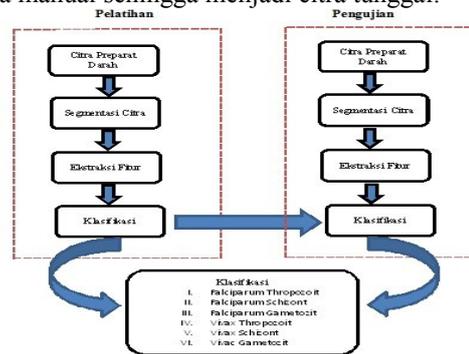
Pada gambar 3 terlihat terdapat empat fungsi keputusan, dari keempat fungsi keputusan diambil fungsi keputusan yang maksimal.

III. METODE PENELITIAN

Tahapan-tahapan dari metode yang diusulkan untuk melakukan klasifikasi dari jenis dan fase *plasmodium falciparum* dan *plasmodium vivax* dari citra medis preparat darah ditunjukkan dalam Gambar 3.

Pada proses pembelajaran, citra medis dari preparat darah adalah file-file yang masing-masing berukuran 50x50 *pixel* yang diperoleh dari hasil segmentasi secara manual. Artinya, setelah file citra preparat darah terinfeksi parasit malaria yaitu citra *falciparum thropozoit*, *falciparum schizont*, *falciparum*

gametocyt, *vivax thropozoit*, *vivax schizont*, *vivax gametocyt* diperoleh *public health image library* (PHIL) dari *phil.cdc.gov*, proses normalisasi seperti *cropping* dan *resize* untuk menyamakan dimensi citra dilakukan secara manual sehingga menjadi citra tunggal.



Gambar 4 Blok diagram desain sistem

Tahap selanjutnya ekstraksi fitur dengan tujuan untuk mendapatkan fitur-fitur yang akan digunakan untuk melakukan klasifikasi. Fitur yang akan digunakan adalah *mean*, *standard deviation*, *kurtosis*, *skewness*, dan *entropy* dari Histogram Warna, Histogram *Grayscale* dan Histogram Tingkat Saturasi. Selanjutnya dilakukan proses klasifikasi. Metoda yang digunakan untuk pengenalan pola dalam penelitian ini adalah *support vector machine (SVM)*. Pada dasarnya SVM hanya dapat digunakan untuk mengklasifikasikan data ke dalam dua kelas (klasifikasi biner). Untuk dapat diterapkan pada permasalahan klasifikasi jenis dan fase dari parasit malaria *plasmodium falciparum* dan *plasmodium vivax* yang terdiri lebih dari dua kelas maka diterapkan metode klasifikasi SVM multikelas yang dibangun dengan cara menggabungkan beberapa klasifikasi biner. Metode pendekatan SVM multikelas yang digunakan dalam penelitian ini adalah SVM *multiclass one against one (OAO)* dan *one against all(OAA)*. Pada proses klasifikasi pelatihan, variabel *hyperplane* untuk setiap pengklasifikasi (*classifier*) yang didapat akan disimpan dan nantinya akan digunakan sebagai data tiap pengklasifikasi dalam proses pengujian, dengan kata lain proses klasifikasi pelatihan adalah proses untuk mencari *support vector*, *alpha* dan *bias* dari data *input* pelatihan yang berupa vektor fitur dari citra *falciparum thropozoit*, *falciparum schizont*, *falciparum gametocyt*, *vivax thropozoit*, *vivax schizont*, *vivax gametocyt* (enam kelas). Sedangkan pada proses pengujian, data citra *plasmodium falciparum* dan *plasmodium vivax* yang digunakan adalah data yang tidak diikutsertakan pada proses pelatihan. Jika kelas yang dihasilkan dari proses klasifikasi pengujian sama dengan kelas data sebenarnya, maka pengenalan dinyatakan benar.

