Data Science : Journal Of Computing And Applied Informatics
Journal homepage: https://jocai.usu.ac.id

Identification Of Malaria Parasites Plasmodium Vivax on Red Blood Cells Using the Probabilistic Neural Network Method

Sindi Lioni Aritonang[1], Viska Mutiawani[2*]
[1]*Faculty of Computer Science and Information Technology, University of Sumatera Utara*
[2]*The University of Western Australia*
[*]Corresponding Author: viska.mutiawani@research.uwa.edu.au

ARTICLE INFO
Article history:
Received 28 January 2025
Revised 19 March 2025
Accepted 24 May 2025
Available online 31 July 2025

E-ISSN: 2580-829X
P-ISSN: 2580-6769

How to cite:
Aritonang S. L., V. Mutiawani, "Identification Of Malaria Parasites Plasmodium Vivax on Red Blood Cells Using the Probabilistic Neural Network Method", Data Science: Journal of Computing and Applied Informatics, vol. V9, no. 2, Jul. 2025, doi: 10.32734/jocai.v9.i2-22535

http://doi.org/10.32734/jocai.v9.i2-22535

ABSTRACT
Malaria is a disease that is infects human red blood cells transmitted through the bite of a female Anopheles mosquito that contains the parasite genus Plasmodium. Plasmodium vivax is one of the types of parasites that causes malaria, which is known as the type of malaria with the widest distribution area, from tropical, subtropical to cold climates. The diagnosis of malaria, basically depend on microscopic analysis of Giemsa-smeared thin and thick films of blood. However, this diagnostic method is time consuming and prone to human error. To overcome this problem, a method is needed to automatically identify malaria parasites on red blood cells. This study proposes to identifying the malaria parasite Plasmodium vivax using the Probabilistic Neural Network method. The steps taken before identification are preprocessing using Green Channel, Contrast Limited Adaptive Histogram Equalization (CLAHE), Morphological Close and Background Exclusion, then segmentation with Otsu Thresholding, next step is post- processing with Connected Component Analyst (CCA) and feature extraction with Invariant Moment. The results of this research showed that the method used was able to identify the malaria parasite plasmodium vivax on microscopic images of reb blood cells with an accuracy rate of 97.14%, and sensitivity of 95%.
Keyword: Connected Component Analyst, Contrast Limited Adaptive Histogram Equalization, Malaria, Plasmodium Vivax, Probabilistic Neural Network.

ABSTRAK
Malaria adalah penyakit yang menyerang sel darah merah manusia dan ditularkan melalui gigitan nyamuk Anopheles betina yang mengandung parasit genus Plasmodium. Plasmodium vivax adalah salah satu jenis parasit yang menyebabkan malaria, yang dikenal sebagai jenis malaria dengan area penyebaran terluas, mulai dari iklim tropis, subtropis hingga iklim dingin. Diagnosis malaria pada dasarnya bergantung pada analisis mikroskopis dari sediaan darah tipis dan tebal yang diwarnai dengan Giemsa. Namun, metode diagnostik ini memakan waktu dan rentan terhadap kesalahan manusia. Untuk mengatasi masalah ini, diperlukan metode untuk secara otomatis mengidentifikasi parasit malaria pada sel darah merah. Studi ini mengusulkan untuk mengidentifikasi parasit malaria Plasmodium vivax menggunakan metode Jaringan Saraf Probabilistik. Langkah-langkah yang dilakukan sebelum identifikasi meliputi pra-pemrosesan menggunakan Green Channel, Contrast Limited Adaptive Histogram Equalization (CLAHE), Morphological Close, dan Background Exclusion, kemudian segmentasi dengan Otsu Thresholding, dilanjutkan dengan pasca-pemrosesan menggunakan Connected Component Analyst (CCA) dan ekstraksi fitur dengan Invariant Moment. Hasil penelitian ini menunjukkan bahwa metode yang digunakan mampu mengidentifikasi parasit malaria Plasmodium vivax pada gambar mikroskop sel darah merah dengan tingkat akurasi 97,14% dan sensitivitas 95%.
Kata kunci: Analisis Komponen Terhubung, Kontras Terbatas, Penyeimbangan Histogram Adaptif, Malaria, Plasmodium Vivax, Jaringan Saraf Probabilistik.

1. Introduction

Malaria is a disease that is infects human red blood cells transmitted through the bite of a female Anopheles mosquito that contains the parasite genus Plasmodium [1]. Plasmodium vivax is one of the types of parasites that causes malaria, which is known as the type of malaria with the widest distribution area, from tropical, subtropical to cold climates. Plasmodium vivax is also known for causing relapse if not treated properly [2]. According to the World Health Organization (WHO) in 2020, 241 million people were infected with malaria and 627 thousand of them had died. In Indonesia, the malaria morbidity rate in 2020 is 254 thousand of which 39% of cases are Plasmodium vivax malaria infections with the highest endemic areas in Papua, West Papua, East Nusa Tenggara, East Kalimantan, and North Sumatra [3]. The diagnosis of malaria, basically depend on microscopic analysis of Giemsa-smeared thin and thick films of blood. Thick blood smears are usually used to determine the density of malaria parasites in red blood cells and thin blood smears are used to determine the species and phase of parasites in malaria. Species and phases are easier to identify because the structure of the parasite on a thin blood smear is more clearly visible [4]. However, this diagnostic method is time consuming and prone to human error [5].

