

## Article

# Discriminant Analysis PCA-LDA Assisted Surface-Enhanced Raman Spectroscopy for Direct Identification of Malaria-Infected Red Blood Cells

Gunganist Kongklad <sup>1</sup>, Ratchapak Chitaree <sup>1,\*</sup> , Tana Taechalertpaisarn <sup>2</sup>, Nathinee Panvisavas <sup>3</sup> and Noppadon Nuntawong <sup>4</sup>

<sup>1</sup> Department of Physics, Faculty of Science, Mahidol University, Bangkok 10400, Thailand; gunganist.kog@student.mahidol.ac.th

<sup>2</sup> Department of Microbiology, Faculty of Science, Mahidol University, Bangkok 10400, Thailand; tana.tac@mahidol.ac.th

<sup>3</sup> Department of Plant, Faculty of Science, Mahidol University, Bangkok 10400, Thailand; nathinee.pan@mahidol.ac.th

<sup>4</sup> National Electronics and Computer Technology Center (NECTEC), 112 Thailand Science Park, Pathum Thani 12120, Thailand; noppadon.nuntawong@nectec.or.th

\* Correspondence: rachapak.chi@mahidol.ac.th

**Abstract:** Various methods for detecting malaria have been developed in recent years, each with its own set of advantages. These methods include microscopic, antigen-based, and molecular-based analysis of blood samples. This study aimed to develop a new, alternative procedure for clinical use by using a large data set of surface-enhanced Raman spectra to distinguish normal and infected red blood cells. PCA-LDA algorithms were used to produce models for separating *P. falciparum* (3D7)-infected red blood cells and normal red blood cells based on their Raman spectra. Both average normalized spectra and spectral imaging were considered. However, these initial spectra could hardly differentiate normal cells from the infected cells. Then, discrimination analysis was applied to assist in the classification and visualization of the different spectral data sets. The results showed a clear separation in the PCA-LDA coordinate. A blind test was also carried out to evaluate the efficiency of the PCA-LDA separation model and achieved a prediction accuracy of up to 80%. Considering that the PCA-LDA separation accuracy will improve when a larger set of training data is incorporated into the existing database, the proposed method could be highly effective for the identification of malaria-infected red blood cells.

**Keywords:** malaria-infected red blood cells; *P. falciparum* (3D7); surfaced-enhanced Raman spectra; PCA-LDA



**Citation:** Kongklad, G.; Chitaree, R.; Taechalertpaisarn, T.; Panvisavas, N.; Nuntawong, N. Discriminant Analysis PCA-LDA Assisted Surface-Enhanced Raman Spectroscopy for Direct Identification of Malaria-Infected Red Blood Cells. *Methods Protoc.* **2022**, *5*, 49. <https://doi.org/10.3390/mps5030049>

Academic Editor: Leonardo Sacconi

Received: 2 May 2022

Accepted: 7 June 2022

Published: 10 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Malaria is a disease that is transmitted by mosquitoes (acting as a vector) and it is caused by a parasite of the *Plasmodium* genus. *Plasmodium falciparum* is usually the source of malaria mortality [1]. Between 2000 and 2020, the World Health Organization estimated that more than 200 million cases of malaria infection were reported each year [2]. The gold standard techniques for malaria diagnosis are rapid diagnostic tests (RDTs) and microscopic analysis of blood-smears [3,4]. The blood smear test is highly sensitive and precise for malaria diagnosis, but it needs to be performed by trained and qualified staff. RDTs are rapid; however, their sensitivity is not comparable to blood smear examination. Another highly sensitive diagnostic approach is polymerase chain reaction (PCR) but it involves a lengthy sample preparation process [5,6]. Various methods have been developed over the last two decades to satisfy the demand for simple and highly sensitive diagnostic techniques.

There are a number of reports on the application of the Raman spectroscopy technique in malaria diagnosis. The majority of these studies have focused on detecting hemozoin, or malaria pigment [7]. After the parasite infects an erythrocyte, the host's hemoglobin is catabolized. A by-product is created when the free heme is converted into a crystalline structure known as "hemozoin". Plasma or red blood cell extract is often used as the sample. Wood et al. first recorded the expression of the hemozoin Raman signal in the *Plasmodium* parasite food vacuole [8], and subsequent investigations of the excitation wavelength led to the enhancement of the Raman signal by using the resonance wavelength activated by hemoglobin and hemozoin [9–12]. In addition to the resonance Raman spectroscopy, tip-enhanced Raman scatter and magnetic fields have been used to increase the Raman intensity for hemozoin detection [13,14]. The level of hemozoin in the blood clearly corresponds to the time of infection. A small volume of hemozoin makes the examination difficult, especially in the early stages of infection [15]. Due to the size and structure of the metallic nanoscale substrate, the surface-enhanced Raman substrate was found to significantly enhance the Raman intensity [16,17]. With this technique, hemozoin in 0.0005% infected red blood cells could be detected by using a gold-coated butterfly wing nanostructure [18]. Chen et al. used a silver nanorod structure as the surface-enhanced Raman substrate to measure the modification of the red blood cell membrane and the expression of the protein component, especially the cytoadherence protein complex in the erythrocyte infection phase [19]. This was the first time that the direct measurement of the malaria-infected red blood cell was accomplished. Thereafter, there was no further work on the direct measurement of malaria-infected red blood cells. However, the Raman spectroscopy approach for malaria detection is still being developed. In 2020, gold substrate nanoparticles were employed for antibody-free malaria detection in whole blood [20]. Then, coupling Raman spectroscopy with mid- and near-infrared spectroscopy for malaria and arboviruses in human blood [21] was reported in 2021.

Apart from Raman spectrum investigation, the discrimination analysis technique is often used as an analytical tool in the Raman spectroscopy research area, particularly in regard to the issue of classification. The Raman spectrum is considered to be multivariate data and multivariate discrimination analysis (e.g., PCA and PC-LDA [21–27]) has been applied to differentiate a variety of cases such as the state of the red blood cell infection [19], the normal and the malaria-infected spleen tissue [28], and dengue and malaria sera Raman spectra [29].

In this work, we report for the first time that the direct identification of the malaria-infected red blood cells can be accomplished by the combination of SERS and principal component analysis coupled with linear discrimination analysis (PCA-LDA). The surface-enhanced Raman spectra were obtained from direct measurements of the normal and 3D7 *P. falciparum*-infected red blood cells deposited on a SERS substrate. A silver nanorod structure was used as the SERS substrate. The characteristic spectra and Raman spectral images of normal and infected red blood cells were presented. Subsequently, owing to the large number of spectra collected, the data set of spectra was used to train the separation model using a PCA-LDA-based machine learning algorithm. The proposed method can achieve a separation accuracy of more than 90%. Finally, the volunteers' blind data sets were used to evaluate the model and provided a prediction accuracy of up to 80%. Importantly, the accuracy of the separation model could be continually extended due to the updateable Raman spectra data set and could have the potential to be developed for the clinical application of malaria diagnosis in the future.

## 2. Material

### 2.1. Blood

Human red blood cells from healthy volunteers prepared at the National Center for Genetic Engineering and Biotechnology (BIOTEC), Thailand were used in this study.

## 2.2. Surface-Enhanced Raman Substrate

The OnSpec chip used in this study is a SERS based on a silver nanorod structure. All substrates were prepared by NECTEC, Thailand with a laboratory-made DC magnetron sputtering system, with a glancing-angle deposition (GLAD) technique. The details of the SERS preparation and fabrication are described in [30,31].

## 2.3. Raman Spectrometer

The Renishaw inVia™ confocal Raman microscope with an He-Cd laser was used. The light source provided an excitation wavelength of 532 nm. The microscope system was a Leica microsystem, which can be equipped with magnification objective lenses of 5x, 20x, 50x and 100x. The CCD camera was fitted with up to four detectors such as electron multiplied (EM) and InGaAs arrays. The Renishaw inVia™ confocal Raman microscope captured the spectra with a spectral resolution of  $0.3 \text{ cm}^{-1}$  (FWHM).