IV. HASIL DAN PEMBAHASAN

Pada bagian ini akan dijabarkan dan dievaluasi efektifitas metode yang digunakan dalam mengatasi permasalahan klasifikasi terhadap jenis parasit yang menyebabkan malaria yaitu *plasmodium falciparum* dan *plasmodium vivax* beserta fasenya.

Ujicoba terhadap metode klasifikasi SVM digunakan 180 data citra preparat darah dengan ukuran 50x50 pixel. Setiap kelas dari klasifikasi terdiri dari 30 data. Dalam proses ujicoba dilakukan dua tahap, tahap yang pertama adalah pelatihan sedangkan tahap yang kedua adalah tahap uji. Tahap pelatihan digunakan untuk mendapatkan koordinat dari *support vector*, *weight*, *bias* dan jarak *support vector*, sedangkan tahap pengujian adalah menggunakan data-data selain data pelatihan untuk mendapatkan hasil klasifikasi, sehingga dapat diketahui tingkat akurasi.

Pengujian ini dilakukan terhadap semua kelas berdasarkan kelompok data pelatihan dan pengujian yang telah dibagi, kelas yang pertama adalah *Falciparum Thropozoit*, kelas kedua adalah *Falciparum Schizont*, kelas ketiga adalah *Falciparum Gametocyt*, kelas keempat adalah *Vivax Thropozoit*, kelas kelima adalah *Vivax Schizont* dan kelas keenam adalah *Vivax Gametocyt*. Secara singkat enam kelompok data ini di rotasi tanpa terjadi overlap sehingga semua citra darah pernah menjadi data pelatihan maupun data pengujian. Distribusi data pelatihan dan pengujian dapat dilihat pada tabel 4.

TABEL 4
DISTRIBUSI DATA PELATIHAN DAN
PENGUJIAN

Data	Kelompok I						Kelompok II						Kelompok III					
	Kelas						Kelas						Kelas					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI	I	II	III	IV	V	VI
Latih	Citra						Citra						Citra					
	1 sd 20						11 sd 30						1 sd 10 dan 21 sd 30					
	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Uji	Citra						Citra						Citra					
	21 sd 30						1 sd 10						11 sd 30					
	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
	60						60						60					

Proses pelatihan dan pengujian data dilakukan berdasarkan urutan langkah sebagai berikut :

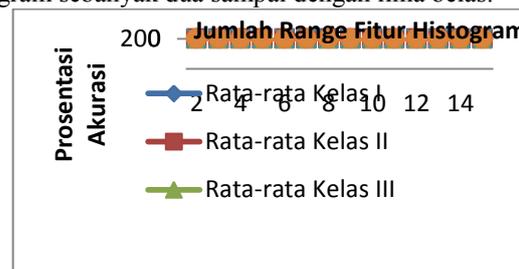
1. Membagi proses pelatihan dan pengujian menjadi 3 buah proses.
2. Proses pertama, pada masing-masing kelas digunakan 20 data awal sebagai data latih dan 10 data terakhir sebagai data uji.
3. Proses kedua, pada masing-masing kelas digunakan data ke 11 sampai 30 sebagai data latih dan data ke 1 sampai 10 sebagai data uji.
4. Proses ketiga, pada masing-masing kelas digunakan data ke 1 sampai 10 dan 21 sampai 30 sebagai data latih dan data ke 11 sampai 20 sebagai data uji.

A. Percobaan Pengaruh Range Fitur Histogram Terhadap Hasil Klasifikasi Menggunakan Support Vector Machine One Against One

Jumlah fitur histogram yang digunakan pada tahap ekstraksi fitur merupakan panjang vektor pola yang menjadi input pada tahap klasifikasi menggunakan *support vector machine* multikelas *one against one*, baik

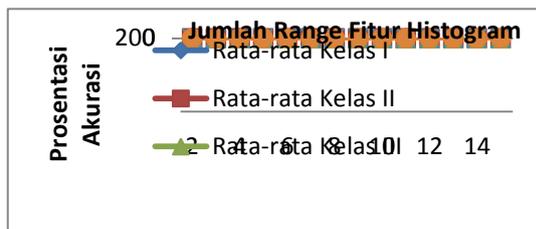
pada proses pelatihan maupun pada tahap pengujian. Dengan demikian pengujian ini bertujuan untuk mengetahui seberapa besar pengaruh jumlah fitur histogram yang digunakan terhadap kinerja dari sistem klasifikasi.