With the development of medical knowledge-based systems, the demand to use computer knowledge systems as an analytical method for diagnosing diseases is becoming increasingly important. This study proposes to identify the malaria parasite Plasmodium Vivax on microscopic images of thin blood smears using the Probabilistic Neural Network method. Probabilistic Neural Network is based on two methods, namely Bayes's theory method and Parzen method which calculates the probability density function of arbitrary variables. PNN is a useful classifier method mapping any input pattern to a large number of classifications. Probabilistic Neural Network (PNN) is used because it has the ability to process data faster and is proven to produce a fairly high level of accuracy in identifying an object [6].

Yunda L., Alarcón A. Millán J. (2011), previously conducted a study on the identification of the malaria parasite Plasmodium vivax on thick blood smears using the Multilayer Neural Network method. Using a combination of Absence of Gradients and Nerstian Equilibrium Stripping (AGNES) and Morphological Gradient techniques as the segmentation process, and Wavelet Transform as feature extraction, with an accuracy rate of 77.19% [7]. In addition, Kristofer E, Rivera P, Naval P. (2017), have also identified two species of Plasmodium vivax and falcifarum through thin blood smears using the Convolutional Neural Network algorithm. The study used Morphological Operations, Binary Thresholding and Connected Components Analysis, with an accuracy of 92.4% and a sensitivity of 95.2% [8].

In conjunction with the techniques used in this study, Gummadadila A, Deepthi A, Sowmya B, Bai A. K, Reddy P. (2020), used the Probabilistic Neural Network to classify 5 types of white blood cells including Lymphocyte, Monocyte, Basophil, Eosinophil, and Neurtophil, with an accuracy of 95% [9]. Another study, Syahputra D (2018) who also uses Probabilistic Neural Network and Invariant Moment as feature extraction to classify facial shapes in men automatically, with 80% accuracy results [10].

2. Method

This study uses the Probabilistic Neural Network as a research method to identify the malaria parasite Plasmodium vivax in human red blood cells through a thin blood smear. Probabilistic Neural Network (PNN) is defined as the application of statistical algorithms to analyze kernel discrimination where the operations are organized into a multi- layer feed forward network with four layers: input layer, pattern layer, summation layer, and output layer. The advantages of this method are: The training data processing process is faster, the structure is parallel and cannot be separated, it shows the optimal classification, and the training data samples can be edited, deleted or added without retraining. Therefore, compared to other neural networks, PNN classification learns faster and is used in many other applications. Based on these advantages, PNN is seen as a supervised neural network that can be applied in pattern recognition and system classification.[6]

Probabilistic Neural Network is a classification with the type of radial basis function (RBF) neural network. Similar to the Radial Basis Function neural network, in the hidden layer PNN activation functions are used to build a local decision function, where the data in the input layer is the focus. After processing the training data, these functions are added up in the summation layer. Then the probability value is obtained which is the total of the number of functions. Then the highest probability value will be classified into a certain class.[11]

Before being identified using PNN, there are several steps that must be done. Started by collecting image datasets and split into training and test sets. After collecting, the images is preprocessed using Green Channel, Contrast Limited Adaptive Histogram Equalization (CLAHE), Morphological Close and Background Exclusion, then segmentation with Otsu Thresholding, next step is post-processing with Connected Component Analyst (CCA) and feature extraction with Invariant Moment. Then it ends with a Probabilistic Neural Network (PNN) to identify the image as a malaria image or a normal image.

These steps begin with collecting normal blood image datasets and Plasmodium vivax parasite blood image datasets as training data images and testing data images. The dataset used in this study was obtained from the Faculty of Medicine, Department of Parasitology, University of Sumatera Utara. With a total of 173 images, consisting of 138 training data (80%) and 35 testing data (20%). The processed image is 500 x 500 pixels with the jpg extension. The Following diagram will show:

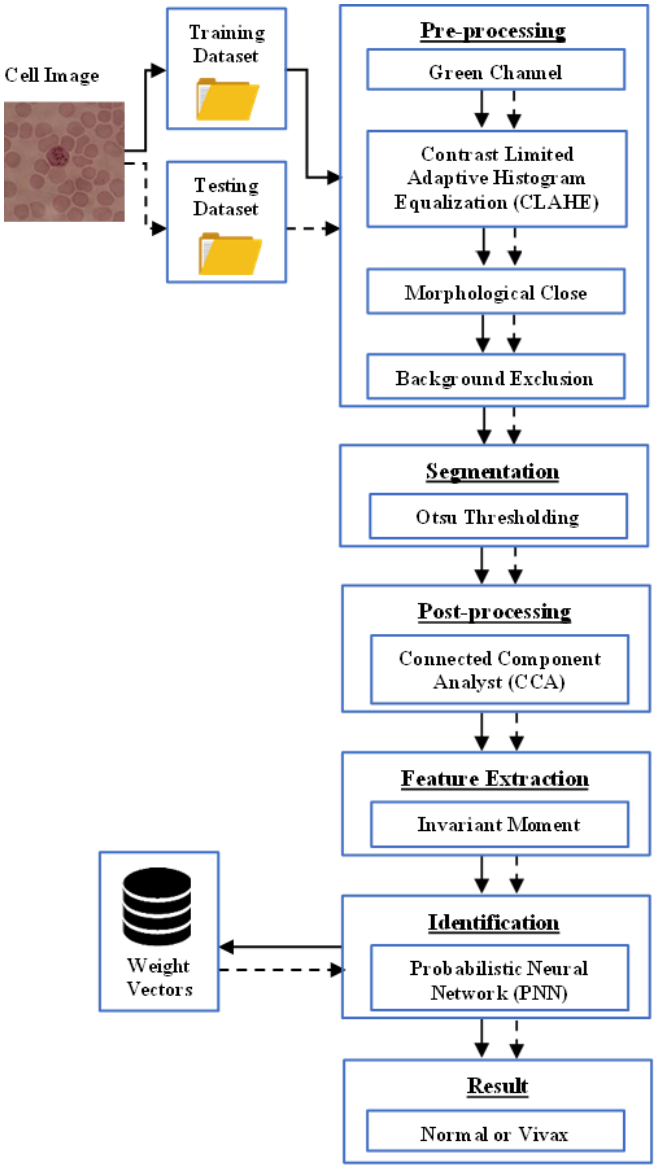


Figure. 1 The concept of Identification of Malaria Parasites Plasmodium Vivax on Red Blood Cells Using The Probabilistic Neural Network Method.

2.1 Preprocessing

2.1.1 Green Channel

The red blood cell image is an RGB image. The green channel was chosen because it produces images of the malaria parasite Plasmodium vivax and the structure of the parasite is more visible than the red or blue channels. To determine the value of the green channel, the following equation is used:

$$I(x, y) = 0.R + 01.G + 0.B \quad (1)$$

With:

I(x,y): green channel image pixels.
R : red value in each pixel.
G : green value in each pixel.
B : blue value in each pixel.

For example, the image size is changed to 6 x 6 pixels, then the red (R), green (G) and blue (B) values of each pixel are calculated. So the result is:

98	100	112	102	103	100
86	94	101	109	100	100
64	86	109	102	109	101
89	103	104	95	108	99
102	100	102	109	114	103
103	99	110	99	105	100

Figure. 2 Green channel process results.

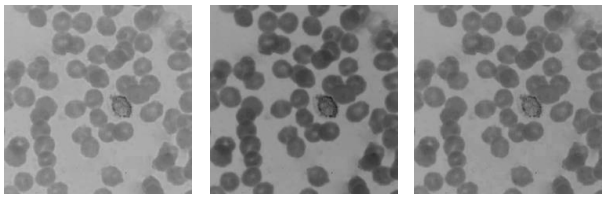


Figure. 3 Comparison between, Green Channel and Blue Channel.

2.1.2 Contrast Limited Adaptive Histogram Equalization (CLAHE)

The images used have various qualities, sometimes the parasitic form found in red blood cells is less clear or difficult to distinguish from non-parasitic blood cells. By increasing the contrast but not excessively showing the shape and structure of red blood cells more clearly so that it can simplify the next process. The CLAHE process uses the clip limit formula which has an independent variable called the clip factor, the clip factor is the limit entered by the user which used to set how high the brightness level is.

$$\beta = \frac{M}{n}\left(1 + \frac{\alpha}{100} S_{max} - 1\right) \qquad (2)$$

With:

β : histogram clip limit value
M : area size
n : grayscale value
α : clip factor (value 1 – 100)

The previous green channel image will be applied to the clip factor formula (2). So the result is:

86	103	155	101	118	50
52	69	121	151	50	50
17	52	134	101	151	67
17	119	136	18	123	53
85	51	85	140	158	88
119	34	152	53	105	70

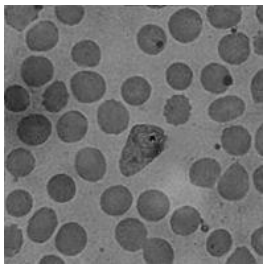
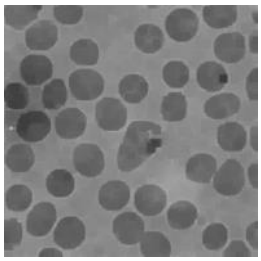


Figure. 4 CLAHE process results

2.1.3 Morphology Close

The image in the morphological close process disguises small holes in the object then adjacent objects will be combined so as to create an effect that is able to smooth out the object's boundaries without changing the object's area. [12] The process of dilation (thickening of pixels) and erosion (thinning of pixels) is applied at this stage.



Figureu. 5 Morphology close process results.

2.1.4 Background Exclusion

Background exclusion applies a subtract operation between the image from the CLAHE process and the image from the morphological close process, so that the malaria parasite Plasmodium vivax is separated from the background.

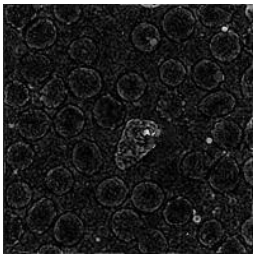


Figure. 6 Background exclusion Process Results.

2.2 Segmentation

The segmentation stage using Otsu Thresholding used to change the grayscale image to a binary image, where the foreground is separated from the background more maximally, so that the parasites are more clearly visible at this stage. The first step in calculating the Otsu Threshold value is to find the cumulative summation (3). Followed by calculating the Cumulative Mean (4), then calculating the Global Intensity (5). The next step is to calculate the value of the variance between class (6), then find the maximum value of the variance between class (7).

$$\omega_{(k)} = \sum_{i=0}^{k} i.p_i \quad (3)$$

$$\mu_{(k)} = \sum_{i=0}^{k} i.p_i \quad (4)$$

$$\mu_{(T)} = \sum_{i=0}^{L} i.p_i \quad (5)$$

$$\sigma_B^2(k) \frac{\left[\mu_{(T)}\omega_{(k)} - \mu_{(k)}\right]^2}{\omega_{(k)}\left[1 - \omega_{(k)}\right]} \quad (6)$$

$$\sigma_B^2(k^*) = max_{1 \le x \le L}\sigma_B^2(k) \quad (7)$$

With:

$\omega(k)$: Cumulative Summation Value
$\mu(k)$: Cumulative Mean Value
$\mu(T)$: Global Intensity Value
σ_B^2 : Threshold Value

The first step is to calculates the pixel probability by dividing the value of each pixel by the total pixels, then calculates the cumulative summation and the cumulative mean.