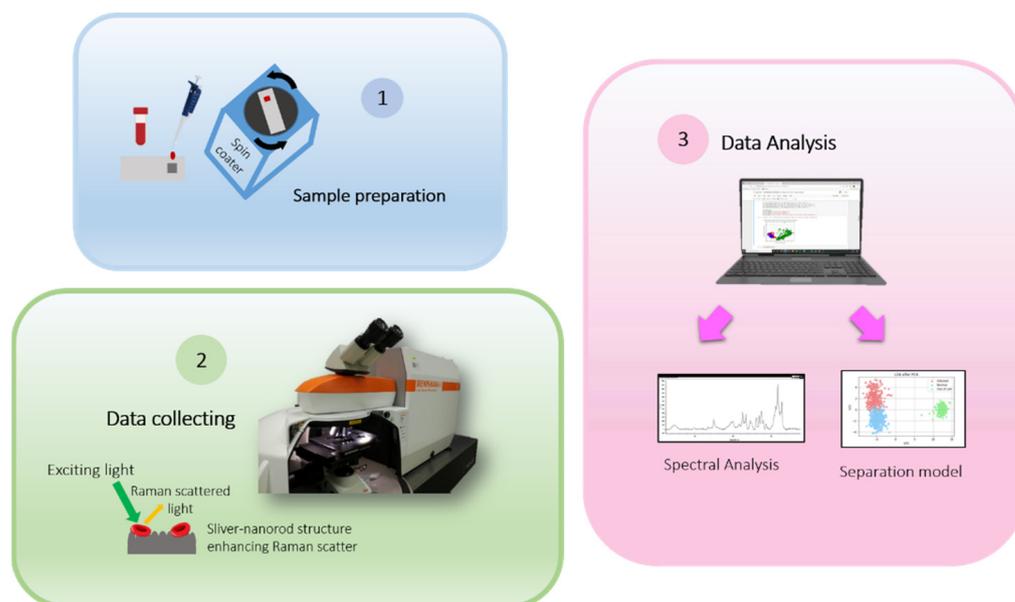
## 2.4. Software

WiRE4.2 software was used for spectral processing including cosmic ray removal, baseline subtraction, smoothing signal and normalization.

The imaging processes and separation model were simulated by Python. The module Hyperspy 1.6.5 was applied to intensify the imaging of the spectra in the area of interest and a collection of machine-learning modules, including scikit-learn 0.23.2 were applied for PCA-LDA discrimination processing.

## 3. Methodology

In this work, there are 3 main parts (see Figure 1). The first part involved the blood sample preparation and then the Raman spectra were collected. Finally, the data were interpreted and used for training the separation model.



**Figure 1.** Diagram of the Methodology.

### 3.1. Sample Preparation

#### 3.1.1. Malaria-Infected Cell Sample

*P. falciparum* (3D7) parasites were cultured according to the routine method [32] in RPMI1640 medium supplemented with 1% Albumax I at 4% hematocrit. The parasite culture was maintained at 37 °C with 5% carbon dioxide. The age of the parasite was around 24–32 h post invasion (this corresponds to the mid–late trophozoite stage to the

early schizont stage of the asexual cycle). The Percoll density gradients method was used to enrich the infected red blood cells. The percentage of parasite-containing red blood cells was confirmed by Giemsa staining. Red blood cells were resuspended in phosphate buffered saline (PBS) at 0.16% hematocrit. Then, an aliquot of 25  $\mu\text{L}$  of the cell suspension was applied to the OnSpec chip. The sample on the surface of the OnSpec chip was flattened by the spin coater.

### 3.1.2. Normal Red Blood Cell Sample

The normal red blood cell sample was prepared with human red blood cells resuspended in PBS. Then, 25  $\mu\text{L}$  of the cell suspension was applied to the OnSpec chip in the same way as the infected sample.

## 3.2. Spectral Collection

### 3.2.1. Measurement Conditions

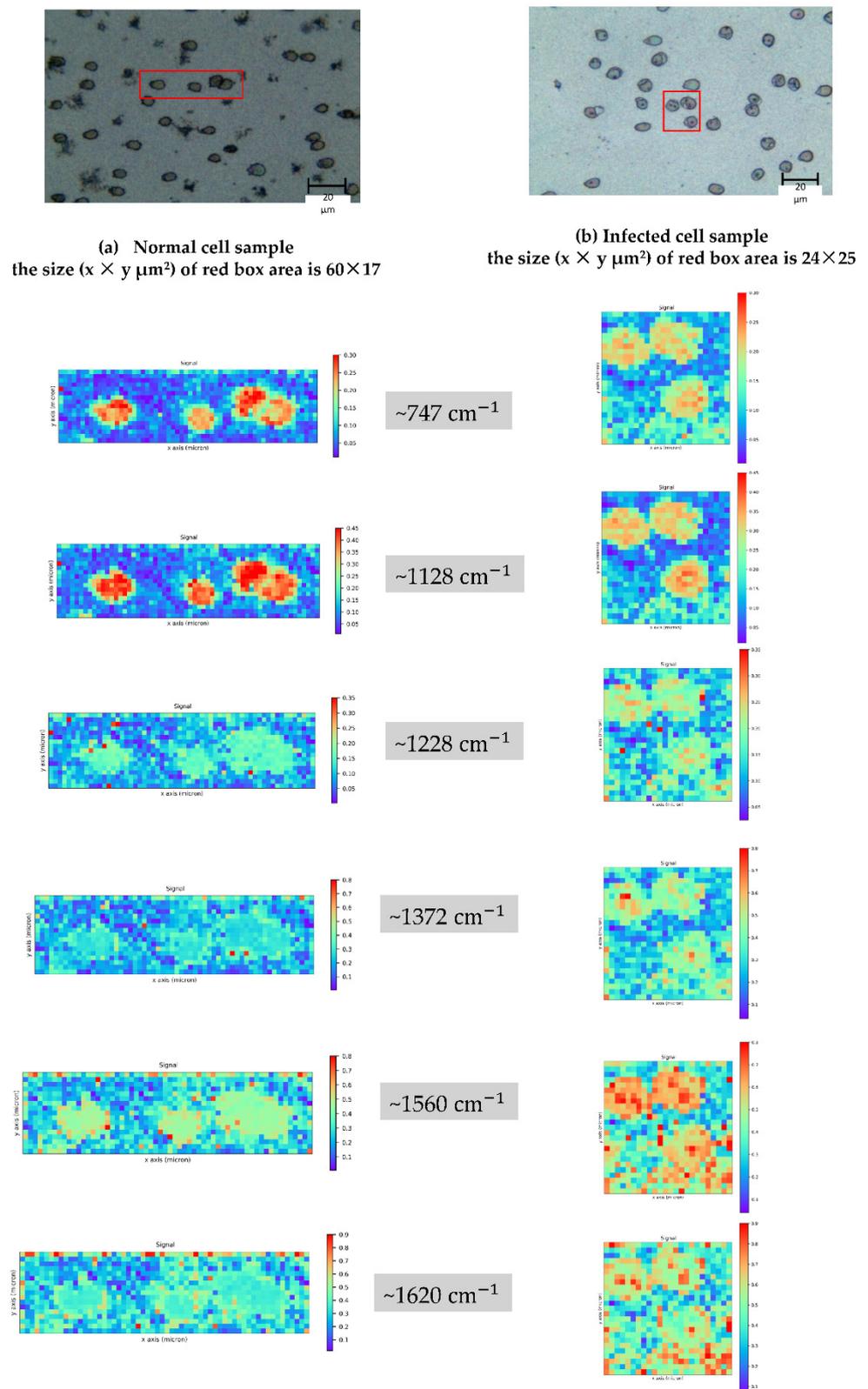
All spectra were acquired in the range of 123–1883  $\text{cm}^{-1}$  with 50x objective lens, and the laser power was set to be 1% (around 3 mW) of the maximum power (the low power of the laser was used to avoid damage to the sample). Each spectrum was presented as the average of three accumulation spectra with a 30 s exposure time.

### 3.2.2. Pointing Spectral Collection

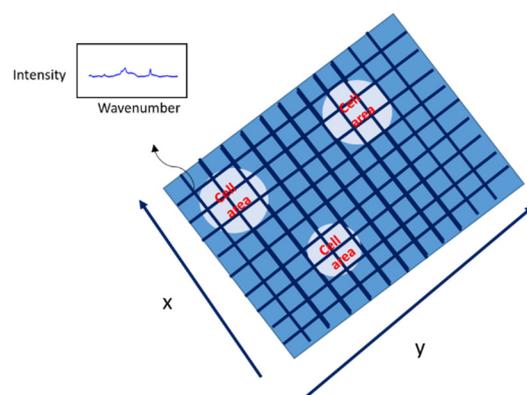
Point spectral collection was applied to select the point in the substrate area. The laser pointed to the pre-selected point. Normal and infected red blood cell spectra were collected from the points in the cell area, while outside the cell, spectra were collected from the points outside the cell area. There were 460 normal red blood cell spectra acquired from 6 normal red blood cell sample sets, 365 infected red blood cell spectra acquired from 4 infected red blood cell sample sets and 350 outside-the-cell spectra acquired from 10 total sample sets. Moreover, blank sample tested only on PBS was dropped on the OnSpec chip. All spectra met the measurement conditions described in Section 3.2.1.

