e-ISSN: 2689-4934

Available online at
http://uppp.uitm.edu.my/online-issues.html

ESTEEM Academic Journal Vol 20, March 2024, 65-81

ESTEEM
Academic Journal
(EAJ)

Surface-enhanced raman spectroscopy with machine learning in non-invasive detection of dengue-NS1 fingerprint

Afaf Rozan Mohd Radzol[1*], Khuan Y Lee[2], Peng Shyan Wong[3], Irene Looi[4], Wahidah Mansor[2]

[1]*Centre for Electrical Engineering Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, Pulau Pinang, Malaysia*
[2] *Center of System Studies, School of Electrical Engineering, Collage of Engineering, Universiti Teknologi MARA, Selangor, Malaysia*
[3]*Infectious Disease Unit, Penang General Hospital, Georgetown, Pulau Pinang, Malaysia*
[4]*Clinical Research Centre, Seberang Jaya Hospital, Seberang Jaya, Perai, Pulau Pinang, Malaysia*

ARTICLE INFO

Article history:
Received 22 January 2024
Revised 12 March 2024
Accepted 23 March 2024
Online first
Published 31 March 2024

Keywords:
SERS
Machine Learning
Dengue
NS1

DOI:
10.24191/esteem.v20iMarch.616.g534

ABSTRACT

The surface-enhanced Raman spectroscopy (SERS) method exploits the plasmonic effect of nano-sized metallic materials to intensify the Raman scattering of the monochromatic light of analyte molecules. This promotes the sensitivity and specificity of the Raman spectroscopy analysis method. This study integrated SERS with machine learning (ML) to detect dengue fever, a disease infecting more than 40% of the world's population. Non-structural protein 1 (NS1), detected in the sera of infected dengue patients during the early infection stage, is currently recognised as a biomarker for the early diagnosis of DF. However, no attempts have been made to detect NS1 in the salivary Raman spectra. Given this situation, this study delves into the potential of SERS as an early, non-invasive DF detection technique using salivary NS1. The SERS spectra of saliva samples (n=289) were collected and subsequently classified as positive and negative for DF, using principal component analysis (PCA) integrated with support vector machine (SVM) models. The PCA-SVM model's performance was benchmarked against two clinical diagnostic NS1-enzyme-linked immunosorbent assay (ELISA) tests recommended by the World Health Organization (WHO). The PCA-SVM model outperformed both tests regarding radial basis function kernel (RBF) and cumulative percent variance (CPV; 83.22% accuracy, 88.27% sensitivity, 78.13% specificity). It is encouraging that the sensitivity level of the PCA-SVM model is above the benchmark set by the saliva-based NS1-ELISA tests proposed by the

[1*] Corresponding author. *E-mail address*: afafrozan944@uitm.edu.my
https://doi.org/10.24191/esteem.v20iMarch.616.g534

WHO, demonstrating the potential of SERS for the non-invasive detection of DF.

1. INTRODUCTION

Dengue fever (DF) is endemic in tropical and subtropical regions. In Malaysia, since 2014, more than 100000 cases have been reported annually, with more than 200 fatalities, rendering this disease a major public health concern. As a vaccine against DF is yet to be developed, early detection is crucial for reducing mortality. Non-structural protein 1 (NS1), an antigen that stimulates the production of antibodies to combat the dengue virus, was recently acknowledged as a new biomarker for the early detection of dengue [1]. The presence of NS1 in the blood circulation of infected individuals has been detected as early as Day 1 of the early infection stage and persists until Day 9 [2-3]. The detection of NS1 usually involves enzyme-linked immunoassay (ELISA) and immune-chromatographic lateral flow tests. Using ELISA, Ander et al. [4] reported the presence of NS1 in saliva with a sensitivity level of 64.7%, which is lower than that for blood. Despite utilising various analysis techniques [5-8] to identify the precise structure and function of NS1, this objective remains beyond reach. The development of an innovative and effective procedure for treating and preventing DF calls for an unexplored approach to discerning the NS1's protein structure.

This undertaking involves the implementation of a surface-enhanced Raman spectroscopy (SERS), which integrates nano-technology, to augment the intensity of Raman scattering by an enhancement factor of 10^{10} - 10^{11}, thus promoting its capacity for detecting low-concentration analyte. In an automated, non-invasive attempt to detect NS1 in saliva, the salivary Raman spectra of positive and negative dengue patients were acquired, processed and classified using machine learning (ML) algorithms, specifically a vector machine (SVM) with linear (Linear), radial basis function (RBF), and multi-layer perceptron (MLP) kernels. Before classification, a feature extraction method, namely principal component analysis (PCA), was used to extract significant features from the spectra.

2. RESEARCH BACKGROUND

2.1 Non-structural protein 1 (NS1)

Non-structural protein 1 (NS1) is among the non-structural proteins encoded by the flavivirus genome. NS1 is not part of the viral particles but is synthesised in the host cell following infection. It is encoded by a 352-amino-acid polypeptide, with 1056 nucleotides in length and 46-55 kDa in molecular weight, depending on its glycosylation status [9-11]. These characteristics are shared by all virus types of the flavivirus genus [12].

NS1 exists in multiple oligomeric forms. Upon flavivirus infection, the hydrophilic NS1 monomer is produced from the signal sequencing of the envelope (E) protein of the viral genome at its C-terminus, during initial translation in the endoplasmic reticulum (ER) lumen. It contains 12 cysteines to form six discrete disulphide bonds. The bonds are deemed important for the structure and function of NS1 [8], [12-13]. Following the addition of high-mannose carbohydrates to the monomer, the cell-membrane-associated NS1 (mNS1) with hydrophobic properties produces NS1 as a dimmer. Parts of the dimmer are anchored with glycosyl-phosphatidylinositol (GPI) from the N-terminus of NS2A, forming GPI-NS1. Both mNS1 and GPI-NS1 are present on the cell surface associated with lipid raft. A portion of mNS1 is trafficked to Golgi from ER and occurs in soluble hexameric form, with lipid cargo stored in its central channel [14]. This lipoprotein is further modified in Golgi, producing a complex form of carbohydrate, which is then secreted from the cell known as sNS1.

https://doi.org/10.24191/esteem.v20iMarch.616.g534

While researchers have made attempts from different disciplines to explain the structure and function of NS1, these explanations remain inconclusive [5-6], [15]. Relevant literature reports the highly immunogenic capacity of secreted NS1 (sNS1) [16]. With dengue, NS1, in an estimated amount of 0.01-50 mg/L [2] is detectable in the patient's blood serum prior to the formation of antibodies IgM and IgG [2-3]. The NS1 amount correlates to the disease pathogenesis, with the traces of NS1 increasing as DF progresses to dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS) [3]. Also, the mNS1 has been observed to co-localise with double-stranded RNA (dsRNA), indicating its role in virus replication [17].