Table 1. Cumulative Summation Value and Cumulative Mean Value

i	p_i	$\omega_{(k)}$	$\mu_{(k)}$
17	0.055	0.055	0.935
18	0.027	0.082	1.421
…	…	…	…
88	0.027	0.519	28.576
101	0.027	0.546	31.303
…	…	…	…
158	0.027	0.981	88.052

After the results of the Cumulative Summation and Cumulative Mean calculation are obtained, the next step is to calculate the Global Intensity (5).

$$\frac{\begin{array}{c}(17 \text{ x } 2) + (18 \text{ x } 1) + (34 \text{ x } 1) + (50 \text{ x } 3) + (51 \text{ x } 1) + \\ (52 \text{ x } 2) + (53 \text{ x } 2) + (67 \text{ x } 1) + (69 \text{ x } 1) + (70 \text{ x } 1) + \\ (85 \text{ x } 2) + (86 \text{ x } 1) + (88 \text{ x } 1) + (101 \text{ x } 1) + (103 \text{ x } 2) \\ + (105 \text{ x } 1) + (118 \text{ x } 1) + (119 \text{ x } 2) + (121 \text{ x } 1) + \\ (123 \text{ x } 1) + (134 \text{ x } 1) + (136 \text{ x } 1) + (140 \text{ x } 1) + (151 \text{ x } \\ 2) + (152 \text{ x } 1) + (155 \text{ x } 1) + (158 \text{ x } 1)\end{array}}{36}$$

$$= 89.89$$

The next step is to calculate the value of variance between classes (6).

Table 2 Variance Between Classes

i	$\sigma_B^2\ (k)$
17	309,21
18	470,29
…	…
88	1.308,98
101	1.274,86
…	…
158	0,907

After obtaining the results of the variance between classes, the maximum value is determined. The highest value is used as the threshold (7). Where the maximum value obtained is 1.308.98 which is the value of 88 pixels. So, the threshold value obtained (k) is 88, so it can be applied to grayscale images with the following conditions:

$$g(x,y) = \begin{cases} 1 \; if \; f(x,y) \geq T \\ 0 \; if \; f(x,y) < T \end{cases}$$

Where g(x,y) is the binary image pixel and f(x,y) is the input image pixel. Where the part of the image that tends to be dark is made darker (black) and has a value of 0, the image that tends to be bright is made brighter (white) and has a value of 1. With a value of T = k = 88. So the result is:

0	1	1	1	1	0
0	0	1	1	0	0
0	0	1	1	1	0
0	1	1	0	1	0
0	0	0	1	1	1
1	0	1	0	1	0

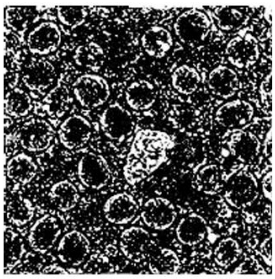


Figure 7. Otsu threshold

2.3 Post Processing

At the segmentation stage, many objects that are not parasites are still detected, therefore it is necessary to eliminate these objects. To eliminate non-parasites can be done by using the Connected Component Analysist. Connected Component Analysist algorithm can work on binary images by using 4 connectivity or 8-connectivity methods.

Figure 8. Connected Component Analysist

2.4 Feature Extraction

Feature extraction is the stage to find the feature value of an object that is unique, one of the methods is using Invariant. This method has seven values in each image object which is calculated using the central moments which are invariant to the image transformation. The first 6 moments are invariant to translation, scale and rotation, and reflection. While the 7th moment changes to image reflection [13]. If the image objects have been found, then the seven invariant moment values are calculated. The first step to get the Moment Invariant value is to determine the moment value. The moment value obtained is the final result of the Connected Component Analysis process measuring 500 x 500 pixels using the original image of red blood cells containing malaria.

The first step to determine the value of the moment invariant is to calculate the moment values m_{00}, m_{10}, and m_{01}.

$$m_{pq} = \sum_{x=0}^{H-1} \sum_{x=0}^{W-1} x^p \, y^q f(x,y) \qquad (8)$$

With:

m_{pq} : moment
W : width
H : height
y and x : coloumn and row
f (x,y) : image intensity value

So the result is:

m_{00} = 1414779.0
m_{10} = 3.68056469E8
m_{01} = 3.71037203E8

After the value of the moment is obtained, then the next step is to determine the value of the center moment (9). Previously calculated the value of ¯and ¯first. It is known that ¯is the central moment of x and ¯is the central moment of y, the value of m00 is the total number of pixels that make up the object, the values of m10 and m01 are the center of mass of the object.

$$\mu_{pq} = \sum_{x=0}^{H-1} \sum_{x=0}^{W-1} (x - \bar{x})^p \, (y - \bar{y})^q f(x,y) \qquad (9)$$

With:

$$\bar{x} = \frac{m_{10}}{m_{00}} = \frac{3.68056469E8}{141779.0} = 260.15121018901186$$

$$\bar{y} = \frac{m_{01}}{m_{00}} = \frac{3.71037203E8}{141779.0} = 262.25806504054697$$

So the result is:

μ_{11} = −2.1499869975960536E9
μ_{20} = 1.499435072275548E10
μ_{02} = 1.6110560238163187E10
μ_{30} = −5.940354027470562E11

μ_{03} = −3.5876500602702131E18
μ_{12} = −1.3321570542842102E11
μ_{21} = −3.3467707899063544E11

After the values of central moment is obtained, then the central moment normalized.