### 3.2.3. Area Spectral Collection

The area spectral collection is the spectral collecting process done by area selection. In the area of interest (the red boxes in Figure 2), the intersection of the grid in x and y (see Figure 3) was the excitement point. The Raman spectrum from each position is the 2D data that contains the information about the intensity and wavenumber. The spectrum from each position was acquired using the same measurement conditions described in Section 3.2.1.



**Figure 2.** Image shows white light images of normal (a) and infected red blood cells. (b) The red boxes indicate the mapped area with the normalized intensities covering the peaks at 747, 1128, 1228, 1372, 1560 and 1620  $\text{cm}^{-1}$ . These maps visually compare the healthy red blood cells and the *P. falciparum* (3D7) infected red blood cells. The increase or decrease in the intensity of the Raman spectrum can be linked with the biological component’s modification in the red blood cell. The information regarding peak assignment is shown in Table 1.



**Figure 3.** Each point at the intersection in the area provided a Raman spectrum. The Raman spectra and spatial data were used for intensity image simulation.

### 3.3. Data Analysis

#### 3.3.1. Spectral Processing

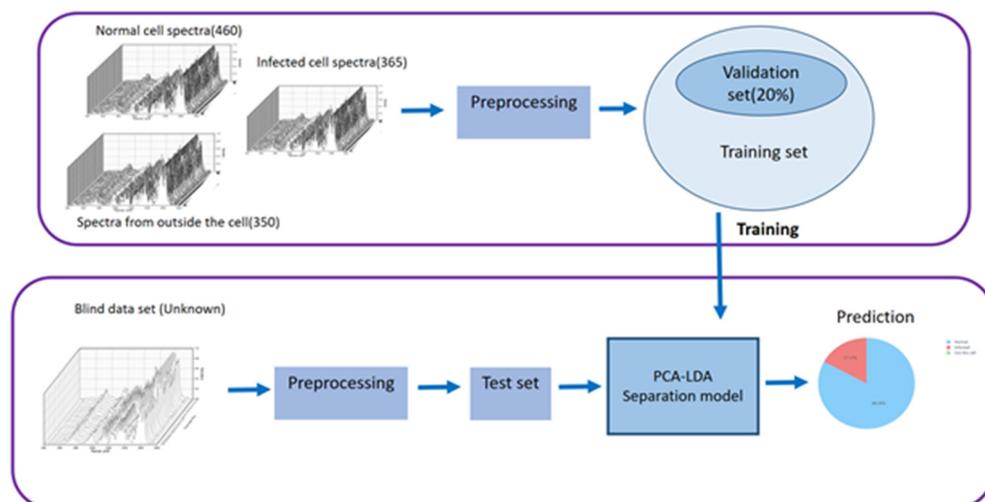
In the WiRE 4.2 program, there are built-in tool for spectral processing. The polynomial order was defined and adjusted automatically for fitting the baseline of the spectrum. The Savitsky–Golay filter [33] was used in the smoothing process. The cosmic ray spikes were identified by the threshold value of the width and height of the intensity peak. Then, the mean value of the intensities around the spike was used as the substitute intensity at the spike position [34,35]. Finally, the spectra were only truncated in the range covering  $450\text{--}1750\text{ cm}^{-1}$ .

#### 3.3.2. PCA-LDA Separation Model

- *PCA-LDA Discrimination Analysis*

Principal component analysis (PCA) is an unsupervised technique that is frequently used in conjunction with Raman spectral analysis [36]. The Raman spectrum is a representation of multidimensional data. PCA is always used to reduce the dimensions of the data and visualize the spectrum by transforming the data set into a new coordinate frame. Apart from dimension reduction, PCA can also be used to extract data features. While reducing the dimension of the data, the information related to the data is still retained as much as possible. Principal components (PCs) were used as new orthogonal coordinates. The first PC (PC1) is responsible for the data's largest variance axis, followed by PC2, and so on.

After spectral processing, 1175 Raman spectra were analyzed using the PCA algorithm: 460 from normal cells, 365 from infected cells, and 350 from outside the cell area. The 350 PCs transformed data that account for >99.9 percent of the total variance was used as the LDA algorithm's input data. LDA is the supervised analysis [37]. Each transformed spectrum from the PCA was labeled according to its cell type: normal, infected, or outside the cell. Figure 4 shows the flow chart of the discrimination analysis. All spectra acquired from the experiment were separated into a training data set (80%) and testing data set (20%). The training spectra were transformed by the PCA and LDA procedures. From the original data, each spectrum containing the information about intensities and wavenumbers (Raman shift) was transformed into data dots in 2D dimensional coordinates of LDA in which each dot represents the information about the spectrum. The PCA-LDA separation model was created by grouping the transformed data in 2D LDA coordinates according to their spectrum class. Then, the testing spectra was transformed in the same way from the original spectra to the data dots in the PCA-LDA separation model for the evaluation. Finally, all spectra classified as the training spectra and the blind spectra were applied as the testing spectra. The pie chart is the output of the blind test, which shows the probabilities of the spectra class.



**Figure 4.** Flow chart of the discrimination analysis.

- *Blind Test*

Six blind samples were prepared according to the method described in Sections 3.1.1 and 3.1.2. Three of these were normal cell samples and the others were infected cell samples. Six samples were randomly distributed for testing by three operators. The operators must maintain the spectral collection parameter in accordance with the manual. The manual defines the method for obtaining the spectra and the measurement parameters such as the laser's power and exposure time (as defined in Section 3.2.1). Each sample contained 30–35 spectra randomly collected from the cells. Preprocessing (e.g., cosmic ray removal, baseline subtraction, normalization, and spectral range selection) was applied to the spectra from each blind spectrum. This process may result in a reduction in the number of usable spectra. The blind spectra of each sample were then transformed into PCA and LDA coordinates.

Besides, two more spectra sets were collected from outside the cell using the same protocols. Without a doubt, the spectra could not be considered as a blind sample due to the point spectral collection process. These two data sets were used exclusively to evaluate the PCA-LDA separation model. However, the outside-the-cell spectra database may be useful in future analysis when area spectral collection is used. Consequently, there were eight data sets in total for the PCA-LDA separation model evaluation.