Percobaan ini dilakukan dengan cara membuat grafik prosentase akurasi sistem klasifikasi terhadap jumlah *range* fitur histogram yang digunakan. Rentang jumlah *range* fitur histogram yang digunakan dalam pengujian ini dibatasi dari dua hingga lima belas *range* fitur. Grafik yang pertama adalah grafik rata-rata akurasi per kelas yang menggambarkan nilai akurasi masing-masing kelas yang didapatkan dari hasil rata-rata akurasi ketiga tahap pengujian, dapat dilihat pada gambar 4. Grafik yang kedua adalah grafik akurasi multi kelas yang digambarkan untuk masing-masing data pelatihan dan pengujian yang menggunakan kombinasi data pelatihan dan pengujian yang berbeda, dapat dilihat pada gambar 5. Dan yang terakhir gambar 6 adalah grafik rata-rata akurasi multi kelas yang merupakan nilai rata-rata dari ketiga tahap pengujian klasifikasi. Perhitungan hasil klasifikasi parasit jenis dan fase malaria *falciparum* dan malaria *vivax* secara keseluruhan yang menggunakan jumlah *range* fitur histogram sebanyak dua sampai dengan lima belas.



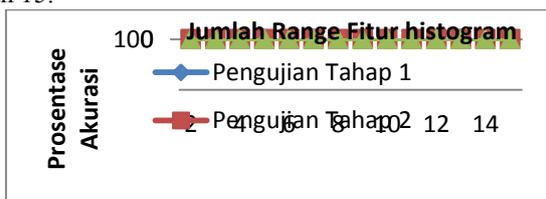
Gambar 5 Grafik Rata-rata Akurasi per kelas svm one vs one

Pada gambar 5 ditunjukkan bahwa prosentase akurasi rata-rata terendah untuk kelas I sebesar 78,07%, kelas II sebesar 81.11%, kelas III sebesar 80.56%, Kelas IV sebesar 72.78%, Kelas V sebesar 77,77% dan Kelas VI sebesar 78,33%. Sedangkan prosentase akurasi rata-rata yang tertinggi untuk kelas I adalah 96,11% didapatkan saat menggunakan jumlah range fitur 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, dan 14. Untuk kelas II adalah 95,56% didapatkan saat menggunakan jumlah range fitur 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, dan 15. Untuk kelas III adalah 93,33% didapatkan saat menggunakan jumlah 2, 3, 4, 5, 6, 7, 8, 9, 10, dan 11. Untuk kelas IV adalah 88,88% didapatkan saat menggunakan jumlah 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, dan 15. Untuk kelas V adalah 92,78% didapatkan saat menggunakan jumlah 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, dan 14. Dan Untuk kelas VI adalah 86,11% didapatkan saat menggunakan jumlah 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, dan 12.



Gambar 5 Grafik Rata-rata Akurasi per kelas svm one vs one

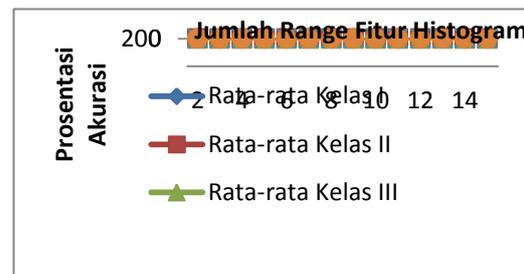
Pada gambar 6 ditunjukkan bahwa prosentase akurasi multi kelas untuk masing-masing tahap pengujian (kelompok data pengujian) yang terendah didapatkan pada saat histogram menggunakan range fitur dua. Prosentase akurasi multi kelas terendah pada pengujian tahap 1 sebesar 40,67%, pengujian tahap 2 sebesar 35,59%, pengujian tahap 3 sebesar 30%. Sedangkan prosentase akurasi multi kelas yang tertinggi pada pengujian tahap 1 adalah 78,33% didapatkan saat menggunakan jumlah range fitur 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, dan 14. Pada pengujian tahap 2 adalah 65% didapatkan saat menggunakan jumlah range fitur 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 dan 15. Pada pengujian tahap 3 adalah 81,35% didapatkan saat menggunakan jumlah range fitur 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 dan 15.