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}\gamma} \quad (10)$$

With:

$$\gamma = \frac{p+q}{2} + 1 \quad \text{and} \quad \mu_{00} = m_{00}$$

So the result is:

η_{11} = −0.0010741343959892314
η_{20} = 0.007491183841970235
η_{02} = 0.008048842578962817
η_{30} = −2.4951164321414634E − 4
η_{03} = −1506.914331495688
η_{12} = −5.5954357955211486E − 5
η_{21} = −1.40573823612027E

Then it ends by determining the value of invariant moments (ϕ), while the value obtained is very low, so the |log(|ϕ|) function is used to define the value, so that the difference between each value can be seen.

$$\phi 1 = \eta 20 + \eta 02$$
$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$
$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} + \eta_{03})^2$$
$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$
$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[\, 3(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})^2]$$
$$\phi_6 = (\eta_{20} - \eta_{02})\,[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}\,(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$
$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2](\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})\,[\, 3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (11)$$

The result of Invariant Moments consists of seven values is:

$\phi_1 = ln\ 0.0074 + 0.008$
$\phi_2 = ln\ (0.0074 - 0.008)^2 + 4 * (-0.001)^2$
$\phi_3 = ln\ ((-2.495) - 3 * (-5.595)) + (3 * (-1.405) + (-1506.9))^2$
$\phi_4 = ln\ (-2.495 + (-5.595))^2 + ((-1.405) + (-1506.9))^2$
$\phi_5 = ln\ ((-2.495) - 3 * (-5.595))\ ((-2.495) + (-5.595))\ [((-2.495) + (-5.595))^2 - 3 * ((-1.405) + (-1506.9))^2]$ $+3 * (-1.405) - (-1506.9) * ((-1.405) + (-1506.9))\ 3* ((-2.495) + -5.595)\ ((-1.405) + (-1506.9))^2$
$\phi_6 = ln\ (0.0074 - 0.008) * [((-2.495) + (-5.595)) - ((-1.405) - 1506.9)^2 + 4 * (0.001)\ ((-2.495) + (-5.595))$ $((-1.405) + (-1506.9))$
$\phi_7 = ln\ (3 * (-1.405) - (-1506.9))\ ((-2.495) + (-5.595))\ [((-2.495) + (-5.595)]^2 - 3 * ((-1.405) + (-1506.9))^2]$ $-((-2.495) - 3 * (-5.595)) * (-3 * (-5.595))\ ((-1.405) + (-1506.9))\ 3 * ((-2.495) + (-5.595))^2 -$ $((-1.405) + (-1506.9))^2$

So the result is:

ϕ_1 = 4.164336233855339
ϕ_2 = 10.883185793260857
ϕ_3 = 14.63563814020601
ϕ_4 = 14.635638886493116
ϕ_5 = 29.271277399842447

ϕ_6 = 7.143873959569183
ϕ_7 = 15.043747370569644

2.5 Identification

The final stage is identification using the Probabilistic Neural Network (PNN) method. This method has two processes, the training and testing process. In this case, the value of feature extraction becomes a reference in both processes. The value of invariant moments used in the training process will be stored in the database as a weight vector. While the value of invariant moments in the testing process is used as an input vector which is then entered into the pattern layer, where the Gaussian kernel function is applied to this layer. Then into the summation layer, in this layer the Gaussian kernel functions from the same class will be summed, and with the amount of training data from each class the average result will be searched. On that layer, the probability density function is used. In the output layer highest probability value will be classified into a certain class using Bayes's decision.

In the training process, the characteristic vector of feature extraction in each training data forms the weight vector of each neuron in the pattern layer. In this process there are 88 images of blood cells infected with Plasmodium vivax and 50 images of normal red blood cells with each training image having 7 weight vectors to be trained. Weight vector of each training image will be stored in the database. The Following diagram will show:

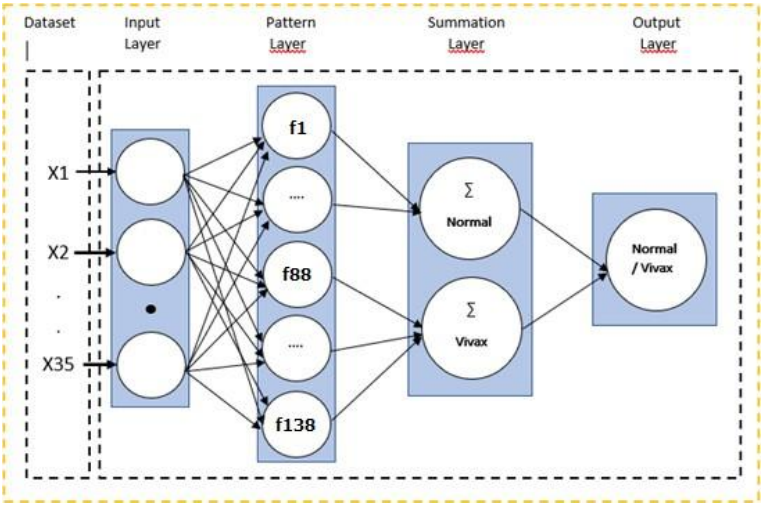


Figure. 9 The architechture of Probability Neural Network

The diagram above shows that the testing process is carried out in the following stages:

2.5.1 Input Layer

In this layer is found the input vector variable, its function as input in the network. In every data that has been tested has a feature extraction value, the value that is expected will be the value of the input vector variable.