2.2 Dengue detection: raman spectroscopy

Raman spectroscopy is specific as it produces the Raman spectrum based on the vibration unique of each molecular structure. As monochromatic light makes contact with the analyte molecules, the interaction between light photons and molecules causes a tiny fraction of the light photons to scatter at a frequency dissimilar to the incident light. This difference in frequency is measured as the Raman shift. The occurrence of this phenomenon is exceedingly subtle, rendering it difficult to capture, particularly in a circumstance where the analyte is of low concentration. This stumbling block can be overcome with SERS, through which Raman scattering is amplified, by attaching the analyte to a noble metal denoted the SERS substrate [18]. The SERS substrate comes as a colloidal or roughened solid surface. The SERS substrate boasts a low detection limit as subtle as a single molecule [19-20]. SERS renders Raman spectroscopy a specific and sensitive method with great potential for many applications, particularly those associated with disease detection [21-25]. A SERS compilation is provided in [26], with the classification performance according to various diagnostic mediums.

The National Institute of Lasers and Optronics of Pakistan (NILOP) explored the potential of Raman spectroscopy for detecting dengue based on the Raman spectra of blood samples from six dengue-positive patients and healthy volunteers. The use of the partial least squares (PLS) technique revealed the coefficient of determination as (R^2) = 0.9998, indicating a strong correlation between the predicted and reference results [27]. Shortly, in 2016, this investigation conducted by the NILOP reported R^2=0.91, with 95 dengue-positive subjects and 123 healthy volunteers, as well as Raman peaks with strong regression coefficients and associated biomolecules, at 736, 776, 1127, 1045, and 1454 cm^{-1}, identified through the PLS technique [28].

A higher normalised intensity of Raman peak was observed at 1002cm^{-1} in dengue blood serum, with 40 dengue-positive patients and 25 healthy volunteers [29]. The dengue detection process then proceeded with classifiers to detect the disease in the blood samples. The use of a principal component analysis (PCA)-support vector machine (SVM), with a second-order polynomial kernel and five principal components (PCs), delivered a classification performance of 85, 73, and 93% against IgM ELISA [30]. Table 1 summarises other areas' efforts [31-34]. To the best of our knowledge, using SERS to detect NS1 in saliva and indicate dengue virus infection remains unexplored.

2.3 Machine learning (ML): principal component analysis (PCA)-support vector machine (SVM) algorithm

Surface-enhanced Raman spectroscopy (SERS) spectra come with multiple peaks from bio-molecules such as blood and saliva. Depending on the region of interest and instrumentation resolution, thousands of points are produced per spectrum. Furthermore, changes in the bio-molecular constituents bring about alterations in the Raman spectrum, which are invisible to the naked eye. As such, feature extraction and ML techniques must be applied to facilitate automated detection from the SERS spectra.

Principal component analysis (PCA) uses an orthogonal transformation to convert a set of observations consisting of correlated variables into ranked uncorrelated variables known as PCs [35]. Hence, Raman

https://doi.org/10.24191/esteem.v20iMarch.616.g534

shifts with significant variance are retained in the higher-ranked PCs, which are useful features for the classification algorithm. The use of three PCA stopping criteria, namely Cattell's scree test (CST), cumulative per cent variance (CPV), and Eigenvalue-one-criterion (EOC), as guidelines for the extraction of significant PCs, served to reduce the dimension of the data.

Table 1. Classification of dengue-positive samples using Raman spectra of blood serum. The following extant studies and [26] examined dengue classification based on Raman spectral analysis. Earlier works can be found in. Notably, the classification performance is higher as the population study was small.

Author, year	Classification algorithm	Number of samples	[Accuracy, Sensitivity, Specificity]
[30] Khan, 2016	PCA-SVM	Positive (31) Negative (53)	85%, 73%, 93%
[31] Khan, 2017	PCA-RF	IgM positive (45) IgM negative (55)	91%, 91%, 91%
[32] Amin, 2017	PCA-LDA	Positive (32) Negative (28)	96.50%, 93.44%, 100%
[33] Mahmood, 2018	PCA-FDA	Positive (17) Negative (17)	NA, 97.38%, 86.18%
[34] Gahlaut, 2020	PCA	Positive (11) Negative (9)	NA

Support vector machine (SVM) is an ML algorithm based on supervised learning. For a given data pool, SVM is trained to maximise the kernel function, separating the d-dimensional data into two classes. The separating plane, known as the hyperplane, is established when the distance between the closest negative and positive samples is equal [36]. A kernel function maps the original data from the input space to the feature space to facilitate the separation of the input data by the hyperplane classifier into different classes with minimal error. Eq. (1), Eq. (2) and Eq. (3) depict the mathematical representation of the SVM kernel functions implemented for this study.

$$K_{LINEAR}(x_i, x_j) = x_i . x_j + 1 \tag{1}$$

$$K_{RBF}(x_i, x_j) = e^{\frac{\|x_j - x_i\|^2}{2\sigma^2}} \tag{2}$$

$$K_{MLP}(x_i, x_j) = \tanh^{-1}(P1 . x_i . x_j^T + P2) \tag{3}$$

where, x_i and x_j are the closest points in the real space, X and σ are the RBF's parameters, P1 is the scaling parameter or the weight, and P2 is a shifting parameter that controls the threshold for mapping of the MLP. The kernel parameters were tuned to obtain the optimum classification performance regarding accuracy, sensitivity and specificity using Eq. (4), Eq. (5) and Eq. (6) respectively.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

https://doi.org/10.24191/esteem.v20iMarch.616.g534

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

where, true positive (TP) is the number of positive cases correctly classified as positive, true negative (TN) is the number of negative samples correctly classified as negative, false negative (FN) is the number of positive cases incorrectly classified as negative, and false positive (FP) is the number of negative cases incorrectly classified as positive. To determine the number of TPs, TNs, FPs, and FP, the predicted results obtained from the algorithm were compared with clinical diagnostic test results obtained using Panbio™ Dengue Early ELISA and/or Bioline™ Dengue Duo Rapid.

3. METHODOLOGY

This investigation involves the performance of three main steps, as portrayed in Fig. 1.

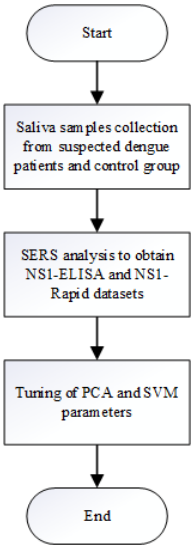
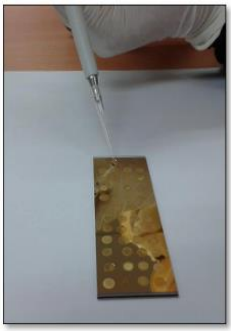


Fig. 1. A comprehensive flowchart of the present study

3.1 Saliva collection and sample preparation

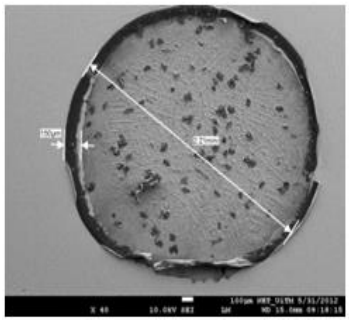
The Raman spectra of 289 saliva samples, collected from 229 suspected dengue patients and 60 healthy volunteers, were analysed. In compliance with a published protocol [37], 3 ml of unstimulated whole saliva was collected in the morning (between 9 am and 11 am). The test subjects were instructed to refrain from consuming food and drinks (other than plain water) for a minimum of one hour prior to the saliva collection exercise. Ten minutes before saliva collection, the test subjects were instructed to gargle thoroughly for one minute to remove debris. The samples were then centrifuged at 14000 rpm for 10 minutes for the extraction of the supernatants, which were subsequently deposited onto gold (Au)-coated substrates and left to dry at room temperature in a desiccator (Fig. 2). When a saliva sample is dropped and dried on a substrate, it forms a circular-shaped deposit, designated the ‘coffee-ring effect’ crystallisation structure (Fig. 2c). This structure comprises tree-like formations located in the centre of the drop, with an amorphous outer ring formation, and volume deposition dispersion, scattered across the drop.

https://doi.org/10.24191/esteem.v20iMarch.616.g534



(a) 10 uL of the saliva supernatant was deposited onto the Au-coated slide.