Gambar 6 Grafik akurasi multi kelas svm one vs one untuk setiap tahap pengujian

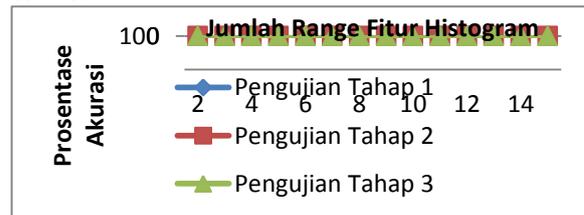
B. Percobaan Pengaruh Range Fitur Histogram Terhadap Hasil Klasifikasi Menggunakan Support Vector Machine One Against All

Pada gambar 7 ditunjukkan bahwa prosentase akurasi rata-rata terendah untuk kelas I sebesar 80%, kelas II sebesar 85,56%, kelas III sebesar 81,67%, Kelas IV sebesar 72,78%, Kelas V sebesar 51,67% dan Kelas VI sebesar 80%. Sedangkan prosentase akurasi rata-rata yang tertinggi untuk kelas I adalah 95,56% didapatkan saat menggunakan jumlah range fitur 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, dan 15. Untuk kelas II adalah 92,78% didapatkan saat menggunakan jumlah range fitur 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, dan 15. Untuk kelas III adalah 91,67% didapatkan saat menggunakan jumlah 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, dan 13. Untuk kelas IV adalah 87,78% didapatkan saat menggunakan jumlah 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, dan 15. Untuk kelas V adalah 90% didapatkan saat menggunakan jumlah 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, dan 13. Dan Untuk kelas VI adalah 84,44% didapatkan saat menggunakan jumlah 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, dan 15.



Gambar 7 Grafik Rata-rata Akurasi per kelas svm one vs all

Pada gambar 8 ditunjukkan bahwa prosentase akurasi multi kelas untuk masing-masing tahap pengujian (kelompok data pengujian) yang terendah didapatkan pada saat histogram menggunakan range fitur dua. Prosentase akurasi multi kelas terendah pada pengujian tahap 1 sebesar 35%, pengujian tahap 2 sebesar 25%, pengujian tahap 3 sebesar 28,33%. Sedangkan prosentase akurasi multi kelas yang tertinggi pada pengujian tahap 1 adalah 65% didapatkan saat menggunakan jumlah range fitur 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, dan 14. Pada pengujian tahap 2 adalah 66,67% didapatkan saat menggunakan jumlah range fitur 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 dan 15. Pada pengujian tahap 3 adalah 81,35% didapatkan saat menggunakan jumlah range fitur 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 dan 15.



Gambar 8 Grafik akurasi multi kelas svm one vs all untuk setiap tahap pengujian

V. KESIMPULAN

Pada penelitian ini telah dikembangkan sistem klasifikasi jenis dan fase parasit malaria plasmodium falciparum dan plasmodium vivax dengan ekstraksi fitur mean, standard deviation, kurtosis, skewness, dan entropy dari histogram warna, histogram grayscale dan histogram tingkat saturasi dari citra parasit malaria dan metode support vector machine sebagai classifier-nya.