ϕ_1 = 4.164336233855339
ϕ_2 = 10.883185793260857
ϕ_3 = 14.63563814020601
ϕ_4 = 14.635638886493116
ϕ_5 = 29.271277399842447
ϕ_6 = 7.143873959569183
ϕ_7 = 15.043747370569644

2.5.2 Pattern Layer

In this layer, the proximity of the distance between the weight vector and the input vector is calculated. The weight vector is the value of the training data for each class that is stored in the database, and the input vector is the value of feature extraction from the test data.

$$f_i(x) = exp\left[\frac{(x - x_{ij})^T - (x - x_{ij})}{2\sigma^2}\right] \quad (12)$$

With:

$f_i(x)$: proximity function
x : testing vector
x_{ij} : training vector j in class i
T : transpose
σ : smoothing

The trial and error techniques is used in determining the value of the smoothing parameter. So the result is:

$$f_i(1) = exp\left[\frac{\begin{array}{c}(4.164336233855339 - 5.903135036243803)^2 + \\ (10.883185793260857 - 14.277785161230357)^2 + \\ (14.63563814020601 - 5.334298846670677)^2 + \\ (14.635638886493116 - 5.334293262805888)^2 + \\ (29.271277399842447 - 10.66858931753487)^2 + \\ (7.143873959569183 - 2.2798198898044513)^2 + \\ (15.043747370569644 - 3.0819973471211446)^2\end{array}}{36}\right]$$

$f_i(1) = 8.040248765682937E\text{-}8$

By using equation (12), it is obtained the proximity value between the input vector and all weight vectors that have been stored in the database, with Normal class having 50 training data and Vivax class having 88 training data.

Table 3. The Proximity Value Between The Input Vector And All Weight Vectors

Class	$f_i(1)$		$f_i(2)$	.	$f_i(49)$	$f_i(50)$	.	$f_i(87)$	$f_i(88)$
Normal	8.04		2.94	.	4.60	5.04	-	-	-
Vivax	8.82		9.41	.	6.77	6.79	.	0.85	0.85

2.5.3 Summation Layer

In this layer the Gaussian kernel functions from the same class will be summed, and with the amount of training data from each class the average result will be searched. Summation Layer is used to find the probability of each class (13).

$$g_i(x) = \frac{1}{2\pi^{d/2}\,\sigma^d} \cdot \frac{1}{N_i} \sum_{i=1}^{N} exp\left[\frac{(x - x_{ij})^T - (x - x_{ij})}{2\sigma^2}\right] \quad (13)$$

With:

$gi(x)$: probability density function
d : vector dimension
N : number of training vectors in class i

As is known, the normal class has 50 training data and in the vivax class has 88 training data.

Normal Class:

$$g_i(1) = \frac{1}{2\pi^{d/2}\,\sigma^2} \cdot \frac{1}{50}\, 0.00504935$$

$$g_i(1) = 1.5052416639449424E - 7$$

Vivax Class:

$$g_i(1) = \frac{1}{2\pi^{d/2}\,\sigma^2} \cdot \frac{1}{88}\, 0.86853837$$

$$g_i(1) = 1.4543471072819024E-5$$

2.5.4 Output Layer

In this layer, the results of these classes will be compared with their probability values. The highest probability value is classified to a certain class by applying Bayes's decision rules. From the results above, it can be concluded that the highest probability value is in the vivax class with a value of 1.4543471072819024E-5, then the results shown are vivax.

Table 4. The Result of Output Layer

No.	Class	Summation Layer
1	Normal	1.5052416639449424E-7
2	*Vivax*	1.4543471072819024E-5

3. Result and Discussion

The methodology presented in this study makes an efficient identification of the image as a malaria image or a normal image. 15 normal images and 20 images of the malaria parasite plasmodium vivax will be tested using training data, 50 normal images and 88 images of the malaria parasite plasmodium vivax. The value of smoothing parameter (σ) used is different in this test, starting from (0.1), (0.3), (0.4), (0.5), (0.9).

In this test, the smoothing parameter (σ) used has different values in order to get the value in order to be able to identify the malaria parasite Plasmodium vivax with a high level of accuracy.