(b) The deposited samples were then dried in a desiccator to avoid contamination.



(c) A FESEM image of the saliva drop on the Au-coated slide.

Fig. 2. The drop-coated sample

The assumed DF patients were recruited from Hospital Pulau Pinang, Hospital Seberang Jaya and district polyclinics in Selangor and tested with Panbio™ Dengue Early ELISA and/or Bioline™ Dengue Duo Rapid. The results derived from the serological tests were used as the benchmark against the classifiers of this study. The relevancy of the procedures employed regarding the human subjects in this study meets the requirements of the National Medical Research Registry (NMRR) of Malaysia.

3.2 Surface-enhanced raman spectroscopy (SERS) analysis

Surface-enhanced Raman spectroscopy (SERS) analysis was conducted with a PeakSeeker Pro™-785 Raman spectrometer integrated with RSIQ™ software in a dark room to minimise spectral noise stemming from stray lights. The spectrometer houses a 785 nm near-infrared diode laser source. A 5 nm chromium (Cr) adhesion layer, sandwiched between an auto-cleavable corrugated high-quality soda-lime glass standard microscope slide and 50nm of Au coating using the vacuum evaporation technique, is the substrate. The spectrometer was operated at its optimal setting [38], interpreted as an exposure time of 30 seconds, a laser power of 100% (300 mW) and a microscope lens magnification of 50 times. With the detection range set to 200-2000 cm^{-1}, measurements were performed at different spots across the drop of every sample by moving the microscope stage at ten repetitions per spot.

The acquired spectra were sorted into two datasets, NS1-ELISA and NS1-Rapid, according to the gold standard tests used as the benchmark against the ML algorithm developed during this study. The NS1-ELISA dataset comprises salivary SERS spectra benchmarked against Panbio™ Dengue Early ELISA, while salivary SERS spectra in the NS1-Rapid dataset are benchmarked against Bioline™ Dengue Duo Rapid.

Non-structural protein 1 (NS1)-ELISA holds 142 positive spectra obtained from 142 dengue-positive patients and 142 negative spectra obtained from healthy volunteers (n = 70), as well as dengue-negative patients (n = 72). Meanwhile, NS1-Rapid holds 105 positive spectra from 105 dengue-positive patients, 105 negative spectra obtained from 72 dengue-negative samples, and 73 healthy volunteers. Spectra from healthy volunteers are used to ensure a balance between the number of positive and negative spectra. An example, a pre-processed SERS spectrum obtained from one of the samples, is presented in Fig. 3.

https://doi.org/10.24191/esteem.v20iMarch.616.g534

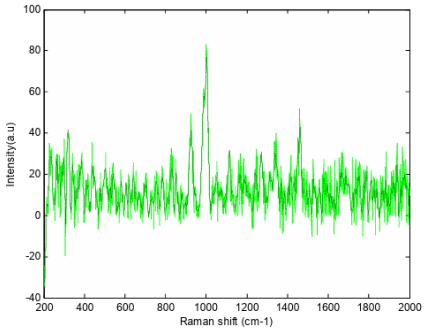


Fig. 3: The pre-processed SERS spectra obtained from one of the saliva samples.

3.3 Tuning the machine learning (ML) algorithm

Fig. 4 shows the overall process for the development of the ML algorithm. Every spectrum in both datasets displays 1801 features corresponding to the range of Raman shift (200-2000 cm^{-1}). This presents a total data dimension of 284 x 1801 for NS1-ELISA, and 210x1801 for NS1-Rapid. Initially, the spectra are pre-processed through an optimised four-stage pre-processing technique, entailing background subtraction, baseline removal [39], smoothing [40], and normalisation. The pre-processed spectra were then analysed using PCA, to produce PCs ranked based on their eigenvalues. Principal components (PCs) with the highest eigenvalues are ranked at the top of the list, followed by those with less prominent eigenvalues. To reduce the data dimension, three PCA stopping criteria, namely Cattell's scree test (CST), CPV and EOC, were applied to select the PCs. CST selects PCs above the knee in the plot of eigenvalues known as the scree plot [41]. The number of PCs retained by the CPV criterion is determined via CPV > 90%, while EOC retains all PCs with eigenvalues >1.

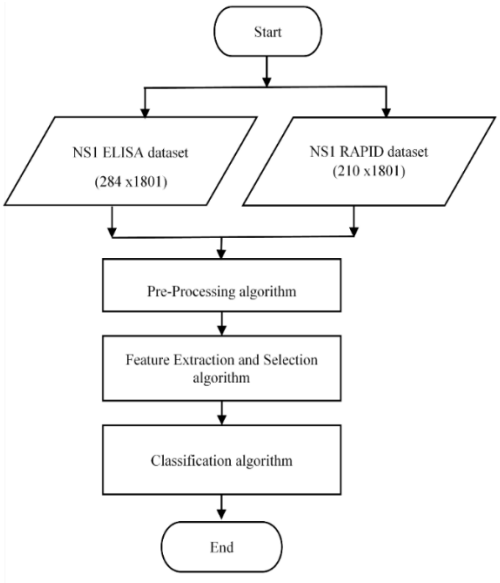


Fig. 4. The automated detection algorithm for classifying dengue-infected samples. It consisted of pre-processing, feature extraction, and selection, followed by SVM tuning. The pre-processing algorithm comprised four procedures: background removal, baseline removal, smoothing, and normalisation [39-40].

https://doi.org/10.24191/esteem.v20iMarch.616.g534

The tuning required in the classification procedure is illustrated in Fig. 5. The number of PCs from the stopping criteria serves as inputs for the SVM classifier. In order to avoid the occurrence of overfitting, a cross-validation technique was implemented with K=3. The three SVM kernels expressed in Eq. (1) to Eq. (3) (Linear, RBF and MLP), were tuned for optimal classification of the dengue-infected samples from the salivary Raman spectra. 300-thousands and more SVM models were evaluated and configured with the three PCA stopping criteria and model parameters (C-, σ-, P1, and P2).

The tuning parameters for the models are tabulated in Table 1; for the Linear kernel, the only tuned parameter is box constraint or C-value. For the RBF kernel, C-values and RBF gamma values (γ) were varied [42]. Meanwhile, for MLP-SVM, P1-value and P2-value were experimented with. P1 corresponds to the slope of the sigmoid function denoted as weight, while P2 corresponds to the sigmoid function intercept value.