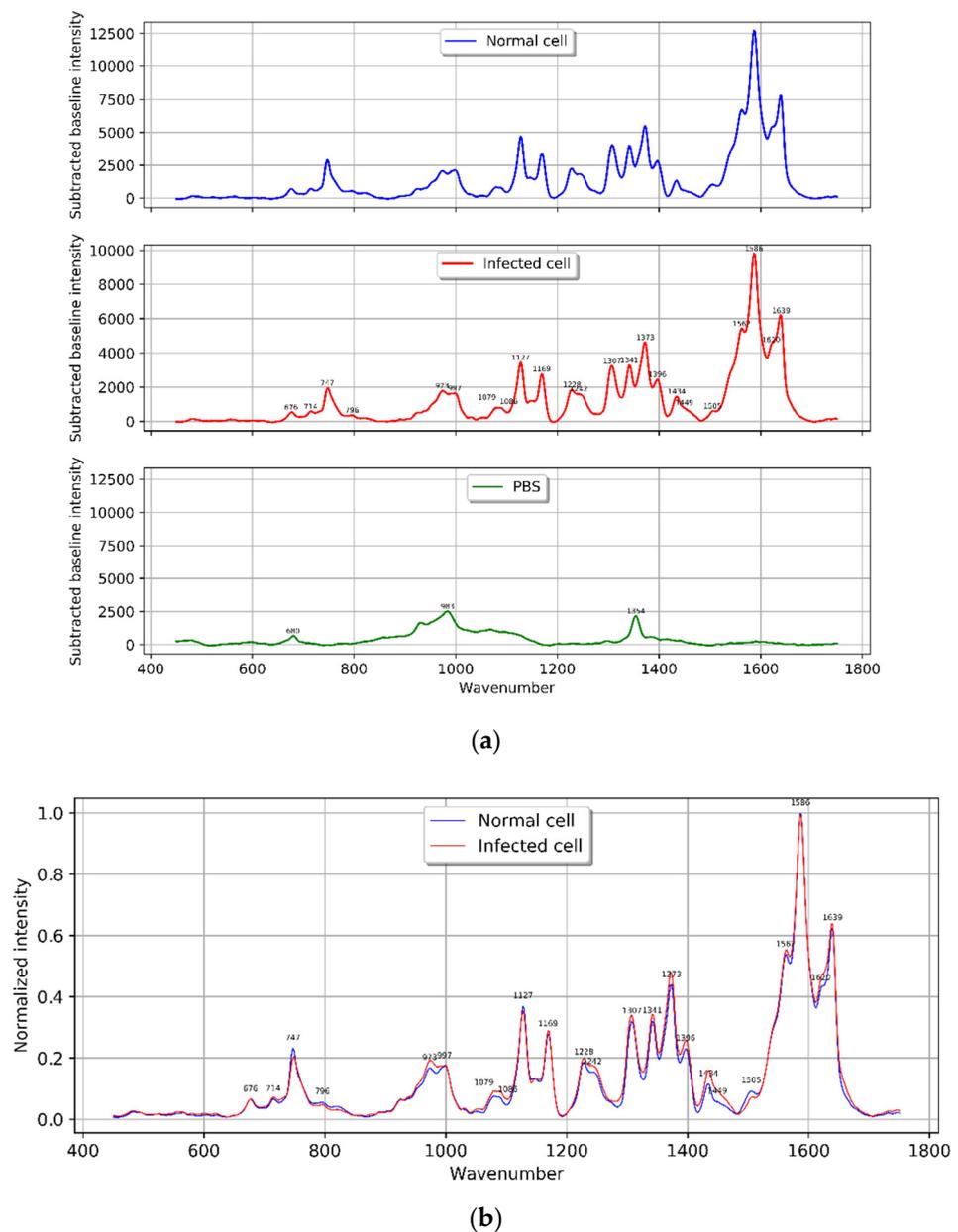
## 4. Results and Discussion

### 4.1. Characterization of Normal and Infected Red Blood Cell Raman Spectra

After the spectral processing, the average spectrum is shown by the blue line in Figure 5a, representing the average of 460 spectra of normal red blood cells, while 365 *P. falciparum* (3D7) infected red blood cell spectra were averaged and are represented by the red line in Figure 5a. Additionally, the spectrum for PBS used as the cell's media is displayed as the green line in Figure 5a. The Raman normalized intensity peak characteristics of normal and infected red blood cells are shown together in Figure 5b.

When normalized to the intensity at  $1586\text{ cm}^{-1}$ , the intensity signal of the Raman spectra from normal and infected red blood cells shows a slight difference, as seen in Figure 5b. This is due to the fact that the intensity at Raman shift  $1586\text{ cm}^{-1}$  is the highest peak in all the spectra acquired from the cell area.

For the intensity imaging, the peaks were focused at 747, 1128, 1228, 1372, 1560 and 1620. Therefore, the normalized intensities in the ranges of 747–749, 1124–1134, 1220–1237, 1361–1383, 1555–1565, and 1615–1629  $\text{cm}^{-1}$  (the wavenumber ranges that cover the peaks of interest) were averaged. The variation in the intensity is represented by the color of the image as shown in Figure 2.



**Figure 5.** Analysis of Raman spectrum intensity for normal and *P. falciparum* (3D7)-infected red blood cell. **(a)** Averaged normal red blood cells spectra ( $n = 460$ ) and infected red blood cells spectra ( $n = 365$ ) in comparison to PBS buffer. **(b)** Averaged, normalized spectra from normal red blood cells ( $n = 460$ ) and infected red blood cells spectra ( $n = 365$ ).

The disposal product, hemozoin is primarily visible between 24 and 32 h after infection, corresponding to the mid-late trophozoite to schizont stages. The hemozoin and the mechanism of its formation are of great interest in the development of anti-malarial drugs. Throughout the parasite's life cycle, hemoglobin in the red blood cell is catabolized, producing a large amount of free heme that is toxic to the parasite. Hemozoin is a byproduct of the parasite's free heme detoxification mechanism [38]. The components' peaks in the obtained Raman spectrum were assigned according to Table 1. Hemoglobin is the primary constituent of normal red blood cells and its presence is represented by the strong band in the Raman spectrum resulting from a 532 nm laser excitation wavelength to normal and infected red blood cells. However, hemoglobin and hemozoin both contain the heme prosthetic group, which provides the resonance enhancement in this wavelength range of excitation [11,28,39]. The Raman spectra of hemoglobin and hemozoin were comparable

with 532 excitation wavelengths [28]. Their distinctive peaks are in close proximity to one another as reported by Frame et al. [28]. The experiments yielded the corresponding results via the characteristic of the averaged normalized Raman spectrum and imaging (in Figures 3 and 5b). The obvious difference between the normal and infected cells is visible at Raman shifts  $\sim 1620, 1560, 1372, 1228, 1128,$  and  $747 \text{ cm}^{-1}$ . When red blood cells are infected, the hemoglobin was converted to the  $\text{Fe}^{3+}$  structure (hemozoin) [7,8,11,40]. The normalized Raman spectrum exhibits an increasing intensity trend of Raman shifts at  $\sim 1620, 1560, 1372$  and  $1228 \text{ cm}^{-1}$ , indicating the  $\text{Fe}^{3+}$  hemozoin characteristic band, while the Raman shift at  $\sim 747$  and  $\sim 1128 \text{ cm}^{-1}$ , which predominantly presents in the hemoglobin band exhibits the opposite trend.

**Table 1.** Raman peak assignment.

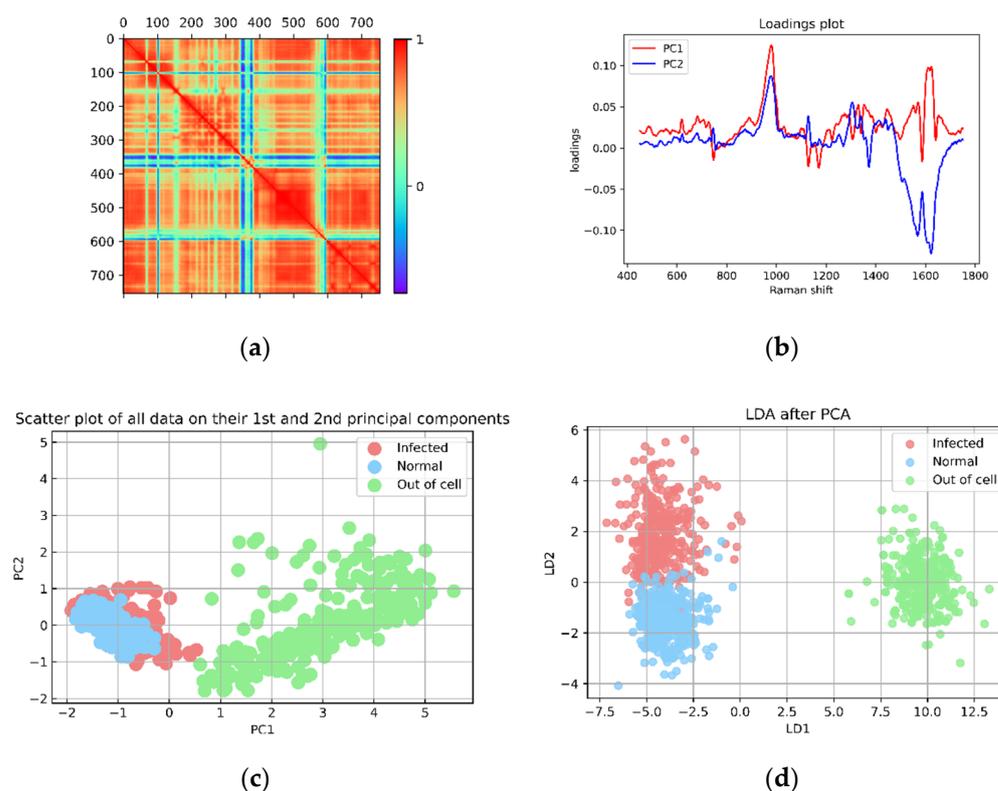
Observed Wavenumber ( $\text{cm}^{-1}$ )	Assignment
676	$\nu_7, \delta(\text{pyr deform})_{\text{sym}}$ of hemoglobin
747	$\nu_{15} \text{ Hb } [\nu(\text{pyr breathing}), \text{Trp}, \text{O-P-O sym Str. (lipid)}$
796	$\nu(\text{pyr breathing})\nu_6$
826	Porphyrin: $\gamma(C_mH)$ , Tyr
923	Amino acids: C-COO <sup>-</sup> stretch
950	C-C Str
973	$\nu_{46}, \delta(\text{pyr deform})_{\text{asym}}$
997	$\nu_{47}, \nu(C_bC_1)_{\text{asym}}, \text{Protein (Phe)}$
1078, 1079	$\delta(=C_bH_2)_4, \text{C-O Str.}$
1127, 1128	$\nu_5$
1171, 1172	$\nu_{30} \nu(\text{Pyr half-ring})_{\text{asym}}$
1225	$\delta(C_mH)$
1228	$\delta(C_mH) (\text{Oxy})$
1244, 1242	Amide III
1247	Amide III (collagen assignment)
1307	$\nu_{21}$
1340	$\nu_{44}(\text{Pyr half-ring})_{\text{sym}}$
1366, 1372	$\nu_4(\text{Pyr half-ring})_{\text{sym}}$
1396	$\nu_{20}$
1432, 1434	$\nu_{28}$
1440	C-H <sub>2</sub> and C-H <sub>3</sub> bend (protein, lipid)
1463	$\nu_3, \text{C-H}_2 \text{ and C-H}_3 \text{ bend}$
1540	$\nu_{11}$
1560, 1562	$\nu(c=c), \text{Trp}$
1586	$\nu_{37}$
1620	$\nu(c=c)$
1639, 1640	$\nu_{10}$