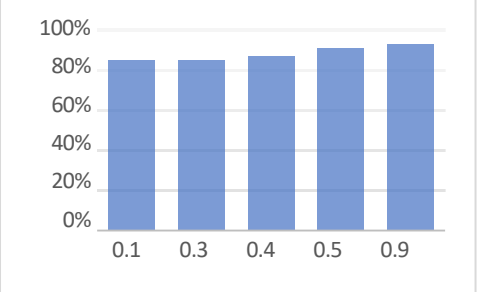
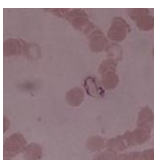
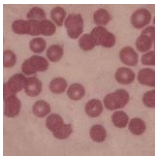
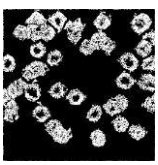
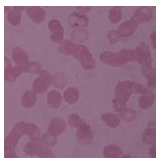
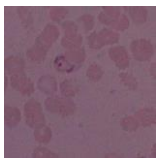
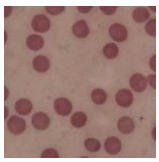
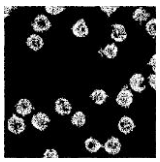
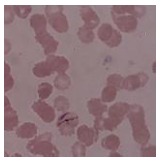
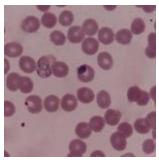
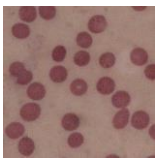
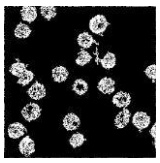
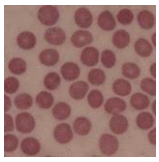
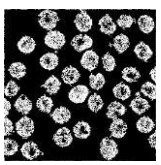
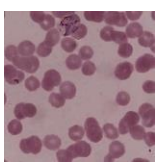


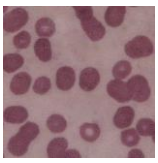
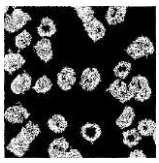
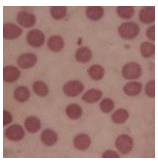
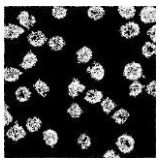
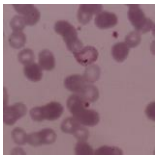
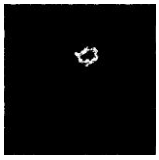
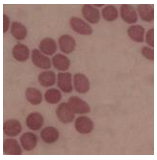
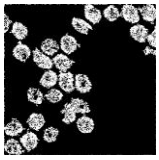
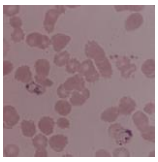
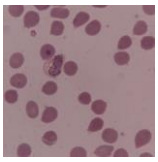
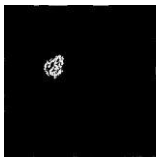
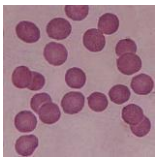
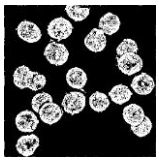
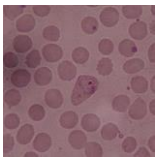
Figure 10. The results of smoothing values

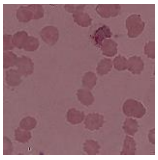
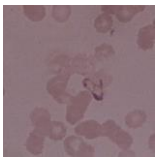
The conclusion drawn from the figure above is that the higher the value of σ, the higher the accuracy obtained. This is because the value of σ is enough to affect the value of the probability density function. On the other hand, the lower the value of σ, the lower the level of accuracy obtained. So that the best accuracy obtained is the value of σ ≥ 0.9.

Table 4. Test Results

No.	Original Image	Image Processing	Result	Status
1.			Normal: 1.574E-23 **Vivax: 2.092E-17**	TRUE
2.			**Normal**: **3.397E-18** Vivax: 1.762-E25	TRUE

No.	Original Image	Image Processing	Result	Status
3.			Normal: 1.599E-20 **Vivax: 1.798E-18**	TRUE
4.			Normal: 4.560E-23 **Vivax: 6.035E-18**	TRUE
5.			**Normal: 1.455E-17** Vivax: 4.449E-20	TRUE
6.			Normal: 1.292E-24 **Vivax**: **7.952E-17**	TRUE
7.			Normal: 5.076E-24 **Vivax**: **8.141E-17**	TRUE
8.			**Normal: 1.691E-18** Vivax: 6.217E-19	TRUE
9.			**Normal: 8.806E-16** Vivax: 1.413E-21	TRUE
10.			Normal: 1.211E-25 **Vivax**: **6.712E-17**	TRUE

No.	Original Image	Image Processing	Result	Status
11.			**Normal: 1.298E-17** Vivax: 2.218E-20	TRUE
12.			**Normal: 2.312E-17** Vivax: 1.910E-20	TRUE
13.			Normal: 7.280E-22 **Vivax: 3.493E-17**	TRUE
14.			**Normal: 8.328E-17** Vivax: 2.993E-18	TRUE
15.			Normal: 5.404E-23 **Vivax: 2.003E-17**	TRUE
16.			Normal: 7.986E-24 **Vivax: 6.432E-17**	TRUE
17.			**Normal: 5.940E-17** Vivax: 1.847E-72	TRUE
18.			Normal: 1.505E-20 **Vivax: 1.454E-17**	TRUE

No.	Original Image	Image Processing	Result	Status
19.			Normal: 6.861E-25 **Vivax: 7.106E-17**	TRUE
20.			**Normal: 8.943E-18** Vivax: 3.661E-18	FALSE

From the total number of system tests that were processed using 35 test data, there was 1 image that failed to identify the red blood cell image. Failure is caused by an unclear image. The blood cell image in the table above at number 20 should have malaria vivax output, but the output given by the system shows that image number 20 is normal. In finding the level of accuracy of system performance, it can be done by using:

$$Percentage\ Accuracy = \frac{TP + TN}{TP + FP + FN + TN}\ x\ 100\% \quad (14)$$

With:

TP : True Positive, indicating the number of cases of vivax malaria patients with positive test results.
TN : True Negative, stating the number of cases without illness with negative test results.
FP : False Positive, states the number of cases that are in fact not sick but have positive test results.
FN : False Negative, states the number of cases that actually suffer vivax malaria but had a negative test result.