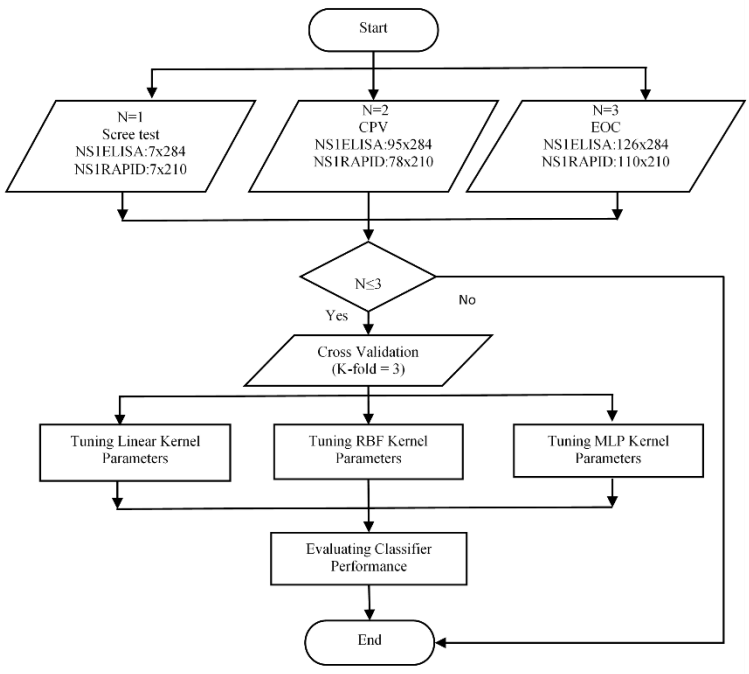


Fig. 5. The tuning for the optimal SVM classifier model.

Table 1. SVM Tuning parameters and values

Kernel	Parameter	Value
Linear	Box constraint (C-value)	0.01, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 100
RBF	Box constraint (C-value)	0.01, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 100
RBF	gamma (γ) value	$\log 10^{-9}$, $\log 10^{-8}$, $\log 10^{-7}$, $\log 10^{-6}$, $\log 10^{-5}$, $\log 10^{-4}$, $\log 10^{-3}$, $\log 10^{-2}$, 0.2, 0.4, 0.8, 1, 2, 5, 10, 20, 100, 1000, 2000
MLP	P1-value	0.01, 0.05, 0.1, 0.5, 5 and 10
MLP	P2-value	0 to 500 at increment at 0.01

https://doi.org/10.24191/esteem.v20iMarch.616.g534

4. RESULTS AND DISCUSSION

4.1 Principal component analysis (PCA) for feature extraction of NS1

The average SERS spectra of saliva, obtained from the positive and negative samples in the NSI-ELISA dataset, are shown in Fig. 6. The non-structural protein 1 (NS1) characteristic peak at 1000 cm^{-1} is invisible to the naked eye due to the low concentration of NS1 in the saliva of positive samples. This renders the unaided extraction of NS1 features from the spectra an exceptional challenge. These spectra were analysed with PCA to extract the difference between the positive and negative spectra, using ranked eigenvalues and PCs to overcome this setback.

Fig. 7 depicts the eigenvalues of the dataset ranked by PCA. The highest eigenvalue (252.66) corresponds to PC1, which carries 26.69% of the total variance of the dataset. As can be gathered from the Scree plot, the eigenvalues of the subsequent PCs are lower.

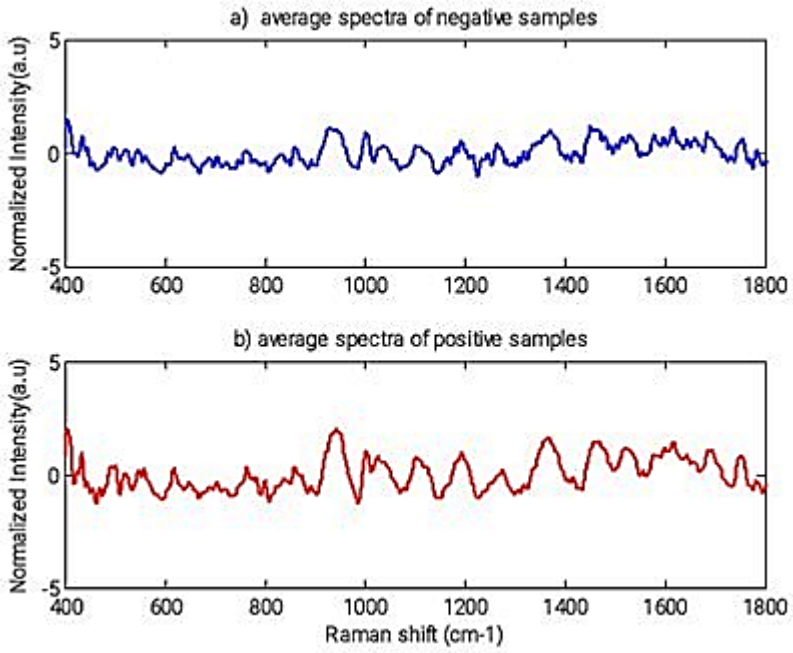


Fig. 6: The average Raman spectra of the NS1-ELISA dataset (a) negative and (b) positive samples

As shown in Fig. 8, with the loading plot of the three most significant PCs for the NS1-ELISA dataset, the slight disparity between positive and negative spectra produces loading plots with minute differences. However, PC3 negative loading value at Raman shift 1000 cm^{-1} matches the signature of NS1 protein as discussed in Fig. 6. This observation suggests that PC3 might carry the NS1 protein features, which can be useful for classifying the positive and negative cases.

A noteworthy situation was observed on the 2D score plot of the two most significant PCs, namely PC1 and PC2 (Fig. 9). As all the positive samples (red +) portray smaller PC1, they are distributed to the left of the plot. As such, the NS1 features of the positive spectra are preserved in PC1. However, 57.75% (82/142) of the negative samples exhibit similar PC1. To realise a better perception of the PC score distribution, PC3 was included in the 3D score plot (Fig. 10). As can be observed in the 2D and 3D score plots, the data is

https://doi.org/10.24191/esteem.v20iMarch.616.g534

not linearly separable, an indication that more PCs need to be included for an improved positive from negative sample classification. However, as too many PCs will result in excessive computation time and load, CST, CPV, and EOC were utilised to determine the number of PCs retained.

According to CST, the inflexion point of the Scree plot curve (Fig. 11) is evident at PC7, suggesting 7 PCs are retained. The bar chart (Fig. 11) depicts the CPV with a 90% threshold selected for the NS1-ELISA dataset. Ninety-five PCs account for 90% of the total variance of the dataset. Regarding the eigenvalues, 126 are ascertained to meet the EOC criterion (Fig. 7).