\* References for the assignment of Raman peaks: [11,19,28,38,39,41–49].

#### 4.2. Discrimination Analysis

Due to the slight differences between the spectra, it is difficult to distinguish the Raman spectra of normal cells from that of infected cells with the naked eye. Moreover, several hours must be spent on an area spectra collection for imaging. As a result, PCA and LDA techniques [50–52] were chosen for the discrimination analysis in this work.

The loadings plot of PC1 and PC2 are shown in Figure 6b, while the spectra in the range of  $450\text{--}1750 \text{ cm}^{-1}$  were reduced in dimension from 755 to 350 and transformed to a visual representation in 2D as shown in Figure 6c (PC1 and PC2). The scattering plot allows the separation of cell spectra (both normal and infected cell spectra) from non-cell spectra (spectra collected from outside the cell area). However, the normal and infected spectra was still misclassified. PC1 represented 69.1% of the variance, while PC2 represented 7.8%. There were 755 dimensions of data in total; nearly all significant features were extracted using 350PCs, and these 350PCs data were then used as the input for the LDA algorithm in a subsequent step.



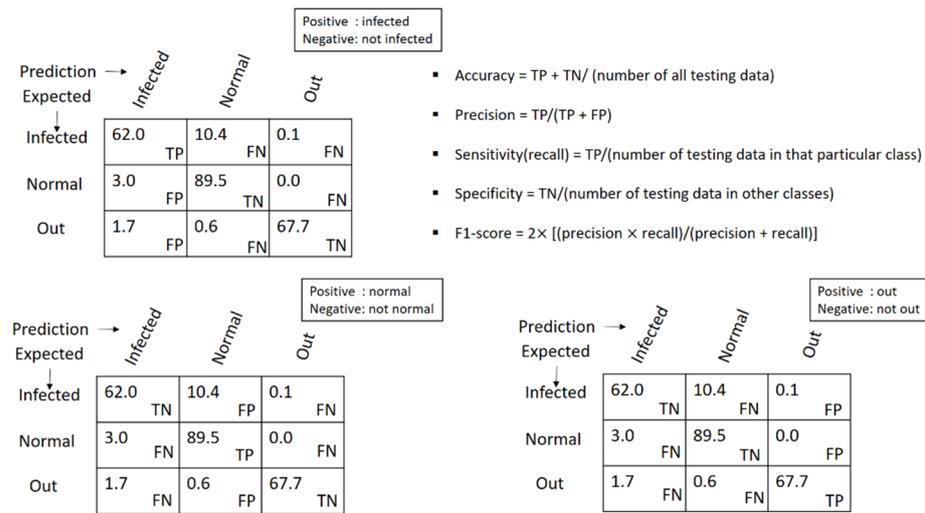
**Figure 6.** PCA-LDA results. (a) The covariance matrix of the PCA plot. (b) The loading plot of PC1 and PC2 after PCA processing of 1175 Raman spectra (460 normal cells, 365 infected cells and 350 outside-the-cell spectra) by using 755 Raman shifts in the range of 450–1750  $\text{cm}^{-1}$  is presented. (c) The PCA result shows the PC1 and PC2 coordinate and 350 PC results from PCA are continually input to LDA processing. (d) The LDA result plots in LD1 and LD2. In (c,d): blue dots are designated as the normal cell transformed spectra, red dots represent the infected cell transformed spectra and green dots are the transformed spectra obtained from the outside-cell area.

The LDA algorithm was used to process 1175 spectra. The discrimination of each type of data is demonstrated clearly in the LD1 and LD2 coordinates in Figure 6d. The pale sky-blue area represents a collection of normal cell spectra. The light-red cluster represents the infected cell, while the light-green area represents the non-cell spectra. The distinguishing area of each spectral class in Figure 6d was used as criteria for predicting the type of unknown data.

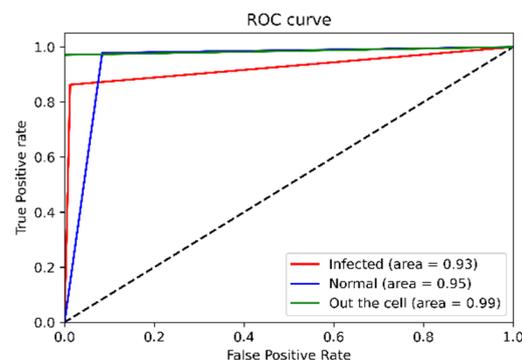
To illustrate the analytical parameters that indicated the PCA-LDA separation model's efficiency, 1175 spectra were split into an 80% training set and a 20% preliminary testing set for the model evaluation. In data splitting, ten random states (0–9) were selected. Thus, ten data sets were considered for the model evaluation, and the final result shown in Table 2 was calculated using the average value of the confusion matrix. The values in the confusion matrix were used to calculate the accuracy, precision, sensitivity (recall), f1-score, and specificity terms for the PCA-LDA separation model, as shown in Figure 7. Additionally, the ROC curve provided an overview of the efficiency of the PCA-LDA separation. If the area under the receiver operating characteristic curve (ROC) is close to a value of one the classification is considered to be well-performed. Due to the fact that this study contained data from three distinct classes, the data was binarized prior to the creation of the ROC curve. The ROC curve in Figure 8 illustrates that the PCA-LDA model performed well in terms of discrimination, with an area under the curve of more than 0.9 for all classes' analysis.

**Table 2.** List of parameters presenting how well the infected cells were distinguished from normal cells and non-cell area spectra (evaluated by 20% split data from training data set).

	Precision	Sensitivity (Recall)	F1-Score	Specificity	Support
Infected	0.97	0.86	0.91	0.98	73
Normal	0.88	0.98	0.93	0.92	92
Out the cell	1.00	0.97	0.99	0.93	70
Accuracy		0.94			235
Macro avg	0.95	0.94	0.94		235
Weight avg	0.94	0.94	0.94		235



**Figure 7.** Confusion matrix and each parameter calculation.



**Figure 8.** ROC curve presenting the efficiency of the model evaluated by 20% split data from training data set.

**4.3. Blind Test**

An example of the blind results is depicted in Figure 9 and is represented by the yellow dots. At the conclusion of the procedure, the pie chart was used to represent the data class prediction. The position of unknown data in the PCA-LDA separation was converted to the probability of data class prediction, which could be easily visualized using a pie chart. The area of the pie chart indicates the sample class’s probability (see Figure 10).