Dari hasil percobaan terhadap 180 data citra parasit malaria dinyatakan bahwa klasifikasi menggunakan support vector machine kernel linier metode one against one jika dibandingkan dengan metode SVM one against all didapatkan tingkat akurasi dari support vector machine kernel linier metode one against one yaitu falciparum thropozoit 95,55%, falciparum schizont 94,44%, falciparum gametocyte 91,11%, vivax thropozoit 88,89%, vivax schizont 93,88% dan vivax gametocyte 90%, sedangkan backpropagation neural network didapatkan tingkat akurasi falciparum thropozoit 92,77%, falciparum schizont 91,11%, falciparum gametocyte 90%, vivax thropozoit 87,22%,

vivax schizont 88,33% dan *vivax gametocyte* 85,55% sehingga dapat disimpulkan bahwa klasifikasi menggunakan *support vector machine* kernel *linier* metode *one against one* memiliki keunggulan dalam tingkat akurasi dibandingkan dengan metode SVM *one against all*.

Allergy and Infectious Diseases (2007), “*Understanding malaria Fighting an Ancient Scourge*”, Bethesda.

- [15] Wahab, Iis Hamsir Ayub (2008), “Identifikasi parasit malaria dalam darah menggunakan segmentasi citra digital dan jaringan syaraf tiruan”.

DAFTAR PUSTAKA

- [1] Ahmad Usman (2005) “Pengolahan Citra Digital dan Teknik Pemrogramannya”. Graha Ilmu, Yogyakarta.
- [2] Albrechtsen, Fritz, *Statistical Texture Measures Computed from Gray Level Cooccurrence* <http://www.ifi.uio.no/in384/info/glcm.ps>, (download: desember 10, 2012).
- [3] Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992). “*A training algorithm for optimal margin classifiers*”. D. Haussler, editor, 5th Annual ACM Workshop on COLT, pages 144–152, Pittsburgh.
- [4] Chunyan Yao, Jianwei Zhang, Houxiang Zhang (2007), “ Blood Cell Identification and Segmentation by means of Statistical Models”. Proceeding of the 7th WSEAS Int Conf. on Signal processing, Computational Geometry & Artificial Vision, Athens Greece, August 24-26, 2007.
- [5] Christiani, Nello, and Taylor, J.S., (2000), “An Introduction to Support Vector Machine and Other kernel-based Learning Methods”, Cambridge University Press.
- [6] Díaz, G., González, Fabio A., Romero, Eduardo (2009),” A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images”, J. of Biomedical Informatics,42:296–307.
- [7] Di Ruberto, Cecilia, Dempster, Andrew, Khan, Shahid, Jarra , Bill (2002), “*Analysis of infected blood cell images using morphological operators*”, Image and Vision Computing, 20:133-146.
- [8] Direktorat Jenderal Pengendalian Penyakit dan Penyehatan Lingkungan Departemen Kesehatan Republik Indonesia (2008), “*Pedoman Ketatalaksanaan Kasus Malaria di Indonesia*”
- [9] Gonzalez, R. C., Woods, R. E. (2010), “*Digital Image Processing*”, Third Edition, Pearson Prentice Hall, Inc.
- [10] Haralick, R.M. et., 1973, “Texture Features for Image Classification, IEEE Transaction on System, Man, and Cybernetics, Vol. SMC-3
- [11] Hall-Beyer, Mryka, *The GLCM Tutorial HomePage*, [Online], URL: http://www.fp.ucalgary.ca/mhallbey/the_glcm.html, (download: Desember 10, 2012)
- [12] Hsu, C.W., and Lin, C.J. (2002), “*A Comparison of Methods for Multi-class Support Vector Machines*”, IEEE Transactions on Neural Networks, Vol. 13, No. 2, hal. 415-425.
- [13] Nugroho, A.S., Witarto, B.A., Handoko, D., (2003), *Support Vector Machine – Teori dan Aplikasinya Dalam Bioinformatika*, Kuliah Umum Ilmu Komputer.com.
- [14] U.S. Department of Health and Human Services National Institutes of Health National Institute of