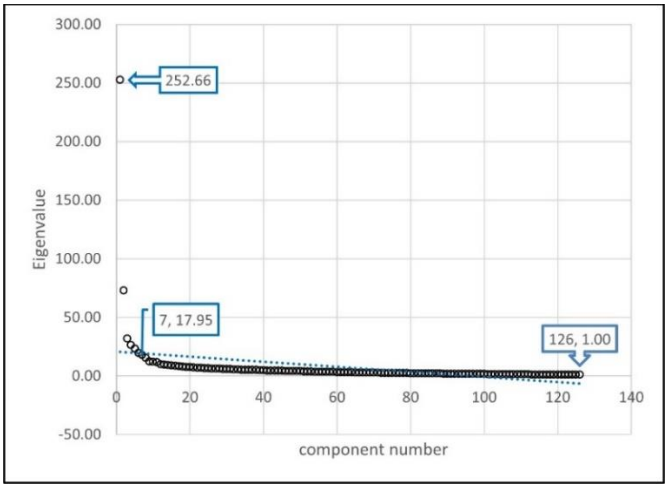


Fig. 7. The Scree plot of the NS1-ELISA dataset

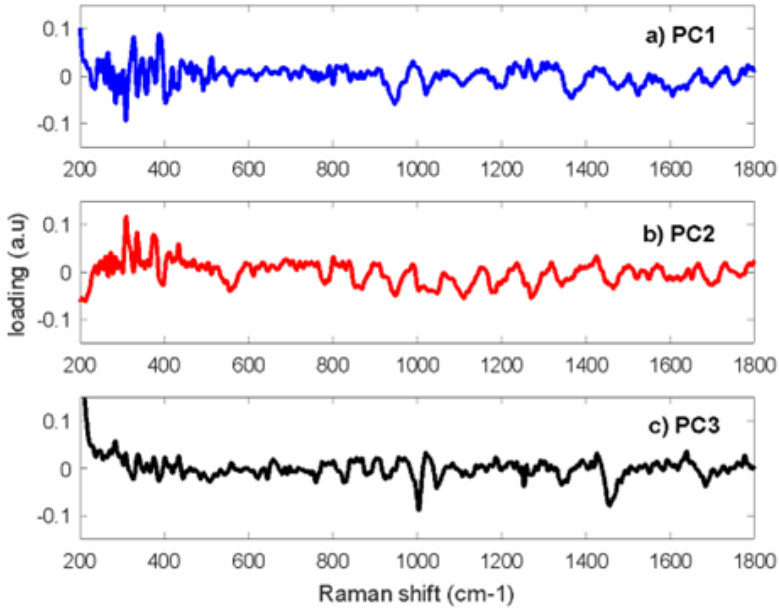


Fig. 8. The loading plots of PC1, PC2, and PC3 for the NS1-ELISA dataset

https://doi.org/10.24191/esteem.v20iMarch.616.g534

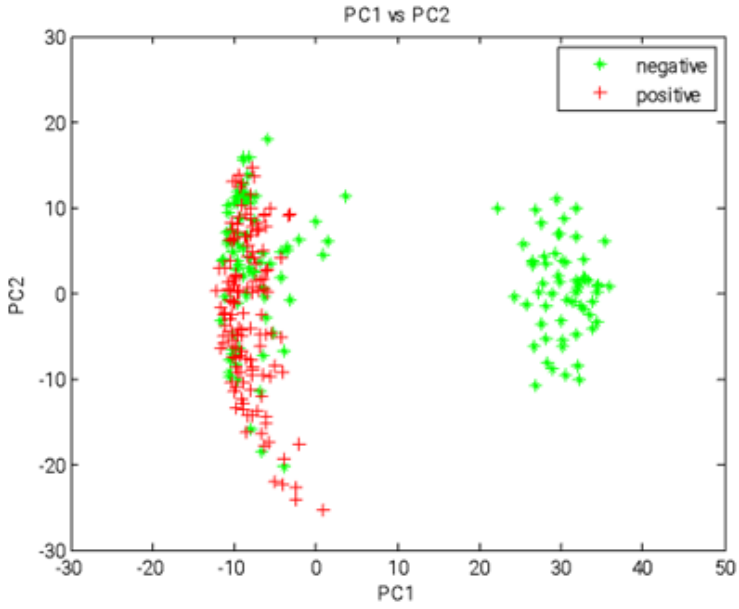


Fig. 9. The 2D score plot of PC1 and PC2 for the NS1-ELISA dataset

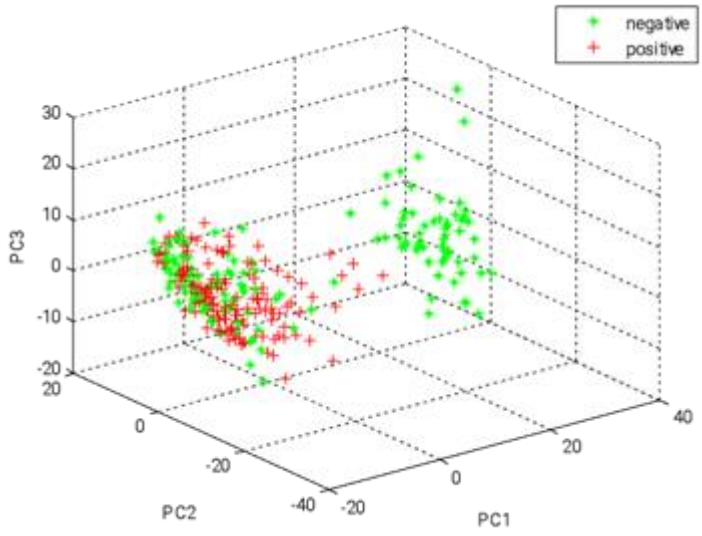


Fig. 10. The 3D score plot for PC1, PC2, and PC3 for the NS1-ELISA dataset

A summary of the dimension reduction capacity, according to the respective criteria, is provided in Table 2. The EOC retained 126 PCs with eigenvalue >1, which carries 94.49% of the total variance, reducing the data dimension by 93.00%. The CPV retained 95 PCs, which carries 90% of the total variance, as set by the threshold, reducing the data dimension by 94.73%. The CST delivered a data dimension reduction of 99.61%, the highest among the three criteria, by retaining only 7 PCs. Similar methods were

https://doi.org/10.24191/esteem.v20iMarch.616.g534

applied to determine the number of retained PCs for the NS1-Rapid dataset. The retained PCs of the NS1-Rapid dataset are presented in Table 3.

For both datasets, the CST heads the list for the least number of PCs to deliver the highest reduction in data dimension. However, fewer PCs denote less cumulative variance, 46.94% for NS1-ELISA and 51.63% for the NS1-Rapid dataset of the original. The CPV and EOC proposed additional PCs, which led to greater cumulative variance. Nevertheless, a data dimension reduction of above 90% was observed. Reducing the data dimension relieves the computation load and time consumption during the subsequent processing stage.