After predicting the type of blind samples using the PCA-LDA separation model generated from 1175 training spectra, the unknown data class prediction results were displayed in the pie chart (see the example in Figure 10, all pie charts of the blind data sets can be seen in the Supplementary Materials). The summary of all sample predictions and their expected outcome are shown in Table 3. All blind sample classes could be accurately predicted. Each prediction class corresponded to an expected class.

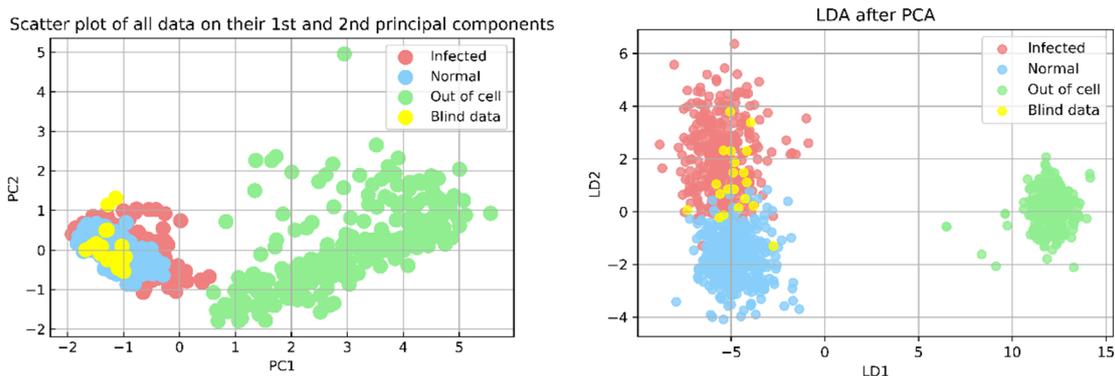


Figure 9. Transformation of blind data to PCA-LDA coordinates.

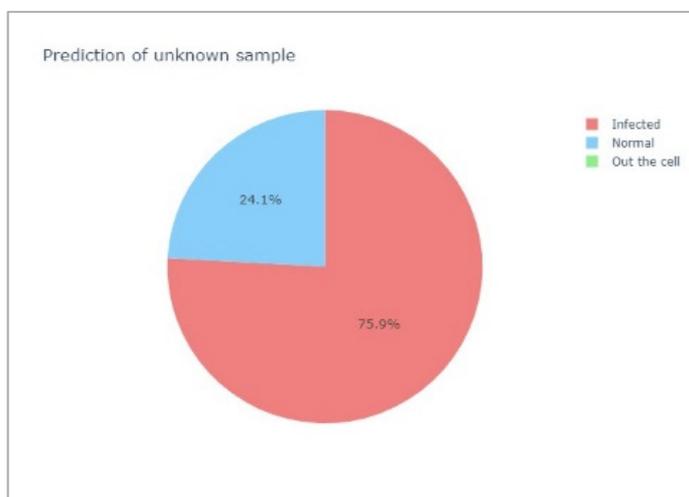


Figure 10. An example of a prediction pie chart.

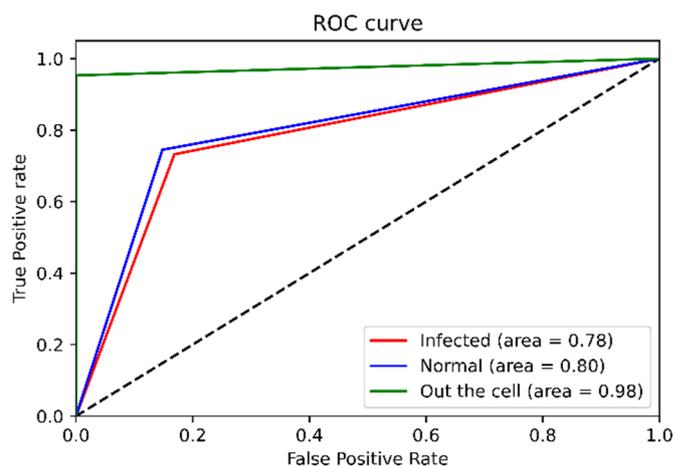
Table 3. Prediction results of blind samples.

Sample No.	Expected	Probability of Sample Class			Prediction
		Normal (%)	Infected (%)	Out (%)	
1	Infected	24.1	75.9	0.0	Infected
2	Normal	60.6	39.4	0.0	Normal
3	Infected	31.8	68.2	0.0	Infected
4	Infected	25.0	75.0	0.0	Infected
5	Normal	82.9	17.1	0.0	Normal
6	Normal	78.6	21.4	0.0	Normal
7	Out	0.0	6.2	93.8	Out
8	Out	3.1	0.0	96.9	Out

Table 4 shows the accuracy, precision, sensitivity (recall), and specificity of the blind data sets. The values dropped compared with using 20% split data from 1175 training spectra to be tested (see Table 2). In Figure 11, the area under the receiver operating characteristic curve shows the same trend, with the infected class having an area under the curve of 0.78 and the normal class having an area under the curve of 0.80. However, the efficiency of the out-of-cell class is quite clearly stable, regardless of whether the model was evaluated using blind data or data split from the 1175 training spectra set. That is, the separation model is highly efficient at classifying the inside- and outside-the-cell information.

**Table 4.** Efficiency of the model evaluated by blind sample data sets (test set).

	Precision	Sensitivity (Recall)	F1-Score	Specificity	Support
Infected	0.65	0.73	0.69	0.83	71
Normal	0.79	0.75	0.77	0.84	102
Out the cell	1.00	0.95	0.98	0.74	64
Accuracy		0.80			237
Macro avg	0.81	0.81	0.81		237
Weight avg	0.81	0.80	0.80		237

**Figure 11.** ROC curve of PCA-LDA model evaluation using blind sample data sets (test set).

In Figure 8, the testing set taken from the 1175 training spectra, the model demonstrates a high ability to discriminate with an average accuracy of over 90%, as illustrated in Table 2, and the area under the receiver operating characteristic curve is nearly one (Figure 8). It should be noted that in blind spectra data sets obtained from various operators, the distribution that was generated in each sample's spectral data varied, despite the fact that measurement parameters such as the laser's power, the time exposed to the sample, the lens, and so on were all controlled, including the procedures used in the sample preparation step. As the distribution in the training data was incompatible with the distribution in the testing data set, the efficacy of discriminating between data classes was reduced. However, the training spectra set in the PCA-LDA separation model is an updateable database and it can be updated infinitely. When a larger set of training data that covers the distribution in the testing set is incorporated into the existing database, the PCA-LDA separation could improve the efficiency of malaria-infected red blood cells identification.

## 5. Conclusions

Nowadays, there is still demand for the development of malaria diagnosis techniques. This study aims to develop an alternative method for malaria detection based on surface-enhanced Raman spectroscopy. Three crucial stages including the sample preparation, collecting the Raman spectra and data analysis are described in detail. Both average normalized spectra and spectral imaging were found to produce comparable results. During infection, the Raman intensity modification of the heme-based composition, which corresponds to the metabolically crystallized byproduct (hemozoin) of hemoglobin digestion, increased at  $\sim 1620$ ,  $1560$ ,  $1372$ , and  $1228\text{ cm}^{-1}$ , and decreased at  $747$  and  $1128\text{ cm}^{-1}$ . When PCA-LDA analysis was applied to the SERS spectra, the separation of infected and normal cell spectra was readily apparent in PCA-LDA coordinates. Additionally, the PCA-LDA coordinate separation could be used to predict the class of the blind sample cell. Each data set yielded the correct response. The separation model's efficiency can still be improved since the spectral database for training the model can be updated. In this study, only the qualitative results of the sample classification were reported. However, the protocols that

were applied can be reproduced by volunteers and all of the blind samples offered accurate predictions. This shows that the proposed approach can be employed in future quantitative malaria diagnosis development studies and can be refined so that it can be applied at a clinical application level. Furthermore, the spectra in this study can be utilized as references for any red blood cell or malaria-related Raman spectroscopy investigation.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/mps5030049/s1>, Figure S1. pie charts of the blind data sets, Supplement for Figure 10.

**Author Contributions:** Conceptualization, G.K., R.C. and T.T.; methodology, G.K. and T.T.; results and analysis, G.K.; discrimination analysis validation, G.K. and N.N.; writing—original draft preparation, G.K.; writing—review and editing, R.C., T.T., N.N. and N.P.; supervision, R.C., T.T., N.P. and N.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was made possible through financial support from the Science Achievement Scholarship of Thailand (SAST) which is under the supervision of the Ministry of Higher Education, Science, Research and Innovation, Thailand.