Table 6. Matrix Confussion

	ACTUAL VALUE	
PREDICTION VALUE	Positive	Negative
Positive	TP = 19	FP = 0
Negative	FN = 1	TN = 15

$$Percentage\ Accuracy = \frac{19 + 15}{35}\ x\ 100\%$$

$$= \frac{34}{35}\ x\ 100\%$$

$$= \mathbf{97.14\%}$$

With these calculations, it is known that the accuracy of the system performance of the Probabilistic Neural Network method in identifying the malaria parasite Plasmodium vivax through microscopic images reaches 97.14%. In pattern recognition, recall, precision, f-score, and specificity are the formulas used to calculate the work of the system being built. The success rate of the system to retrieve information is called recall. The level of accuracy in the information desired by the user with the results predicted by the system is called precision. The combination of the average of the precision and recall values is called the F-score. Meanwhile, the system's ability to identify the data being tested is normal data or data that does not contain the vivax malaria parasite is called specificity.

Table 7. Percentage Of Recall, Precision, Specificity, F Score

Recall / Sensitivity $\frac{TP}{TP + FN}100\%$	**Precision** $\frac{TP}{TP + FP}100\%$	**Specificity** $\frac{TN}{TN + FP}100\%$	**F-Score** $\frac{Precision + Recall}{2}$
95%	100%	100%	97.5%

4. Conclusion

Based on the test results on the Plasmodium vivax malaria parasite identification system using the Probabilistic Neural Network method, the following conclusions can be drawn: The Probabilistic Neural Network method is able to identify the malaria parasite Plasmodium vivax on microscopic red blood cell images. Based on 35 tested image data, 20 images were identified as malaria vivax and 15 images were identified as normal, with training data 138 images resulting in an accuracy rate of 97.14%, sensitivity 95% and specificity 100%.

From the results of system testing, the value of the smoothing parameter (σ) used affects the level of accuracy. After going through several tests with different neuron values, it can be concluded that the lower the value of σ, the lower the probability value obtained and the higher the value of σ, the higher the probability value obtained. The best smoothing parameter value to identify the malaria parasite Plasmodium vivax using the Probabilistic Neural Network is $\sigma \geq 0.9$. In the future, it is hoped that research will be expanded to identify various types of malaria parasites more accurately.

References

[1] Hiswani. (2004). Gambaran Penyakit Dan Vektor Malaria Di Indonesia. Fakultas Kesehatan Masyarakat Universitas Sumatra Utara.

[2] Fitriany, J., Sabiq, A. (2018). Malaria. *Jurnal Averrous* Vol 4(2).

[3] Kemenkes RI. (2020). Pusat Data Dan Informasi Penyakit Malaria 2020 (pp. 1–7).

[4] Seman, N. A., Ashidi, N., Isa, M., Li, L. C., & Mohamed, Z. (2008). Classification Of Malaria Parasite Species Based On Thin Blood Smears Using Multilayer Perceptron Network. *International Journal of the Computer, the Internet and Management*, 16(1), pp 46–52.

[5] Ghosh, M., Das, Devkumar., Chakraborty, C., Ray, A.K. (2011). *Plasmodium vivax* segmentation using modified divergence. International Confrence on Image Information Processing (ICIIP 2011).

[6] Shahana,S, Das, D.C.N. 2016. Probabilistic Neural Network Assisted Cell Tracking and Classification, International Research Journal of Engineering and Technology (IRJET) 3(8):1155-1161.

[7] Yunda, L., Alarcón, A. & Millán, J. (2011). Automated Image Analysis Method for p-vivax Malaria Parasite Detection in Thick Film Blood Images. Revista S&T, 10(20), 9-25

[8] Kristofer, E. P. D., Rivera, P. T., & Naval, P. C. (2017). Malaria Parasite Detection and Species Identification on Thin Blood Smears Using a Convolutional Neural Network. 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). doi:10.1109/chase.2017.51

[9] Gummadadila, A., Deepthi, A., Sowmya, B., Bai, A. K., & Reddy, P. A. (2020). Classification of White Blood Cell Images Using Probabilistic Neural Networks. 3(10), 167–172.

[10] Saputra, Dian. (2018). Probabilistic Neural Network dan Invarinat Moment Untuk Klasifikasi Bentuk Wajah Pada Pria. Skripsi. Universitas Sumatera Utara

[11] Lotfi, A., Benyettou, A. 2014. A reduced probabilistic neural network for the classi_cation of large databases. Turkish Journal of Electrical Engineering & Computer Sciences 22 : 979-989.

[12] Kaur, R., Kaur, S. 2013. Object Extraction and Boundary Tracing Algorithms for Digital Image Processing: Comparative Analysis: A Review. International Journal of Advanced Research in Computer Science and Sofware Engineering 3(5):263-268.

[13] Chairani, R. 2016. Identifikasi Kesuburan Pria Melalui Kelainan Sperma Berdasarkan Morfologi (Teratospermia) Menggunakan Invariant Moment. Skripsi. Universitas Sumatera Utara.