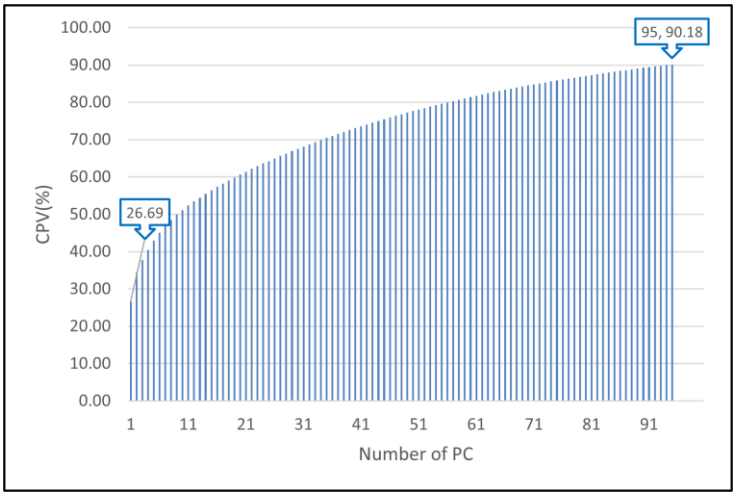


Fig. 11. The CPVs of the NS1-ELISA dataset

Table 2. The NS1-ELISA dataset as benchmarked against the Panbio™ Dengue Early ELISA and the number of PCs, data dimension, and CPV decreased in order of the stopping criteria.

	Number of PCs	Dimension of Significant Features	% of Cumulative Variance	% of Reduction
Raw Data	-	284 x 1801	-	-
EOC	126	284 x 126	94.49	93.00%
CPV	95	284 x 95	90.18	94.73%
CST	7	284 x 7	46.94	99.61%

Table 3. The NS1-Rapid dataset as benchmarked against the Bioline™ Dengue Duo Rapid and the number of PCs, data dimension, and CPV decreased in order of the stopping criteria.

	Number of PCs	Dimension of Significant Features	% of Cumulative Variance	% of Reduction
Raw Data	-	210 x 1801	-	-
EOC	110	210 x 110	95.17	93.89%
CPV	78	210 x 78	90.00	95.67%
CST	7	210 x 7	51.63	99.56%

https://doi.org/10.24191/esteem.v20iMarch.616.g534

4.2 Support vector machine (SVM) for classification of NS1

Table 4 depicts the optimal performance of classifiers of different kernels and PCA stopping criteria for the NS1-ELISA dataset. An increase in the number of PCs improves specificity while deteriorating sensitivity. The highest level of accuracy is achieved with the application of the PCs retained by CPV. However, accuracy is not positively associated with more PCs, as it decreased in the EOC situation due to model overfitting. The best accuracy, sensitivity, and specificity performance was recorded as 83.22, 88.27, and 78.13%, respectively, with RBF and 95 PCs.

Table 4. The optimal performance of the classifier models for the NS1-ELISA dataset when the classifier models were optimally tuned

	PCA stopping criteria	Number of PCs	Accuracy	Sensitivity	Specificity
Linear-SVM	CST	7	72.53	99.29	45.77
RBF-SVM	CST	7	73.84	85.9	61.77
MLP-SVM	CST	7	72.54	99.3	45.8
Linear-SVM	CPV	95	76.41	88.03	64.79
RBF-SVM	CPV	95	83.22	88.27	78.13
MLP-SVM	CPV	95	76.88	85.92	67.83
Linear-SVM	EOC	126	73.58	81.71	65.46
RBF-SVM	EOC	126	79.23	79.09	79.33
MLP-SVM	EOC	126	74.3	79.56	69.05

The performances of the optimal SVM classifier models for the NS1-Rapid dataset are summarised in Table 5. Like the NS1-ELISA dataset, increasing the number of PCs raises the specificity and accuracy levels. This implies that more of the lower-ranked PCs carry useful information. As the number of PCs retained for the NS1-Rapid dataset is less than that for the NS1-ELISA dataset, model overfitting does not occur for the former. The best performance of 81.90%, 80.32% and 83.49% for accuracy, sensitivity, and specificity, respectively, is achieved by the SVM model with RBF, with 110 retained PCs from the EOC stopping criterion. Overall, the performance level of the NS1-Rapid dataset is below that of the NS1-ELISA.

Table 5. The optimal performance of the classifier models for the NS1-Rapid dataset when the classifier models were optimally tuned

Classifier	PCA stopping criteria	Number of PCs	Accuracy	Sensitivity	Specificity
Linear-SVM	CST	7	71.11	81.27	60.95
RBF-SVM	CST	7	70.3	81.6	59
MLP-SVM	CST	7	68.41	77.46	59.37
Linear-SVM	CPV	78	73.17	74.92	71.43
RBF-SVM	CPV	78	79.05	76.83	81.27
MLP-SVM	CPV	78	76.19	81.27	71.11
Linear-SVM	EOC	110	76.03	80.32	71.75
RBF-SVM	EOC	110	81.9	80.32	83.49
MLP-SVM	EOC	110	78.41	74.92	81.9

The results in Tables 4 and Table 5 indicate that the PCA-SVM classifiers efficiently distinguished between the positive and negative DF cases, which would otherwise have been indistinguishable to the naked eye. It is heartening to realise that the sensitivity level, benchmarked against both the WHO-recommended clinical tests, is above the 64.7% reported by Ander [4], and the 73% reported by Khan [30],

https://doi.org/10.24191/esteem.v20iMarch.616.g534

with the accuracy level also comparable to that of the latter, with blood as the medium. Their accuracy was comparable to the latter, with blood as the medium. It is undeniable that blood serum contains a higher content of biomarkers than saliva, a fact demonstrated by the number of PCs retained. However, the DF diagnosis process, by way of blood serum, can turn out to be a painful or even traumatic experience for the patient. As with all serum-based tests, it is prone to blood-borne infection, and its collection procedure is complicated, necessitating the involvement of certified staff. Between the two recommended clinical tests, the accuracy and sensitivity levels were reported as higher with fewer PCs than the ELISA test, the gold standard for clinical diagnosis of DF, except that in this case, an antigen was used in place of antibodies.

5. CONCLUSION

In this study, surface-enhanced Raman spectroscopy (SERS), a sensitive and specific analysis technique, was integrated with SVM, an ML technique, to determine if dengue positive and negative cases can be distinguished, based on the salivary NS1 fingerprint from the SERS spectra. Optimally-tuned SVM classifiers with Linear, RBF, and MLPs were benchmarked against the two WHO-recommended clinical tests. Before SVM, PCA transformed SERS spectra obtained from salivary samples into PCs. Three PCA-stopping criteria (EOC, CPV and CST) were used to exclude the insignificant PCs from the input dataset for SVM. Our findings reveal optimal performances of 83.22%, 88.27% and 78.13% for accuracy, sensitivity and specificity, respectively, for the NS1- ELISA dataset using SVM with RBF and CPV (95 PCs), and 81.9%, 80.32%, and 83.49 for accuracy, sensitivity and specificity respectively, for the NS1-Rapid dataset using SVM with RBF and EOC (110 PCs). Both datasets reported higher sensitivity levels than the previously reported sensitivity levels of 64.7% [4] and 73% [28]. We recommend that further research in this area focus on enhancing the performance of the SERS substrate, specifically for detecting NS1.

6. ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Science, Technology, and Innovation of Malaysia for providing Grant ERGS/1/2013/TK02/UITM/02/01 and the Research Management Centre and Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia, for their support and assistance in carrying out this research. The authors would also like to thank the microbiologists and specialists at the Infectious Disease Unit, Pulau Pinang General Hospital and Dr Michelle M. Samy from the Clinical Research Centre, Seberang Jaya General Hospital, for helping collect patient data and providing valuable clinical advice.

7. CONFLICT OF INTEREST STATEMENT

The authors agree that this research was conducted in the absence of any self-benefits, commercial or financial conflicts and declare the absence of conflicting interests with the funders.

8. AUTHORS' CONTRIBUTION

Afaf Rozan Mohd Radzol: Contributed to the experimental design, focusing on sample preparation, data collection methods, optimization of machine learning algorithms, and manuscript writing. **Khuan Y. Lee**: Secured funding for the research project, including grants and resources necessary for conducting experiments and analysis. She also was involved in writing and reviewing the manuscript. **Wahidah Mansor:** Contributed to reviewing and editing the manuscript, providing insights into the validation process, and ensuring the accuracy and integrity of the research findings. **Peng Shyan Wong and Irene**

https://doi.org/10.24191/esteem.v20iMarch.616.g534

Looi: Collaborators from Kementerian Kesihatan Malaysia, responsible for assisting in the collection of samples from dengue patients. Their assistance in sample collection was integral to the success of the study.

9. REFERENCES

[1] W. H. Organization, *Dengue guidelines for diagnosis, treatment, prevention and control: new edition*, World Health Organization, 2009.

[2] S. Alcon, A. Talarmin, M. Debruyne, A. Falconar, V. Deubel, and M. Flamand, "Enzyme-linked immunosorbent assay specific to Dengue virus type 1 nonstructural protein NS1 reveals circulation of the antigen in the blood during the acute phase of disease in patients experiencing primary or secondary infections," *J Clin Microbiol*, vol. 40, no. 2, pp. 376–381, 2002. Available: https://doi.org/10.1128/jcm.40.02.376-381.2002

[3] D. H. Libraty et al., "High circulating levels of the dengue virus nonstructural protein NS1 early in dengue illness correlate with the development of dengue hemorrhagic fever," *J Infect Dis*, vol. 186, no. 8, pp. 1165–1168, 2002. Available: https://doi.org/10.1086/343813

[4] K. L. Anders et al., "An evaluation of dried blood spots and oral swabs as alternative specimens for the diagnosis of dengue and screening for past dengue virus exposure," *Am J Trop Med Hyg*, vol. 87, no. 1, p. 165, 2012. Available: https://doi.org/10.4269/ajtmh.2012.11-0713

[5] I. Gutsche et al., "Secreted dengue virus nonstructural protein NS1 is an atypical barrel-shaped high-density lipoprotein," *Proceedings of the National Academy of Sciences*, vol. 108, no. 19, pp. 8003–8008, 2011. Available: https://doi.org/10.1073/pnas.1017338108

[6] D. A. Muller et al., "Structure of the dengue virus glycoprotein non-structural protein 1 by electron microscopy and single-particle analysis," *Journal of General Virology*, vol. 93, no. 4, pp. 771–779, 2012. Available: https://doi.org/10.1099/vir.0.039321-0

[7] D. L. Akey, W. C. Brown, J. Jose, R. J. Kuhn, and J. L. Smith, "Structure-guided insights on the role of NS1 in flavivirus infection," *Bioessays*, vol. 37, no. 5, pp. 489–494, 2015. Available: https://doi.org/10.1002/bies.201400182

[8] V. Deubel, R. M. Kinney, and D. W. Trent, "Nucleotide sequence and deduced amino acid sequence of the nonstructural proteins of dengue type 2 virus, Jamaica genotype: comparative analysis of the full-length genome," *Virology*, vol. 165, no. 1, pp. 234–244, 1988. Available: https://doi.org/10.1016/0042-6822(88)90677-0

[9] W. H. Attatippaholkun, M. K. Attatippaholkun, A. Nisalak, D. W. Vaughn, and B. L. Innis, "Nucleotide sequence and deduced amino acid sequence of the nonstructural proteins of dengue type 3 virus, Bangkok genotype.," *Southeast Asian J Trop Med Public Health*, vol. 29, no. 2, pp. 361–366, 1998.

[10] P. Y. Yang, I. Kautner, C. L. Koh, and S. K. Lam, "Nucleotide and deduced amino acid sequences of genes encoding the structural and nonstructural NS1 proteins of a dengue-2 virus isolated in China," *Virus Genes*, vol. 8, no. 1, pp. 71–74, 1994. Available: https://doi.org/10.1007/BF01703603

[11] P. W. Mason, P. C. McAda, T. L. Mason, and M. J. Fournier, "Sequence of the dengue-1 virus genome in the region encoding the three structural proteins and the major nonstructural protein NS1," *Virology*, vol. 161, no. 1, pp. 262–267, 1987. Available: https://doi.org/10.1016/0042-6822(87)90196-6

[12] B. J. Blitvich, D. Scanlon, B. J. Shiell, J. S. Mackenzie, K. Pham, and R. A. Hall, "Determination of the intramolecular disulfide bond arrangement and biochemical identification of the glycosylation sites of the nonstructural protein NS1 of Murray Valley encephalitis virus.," *J Gen Virol*, vol. 82, no. Pt 9, pp. 2251–2256, 2001. Available: https://doi.org/10.1099/0022-1317-82-9-2251

https://doi.org/10.24191/esteem.v20iMarch.616.g534

[13] T. P. Wallis, C. Y. Huang, S. B. Nimkar, P. R. Young, and J. J. Gorman, "Determination of the disulfide bond arrangement of dengue virus NS1 protein," *Journal of Biological Chemistry*, vol. 279, no. 20, pp. 20729–20741, 2004. Available: https://doi.org/10.1074/jbc.M312907200

[14] M. Flamand, F. Megret, M. Mathieu, J. Lepault, F. A. Rey, and V. Deubel, "Dengue virus type 1 nonstructural glycoprotein NS1 is secreted from mammalian cells as a soluble hexamer in a glycosylation-dependent fashion," *J Virol*, vol. 73, no. 7, pp. 6104–6110, 1999. Available: https://doi.org/10.1128/jvi.73.7.6104-6110.1999

[15] D. L. Akey et al., "Flavivirus NS1 Structures Reveal Surfaces for Associations with Membranes and the Immune System," *Science* (1979), vol. 343, no. 6173, pp. 881–885, 2014. Available: https://www.science.org/doi/10.1126/science.1247749

[16] P. Avirutnan et al., "Vascular Leakage in Severe Dengue Virus Infections: A Potential Role for the Nonstructural Viral Protein NS1 and Complement," *J Infect Dis*, vol. 193, pp. 1078–1088, 2006. Available: https://doi.org/10.1086/500949

[17] J. M. Mackenzie, M. K. Jones, and P. R. Young, "Immunolocalization of the dengue virus nonstructural glycoprotein NS1 suggests a role in viral RNA replication.," *Virology*, vol. 220, no. 1, pp. 232–240, 1996. Available: https://doi.org/10.1006/viro.1996.0307

[18] M. Fleischmann, P. J. Hendra, and A. J. McQuillan, "Raman Spectra of Pyridine Adsorbed at a Silver Electrode," *Chem Phys Lett*, vol. 26, no. 2, 1974. Available: https://doi.org/10.1016/0009-2614(74)85388-1

[19] K. Kneipp et al., "Single molecule detection using surface-enhanced Raman scattering (SERS)," *Phys Rev Lett*, vol. 78, no. 9, pp. 1667, 1997. Available: https://doi.org/10.1103/PhysRevLett.78.1667

[20] S. Nie and S. R. Emory, "Probing Single Molecules and Single Nanoparticles by Surface-Enhanced Raman Scattering," *Science* (1979), pp. 1102–1106, 1997. Available: https://www.science.org/doi/10.1126/science.275.5303.1102

[21] Feng et al., "Surface-enhanced Raman spectroscopy of saliva proteins for the noninvasive differentiation of benign and malignant breast tumors," *Int J Nanomedicine*, pp. 537–547, 2015.