**Institutional Review Board Statement:** Human sera and erythrocytes from donors used for *P. falciparum* culture were obtained in accordance with the regulations of the Ethics Committee for Human Research, National Science and Technology Development Agency (NSTDA), Thailand, following an approved protocol (document number 0024/2561).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The author would like to thank the Science Achievement Scholarship of Thailand (SAST) for their financial contribution to this study and the National Electronics and Computer Technology Center (NECTEC, Thailand) for supporting the program and the use of the instruments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. NHS. Malaria-Causes-NHS. Available online: <https://www.nhs.uk/conditions/malaria/causes/> (accessed on 11 November 2019).
2. WHO. World Malaria Report. 2021. Available online: [https://www.mmv.org/newsroom/publications/world-malaria-report-2021?gclid=EA1aIQobChMI4-fHgPGf-AIVy24qCh0\\_7gAREAAYASAAEgJqNfD\\_BwE](https://www.mmv.org/newsroom/publications/world-malaria-report-2021?gclid=EA1aIQobChMI4-fHgPGf-AIVy24qCh0_7gAREAAYASAAEgJqNfD_BwE) (accessed on 3 February 2022).
3. Wongsrichanalai, C.; Barcus, M.J.; Muth, S.; Sutamihardja, A.; Wernsdorfer, W.H. A review of malaria diagnostic tools: Microscopy and rapid diagnostic test (RDT). *Am. J. Trop. Med. Hyg.* **2007**, *77* (Suppl. S6), 119–127. [[CrossRef](#)] [[PubMed](#)]
4. Bisoffi, Z.; Gobbi, F.; Ende, J.V.D. Rapid diagnostic tests for malaria. *BMJ* **2014**, *348*, g3846. [[CrossRef](#)] [[PubMed](#)]
5. Hermsen, C.C.; Telgt, D.S.; Linders, E.H.; van de Locht, L.A.; Eling, W.M.; Mensink, E.J.; Sauerwein, R.W. Detection of Plasmodium falciparum malaria parasites in vivo by real-time quantitative PCR. *Mol. Biochem. Parasitol.* **2001**, *118*, 247–251. [[CrossRef](#)]
6. Snounou, G.; Viriyakosol, S.; Zhu, X.P.; Jarra, W.; Pinheiro, L.; Rosario, V.E.D.; Thaithong, S.; Brown, K. High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. *Mol. Biochem. Parasitol.* **1993**, *61*, 315–320. [[CrossRef](#)]
7. Baptista, V.; Costa, M.S.; Calçada, C.; Silva, M.; Gil, J.P.; Veiga, M.I.; Catarino, S.O. The Future in Sensing Technologies for Malaria Surveillance: A Review of Hemozoin-Based Diagnosis. *ACS Sens.* **2021**, *6*, 3898–3911. [[CrossRef](#)]
8. Wood, B.R.; Langford, S.J.; Cooke, B.M.; Glenister, F.K.; Lim, J.; McNaughton, D. Raman imaging of hemozoin within the food vacuole of Plasmodium falciparum trophozoites. *FEBS Lett.* **2003**, *554*, 247–252. [[CrossRef](#)]
9. Wood, B.R.; McNaughton, D. Resonance Raman Spectroscopy of erythrocytes. *Appl. Life Pharm. Neutral Sci. Med. Diagn.* **2008**. [[CrossRef](#)]
10. Chen, K.; Yuen, C.; Aniweh, Y.; Preiser, P.; Liu, Q. Towards ultrasensitive malaria diagnosis using surface enhanced Raman spectroscopy. *Sci. Rep.* **2016**, *6*, 20177. [[CrossRef](#)]
11. Wood, B.R.; McNaughton, D. Resonance Raman spectroscopy in malaria research. *Expert Rev. Proteom.* **2006**, *3*, 525–544. [[CrossRef](#)]
12. Frosch, T.; Koncarevic, S.; Becker, K.; Popp, J. Morphology-sensitive Raman modes of the malaria pigment hemozoin. *Analyst* **2009**, *134*, 1126–1132. [[CrossRef](#)]
13. Yuen, C. Magnetic field enriched surface enhanced resonance Raman spectroscopy for early malaria diagnosis. *J. Biomed. Opt.* **2012**, *17*, 017005. [[CrossRef](#)] [[PubMed](#)]
14. Wood, B.R.; Bailo, E.; Khiavi, M.A.; Tilley, L.; Deed, S.; Deckert-Gaudig, T.; McNaughton, D.; Deckert, V. Tip-Enhanced Raman Scattering (TERS) from Hemozoin Crystals within a Sectioned Erythrocyte. *Nano Lett.* **2011**, *11*, 1868–1873. [[CrossRef](#)] [[PubMed](#)]

15. Grüning, C.; Heiber, A.; Kruse, F.; Ungefehr, J.; Gilberger, T.-W.; Spielmann, T. Development and host cell modifications of *Plasmodium falciparum* blood stages in four dimensions. *Nat. Commun.* **2011**, *2*, 165. [CrossRef] [PubMed]
16. Pilot, R.; Signorini, R.; Durante, C.; Orian, L.; Bhamidipati, M.; Fabris, L. A Review on Surface-Enhanced Raman Scattering. *Biosensors* **2019**, *9*, 57. [CrossRef]
17. Ding, S.-Y.; Yi, J.; Li, J.-F.; Ren, B.; Wu, D.-Y.; Panneerselvam, R.; Tian, Z.-Q. Nanostructure-based plasmon-enhanced Raman spectroscopy for surface analysis of materials. *Nat. Rev. Mater.* **2016**, *1*, 16021. [CrossRef]
18. Garrett, N.L.; Sekine, R.; Dixon, M.W.A.; Tilley, L.; Bambery, K.R.; Wood, B.R. Bio-sensing with butterfly wings: Naturally occurring nano-structures for SERS-based malaria parasite detection. *Phys. Chem. Chem. Phys.* **2014**, *17*, 21164–21168. [CrossRef]
19. Chen, F.; Flaherty, B.R.; Cohen, C.E.; Peterson, D.S.; Zhao, Y. Direct detection of malaria infected red blood cells by surface enhanced Raman spectroscopy. *Nanomed. Nanotechnol. Biol. Med.* **2016**, *12*, 1445–1451. [CrossRef]
20. Wang, W.; Dong, R.-L.; Gu, D.; He, J.-A.; Yi, P.; Kong, S.-K.; Ho, H.-P.; Loo, J.F.-C.; Wang, W.; Wang, Q. Antibody-free rapid diagnosis of malaria in whole blood with surface-enhanced Raman Spectroscopy using Nanostructured Gold Substrate. *Adv. Med. Sci.* **2020**, *65*, 86–92. [CrossRef]
21. Goh, B.; Ching, K.; Magalhães, R.J.S.; Ciocchetta, S.; Edstein, M.D.; Maciel-De-Freitas, R.; Sikulu-Lord, M.T. The application of spectroscopy techniques for diagnosis of malaria parasites and arboviruses and surveillance of mosquito vectors: A systematic review and critical appraisal of evidence. *PLoS Neglected Trop. Dis.* **2021**, *15*, e0009218. [CrossRef]
22. Ramoji, A.; Ryabchykov, O.; Galler, K.; Tannert, A.; Markwart, R.; Requardt, R.P.; Rubio, I.; Bauer, M.; Bocklitz, T.; Popp, J.; et al. Raman Spectroscopy Follows Time-Dependent Changes in T Lymphocytes Isolated from Spleen of Endotoxemic Mice. *ImmunoHorizons* **2019**, *3*, 45–60. [CrossRef]
23. Kolašinac, S.; Pečinar, I.; Danojević, D.; Stevanović, Z.D. Raman spectroscopy coupled with chemometric modeling approaches for authentication of different paprika varieties at physiological maturity. *LWT* **2022**, *162*, 113402. [CrossRef]
24. Liu, W.; Sun, Z.; Chen, J.; Jing, C. Raman Spectroscopy in Colorectal Cancer Diagnostics: Comparison of PCA-LDA and PLS-DA Models. *J. Spectrosc.* **2016**, *2016*, 1–6. [CrossRef]
25. Song, D.; Yu, F.; Chen, S.; Chen, Y.; He, Q.; Zhang, Z.; Zhang, J.; Wang, S. Raman spectroscopy combined with multivariate analysis to study the biochemical mechanism of lung cancer microwave ablation. *Biomed. Opt. Express* **2020**, *11*, 1061. [CrossRef] [PubMed]
26. Zong, M.; Zhou, L.; Guan, Q.; Lin, D.; Zhao, J.; Qi, H.; Harriman, D.; Fan, L.; Zeng, H.; Du, C. Comparison of Surface-Enhanced Raman Scattering Properties of Serum and Urine for the Detection of Chronic Kidney Disease in Patients. *Appl. Spectrosc.* **2021**, *75*, 412–421. [CrossRef] [PubMed]
27. Zepeda-Zepeda, M.A.; Picquart, M.; Irigoyen-Camacho, M.E.; Mejía-Gózález, A.M. Diagnosis of Dental Fluorosis Using Micro-Raman Spectroscopy Applying a Principal Component-Linear Discriminant Analysis. *Int. J. Environ. Res. Public Heal.* **2021**, *18*, 10572. [CrossRef]
28. Frame, L.; Brewer, J.; Lee, R.; Faulds, K.; Graham, D. Development of a label-free Raman imaging technique for differentiation of malaria parasite infected from non-infected tissue. *Analyst* **2018**, *143*, 157–163. [CrossRef]
29. Patel, S.K.; Rajora, N.; Kumar, S.; Sahu, A.; Kochar, S.K.; Krishna, C.M.; Srivastava, S. Rapid Discrimination of Malaria- and Dengue-Infected Patients Sera Using Raman Spectroscopy. *Anal. Chem.* **2019**, *91*, 7054–7062. [CrossRef]
30. Nuntawong, N.; Eiamchai, P.; Limwichean, S.; Wong-Ek, B.; Horprathum, M.; Patthanasettakul, V.; Leelapojanaporn, A.; Nakngonthong, S.; Chindaudom, P. Trace detection of perchlorate in industrial-grade emulsion explosive with portable surface-enhanced Raman spectroscopy. *Forensic Sci. Int.* **2013**, *233*, 174–178. [CrossRef]
31. Nuntawong, N.; Eiamchai, P.; Somrang, W.; Denchitcharoen, S.; Limwichean, S.; Horprathum, M.; Patthanasettakul, V.; Chaiya, S.; Leelapojanaporn, A.; Saiseng, S.; et al. Detection of methamphetamine/amphetamine in human urine based on surface-enhanced Raman spectroscopy and acidulation treatments. *Sens. Actuators B Chem.* **2017**, *239*, 139–146. [CrossRef]
32. Trager, W.; Jensen, J.B. Human Malaria Parasites in Continuous Culture. *Science* **1976**, *193*, 673–675. [CrossRef]
33. Wise, D. Savitzky-Golay Filters. 5 January 2019. Available online: <https://dkwise.wordpress.com/2019/01/05/savitzky-golay-filters/> (accessed on 31 January 2022).
34. James, T.M.; Schlösser, M.; Lewis, R.J.; Fischer, S.; Bornschein, B.; Telle, H.H. Automated quantitative spectroscopic analysis combining background subtraction, cosmic ray removal, and peak fitting. *Appl. Spectrosc.* **2013**, *67*, 949–959. [CrossRef] [PubMed]
35. Afseth, N.K.; Segtnan, V.H.; Wold, J.P. Raman Spectra of Biological Samples: A Study of Preprocessing Methods. *Appl. Spectrosc.* **2006**, *60*, 1358–1367. [CrossRef] [PubMed]
36. Gautam, R.; Vanga, S.; Ariese, F.; Umopathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Tech. Instrum.* **2015**, *2*, 8. [CrossRef]
37. Raschka, S. Linear Discriminant Analysis. 3 August 2014. Available online: [https://sebastianraschka.com/Articles/2014\\_python\\_lda.html](https://sebastianraschka.com/Articles/2014_python_lda.html) (accessed on 31 January 2021).
38. Perez-Guaita, D.; Marzec, K.M.; Hudson, A.; Evans, C.; Chernenko, T.; Matthäus, C.; Miljkovic, M.; Diem, M.; Heraud, P.; Richards, J.S.; et al. Parasites under the Spotlight: Applications of Vibrational Spectroscopy to Malaria Research. *Chem. Rev.* **2018**, *118*, 5330–5358. [CrossRef] [PubMed]
39. Carter, W. Raman Spectroscopy Study of Single Red Blood Cells Infected by the Malaria Parasite *Plasmodium falciparum* and the Novel Use of Micro-Capillaries. Master's Thesis, University of Central Florida, Orlando, FL, USA, 2007.

40. Sigala, P.A.; Crowley, J.R.; Hsieh, S.; Henderson, J.P.; Goldberg, D.E. Direct Tests of Enzymatic Heme Degradation by the Malaria Parasite *Plasmodium falciparum*. *J. Biol. Chem.* **2012**, *287*, 37793–37807. [[CrossRef](#)] [[PubMed](#)]
41. Wood, B.R.; McNaughton, D. Raman excitation wavelength investigation of single red blood cells *in vivo*. *J. Raman Spectrosc.* **2002**, *33*, 517–523. [[CrossRef](#)]
42. Atkins, C.; Buckley, K.; Blades, M.W.; Turner, R.F. Raman Spectroscopy of Blood and Blood Components. *Appl. Spectrosc.* **2017**, *71*, 767–793. [[CrossRef](#)]
43. Barkur, S.; Chidangil, S. Surface-enhanced Raman spectroscopy study of red blood cells and platelets. *J. Biomol. Struct. Dyn.* **2019**, *37*, 1090–1098. [[CrossRef](#)]
44. Barkur, S. Raman and Surface Enhanced Raman Spectroscopy Study of Blood Components. Manipal. 2018. Available online: <https://shodhganga.inflibnet.ac.in/handle/10603/209011> (accessed on 12 December 2019).
45. Goheen, S.C.; Lis, L.J.; Kucuk, O.; Westerman, M.P.; Kauffman, J.W. Compositional dependence of spectral features in the Raman spectra of erythrocyte membranes. *J. Raman Spectrosc.* **1993**, *24*, 275–279. [[CrossRef](#)]
46. Drescher, D.; Büchner, T.; McNaughton, D.; Kneipp, J. SERS reveals the specific interaction of silver and gold nanoparticles with hemoglobin and red blood cell components. *Phys. Chem. Chem. Phys.* **2013**, *15*, 5364–5373. [[CrossRef](#)]
47. Brazhe, N.A.; Abdali, S.; Brazhe, A.R.; Luneva, O.G.; Bryzgalova, N.Y.; Parshina, E.Y.; Sosnovtseva, O.V.; Maksimov, G.V. New Insight into Erythrocyte through *In Vivo* Surface-Enhanced Raman Spectroscopy. *Biophys. J.* **2009**, *97*, 3206–3214. [[CrossRef](#)] [[PubMed](#)]
48. Wood, B.R.; Langford, S.J.; Cooke, B.M.; Lim, J.; Glenister, F.K.; Duriska, M.; Unthank, J.K.; McNaughton, D. Resonance Raman Spectroscopy Reveals New Insight into the Electronic Structure of  $\beta$ -Hematin and Malaria Pigment. *J. Am. Chem. Soc.* **2004**, *126*, 9233–9239. [[CrossRef](#)] [[PubMed](#)]
49. Nikelshparg, E.I.; Baizhumanov, A.A.; Bochkova, Z.V.; Novikov, S.M.; Yakubovsky, D.I.; Arsenin, A.V.; Volkov, V.S.; Goodilin, E.A.; Semenova, A.A.; Sosnovtseva, O.; et al. Detection of Hypertension-Induced Changes in Erythrocytes by SERS Nanosensors. *Biosensors* **2022**, *12*, 32. [[CrossRef](#)] [[PubMed](#)]
50. Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassanien, A.E. Linear discriminant analysis: A detailed tutorial. *AI Commun.* **2017**, *30*, 169–190. [[CrossRef](#)]
51. Smith, L.I. A tutorial on Principal Components Analysis (Computer Science Technical Report No. OUCS-2002-12). 2002. Available online: [http://www.cs.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf) (accessed on 3 February 2022).
52. Jolliffe, I.T.; Cadima, J.; Cadima, J. Principal component analysis: A review and recent developments Subject Areas. *Philos. Trans. R. Soc. A* **2016**, *374*, 1–16. [[CrossRef](#)] [[PubMed](#)]