[22] J. C. Y. Kah et al., "Early diagnosis of oral cancer based on the surface plasmon resonance of gold nanoparticles," *Int J Nanomedicine*, vol. 2, no. 4, pp. 785, 2007. Available: https://doi.org/10.2147/IJN.S71811

[23] C. Anyu et al., "Detecting Narcotic Usage Using Surface-Enhanced Raman Spectroscopy on Saliva Samples," in *World Congress on Medical Physics and Biomedical Engineering* September 7-12, 2009, pp. 71. Available: https://doi.org/10.1007/978-3-642-03885-3_20

[24] Y. Wang et al., "A feasibility study of early detection of lung cancer by saliva test using Surface Enhanced Raman scattering," in *2012 5th International Conference on BioMedical Engineering and Informatics*, IEEE, 2012, pp. 135–139. Available: 10.1109/BMEI.2012.6513160

[25] E. Widjaja, W. Zheng, and Z. Huang, "Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines," *Int J Oncol*, vol. 32, no. 3, pp. 653–662, 2008.

[26] A. R. M. Radzol, K. Y. Lee, W. Mansor, and F. M. T. Tawi, "Signal processing for raman spectra for disease detection," *Int. J. Pharm. Pharm*. Sci, vol. 8, no. 6, pp. 4–10, 2016.

[27] M. Saleem, M. Bilal, S. Anwar, A. Rehman, and M. Ahmed, "Optical diagnosis of dengue virus infection in human blood serum using Raman spectroscopy," *Laser Phys Lett*, vol. 10, no. 3, p. 035602, 2013. Available: https://iopscience.iop.org/article/10.1088/1612-2011/10/3/035602

[28] M. Bilal et al., "Raman spectroscopy based discrimination of NS1 positive and negative dengue virus infected serum," *Laser Phys Lett*, vol. 13, no. 9, pp. 095603, 2016. Available: https://iopscience.iop.org/article/10.1088/1612-2011/13/9/095603

[29] S. Khan et al., "Raman spectroscopic analysis of dengue virus infection in human blood sera," *Optik* (Stuttg), vol. 127, no. 4, pp. 2086–2088, 2016. Available: https://doi.org/10.1016/j.ijleo.2015.11.060

https://doi.org/10.24191/esteem.v20iMarch.616.g534

[30] S. Khan, R. Ullah, A. Khan, N. Wahab, M. Bilal, and M. Ahmed, "Analysis of dengue infection based on Raman spectroscopy and support vector machine (SVM)," *Biomed Opt Express*, vol. 7, no. 6, pp. 2249–2256, 2016. Available: https://doi.org/10.1364/BOE.7.002249

[31] S. Khan et al., "Random forest-based evaluation of Raman spectroscopy for dengue fever analysis," *Appl Spectrosc*, vol. 71, no. 9, pp. 2111–2117, 2017.

[32] A. Amin, N. Ghouri, S. Ali, M. Ahmed, M. Saleem, and J. Qazi, "Identification of new spectral signatures associated with dengue virus infected sera," *Journal of Raman Spectroscopy*, vol. 48, no. 5, pp. 705–710, 2017. Available: https://doi.org/10.1177/0003702817695570

[33] T. Mahmood et al., "Raman spectral analysis for rapid screening of dengue infection," *Spectrochim Acta A Mol Biomol Spectrosc*, vol. 200, pp. 136–142, 2018. Available: https://doi.org/10.1016/j.saa.2018.04.018

[34] S. K. Gahlaut, D. Savargaonkar, C. Sharan, S. Yadav, P. Mishra, and J. P. Singh, "SERS platform for dengue diagnosis from clinical samples employing a hand held Raman spectrometer," *Anal Chem*, vol. 92, no. 3, pp. 2527–2534, 2020. Available: https://doi.org/10.1021/acs.analchem.9b04129

[35] I. T. Jolliffe, "Principal component analysis: a beginner's guide—I. Introduction and application," *Weather*, vol. 45, no. 10, pp. 375–382, 1990. Available: https://doi.org/10.1002/j.1477-8696.1990.tb05558.x

[36] V. Vapnik, "Support-vector networks," *Mach Learn*, vol. 20, pp. 273–297, 1995. Available: https://doi.org/10.1007/BF00994018

[37] M. Navazesh, "Methods for collecting saliva," *Ann N Y Acad Sci*, vol. 694, no. 1, pp. 72–77, 1993. Available: https://doi.org/10.1111/j.1749-6632.1993.tb18343.x

[38] N. H. Othman, K. Y. Lee, A. R. M. Radzol, and W. Mansor, "Optimized Raman Setting of Objective Lens, Laser Power and Integration Time for High and Low Concentration of Nonstructural Protein 1," *in EMBEC & NBC 2017: Joint Conference of the European Medical and Biological Engineering Conference (EMBEC) and the Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC)*, Tampere, Finland, June 2017, Springer, 2018, pp. 482–485. Available: https://doi.org/10.1007/978-981-10-5122-7_121

[39] A. R. M. Radzol, K. Y. Lee, W. Mansor, and A. Azman, "Optimization of Savitzky-Golay smoothing filter for salivary surface enhanced Raman spectra of non structural protein 1," *in TENCON 2014-2014 IEEE Region 10 Conference*, IEEE, 2014, pp. 1–6. Available: https://ieeexplore.ieee.org/document/7022409

[40] A. R. Radzol, K. Y. Lee, W. Mansor, and N. Saadun, "Optimized Automated Baseline Correction for NS1 Adulterated Salivary Raman Spectra," *Adv Sci Lett*, vol. 24, no. 2, pp. 1182–1186, 2018. Available: https://doi.org/10.1166/asl.2018.10712

[41] R. B. Cattell, "The scree test for the number of factors," *Multivariate Behav* Res, vol. 1, no. 2, pp. 245–276, 1966. Available: https://doi.org/10.1207/s15327906mbr0102_10

[42] S. Han, C. Qubo, and H. Meng, "Parameter selection in SVM with RBF kernel function," *in World Automation Congress 2012*, IEEE, 2012, pp. 1–4.

https://doi.org/10.24191/esteem.v20iMarch.616.g534