Analysis of dengue infection based on Raman spectroscopy and support vector machine (SVM)

Saranjam Khan,1,* Rahat Ullah,1 Asifullah Khan,2 Noorul Wahab,2 Muhammad Bilal,1 and Mushtaq Ahmed1

1Agri-Biophotonics Division, National Institute for Lasers and Optronics (NILOP), Nilore, Islamabad 45650, Pakistan
2Pattern Recognition Lab, DCIS, Pakistan Institutes of Engineering and Applied Sciences (PIEAS), Nilore, Islamabad 45650, Pakistan
*k.saranjam@yahoo.com

Abstract: The current study presents the use of Raman spectroscopy combined with support vector machine (SVM) for the classification of dengue suspected human blood sera. Raman spectra for 84 clinically dengue suspected patients acquired from Holy Family Hospital, Rawalpindi, Pakistan, have been used in this study. The spectral differences between dengue positive and normal sera have been exploited by using effective machine learning techniques. In this regard, SVM models built on the basis of three different kernel functions including Gaussian radial basis function (RBF), polynomial function and linear function have been employed to classify the human blood sera based on features obtained from Raman Spectra. The classification model have been evaluated with the 10-fold cross validation method. In the present study, the best performance has been achieved for the polynomial kernel of order 1. A diagnostic accuracy of about 85% with the precision of 90%, sensitivity of 73% and specificity of 93% has been achieved under these conditions.

OCIS codes: (300.0300) Spectroscopy; (300.6450) Spectroscopy, Raman.

References and links
1. S. M. Paranjape and E. Harris, “Control of dengue virus translation and replication,” Curr. Top. Microbiol. Immunol. 338, 15–34 (2010).
2. D. J. Gubler, “Dengue and dengue hemorrhagic fever,” Clin. Microbiol. Rev. 11(3), 480–496 (1998).
3. D. Guha-Sapir and B. Schimmer, “Dengue fever: new paradigms for a changing epidemiology,” Emerg. Themes Epidemiol. 2(1), 1 (2005).
4. R. Rico-Hesse, “Dengue virus markers of virulence and pathogenicity,” Future Virol. 4(6), 581–589 (2009).
5. N. E. Murray, M. B. Quam, and A. Wilder-Smith, “Epidemiology of dengue: past, present and future prospects,” Clin. Epidemiol. 5, 299–309 (2013).
6. World Health Organisation, “Dengue and severe dengue,” WHO Report, (2015).
7. World Health Organisation Report, “Guideline for diagnosis, treatment, prevention and control,”WHO Report, (2009).
8. V. Wiwanitkit, “Dengue fever: diagnosis and treatment,” Expert Rev. Anti Infect. Ther. 8(7), 841–845 (2010).
9. R. W. Peeling, H. Artsob, J. L. Pelegrino, P. Buchy, M. J. Cardosa, S. Devi, D. A. Enria, J. Farrar, D. J. Gubler, M. G. Guzman, S. B. Halstead, E. Hunsperger, S. Kliks, H. S. Margolis, C. M. Nathanson, V. C. Nguyen, N. Rizzo, S. Vázquez, and S. Yoksan, “Evaluation of diagnostic tests: dengue,” Nat. Rev. Microbiol. 8(12 Suppl), S30–S37 (2010).
10. J. J. Laserna, “An Introduction to Raman spectroscopy: introduction and basic principles,” Edu. Art. (2014).
11. A. Lorincz, D. Haddad, R. Naik, V. Naik, A. Fung, A. Cao, P. Manda, A. Panda, G. Auner, R. Rabah, S. E. Langenburg, and M. D. Klein, “Raman spectroscopy for neoplastic tissue differentiation: a pilot study,” J. Pediatr. Surg. 39(6), 953–956 (2004).
12. K. Kong, C. Kendall, N. Stone, and I. Notingher, “Raman spectroscopy for medical diagnostics—From in-vitro biofluid assays to in-vivo cancer detection,” Adv. Drug Deliv. Rev. 89, 121–134 (2015).
13. P. Crow, J. S. Uff, J. A. Farmer, M. F. Wright, and N. Stone, “The use of Raman spectroscopy to identify and characterize transitional cell carcinoma in vitro,” BJU Int. 93(9), 1232–1236 (2004).
14. A. Sikirzhychska, V. Sikirzhychski, and I. K. Lednev, “Raman spectroscopy coupled with advanced statistics for differentiating menstrual and peripheral blood,” J. Biophotonics 7(1-2), 59–67 (2014).
Dengue fever is a very complex disease because its clinical symptoms mimic other disease like malaria, and most of the time it is misdiagnosed. Likewise, there is no specific medicine for its treatment. Doctors can only recommend supporting medicines for controlling fever and advice intake of plenty of fluids. The prompt and accurate detection of the disease not only reduces its severity but also helps in proper clinical management. Ideally, a diagnostics test with high sensitivity and specificity is a desired task. Currently, many chemistry based laboratory tests are in practice for the confirmation and detection of dengue disease. These techniques are based on the detection of virus itself (cell culture methods), viral RNA detection (RT-PCR) and the detection of Immunoglobulin M (IgM), Immunoglobulin G (IgG) in the human blood serum (ELISA).

Virus and nucleic acid detection methods are quite effective in the initial five days after the onset of an infection. Virus isolation method is quite effective, but the non-availability of cell culture limits its uses for the clinical purpose. Transcriptase-polymerase chain reaction
(RT-PCR) has better sensitivity but it is not cost effective as it requires expensive equipment and chemical reagents. One cannot use it for the screening of large numbers of suspected patients. ELISA method is used for the detection of IgG and IgM anti-bodies produced in response to dengue infection. This test is effective only when done after five days of the onset of the infection, because IgM is produced after 5 days whereas IgG is produced after 10 days in the primary infection. This approach has better sensitivity and specificity but has problem of false positive results [7–9].

Interaction of light with matter gives rise to different types of spectroscopic techniques based on scattering, absorption, reflection and fluorescence. Raman spectroscopy is one such technique which is arising from inelastic scattering of laser light by the molecular vibration inside the sample. As a result, the scattered photons are emitted with the different frequency (energy). This difference in frequency between incident and emitted protons provides finger print about the rotational, vibrational and other low frequency transitions in molecule. Thus Raman spectrum, which is the plot of intensity as function of Raman shift, provides a lot of useful information for biochemical analysis and pathological investigation based on the molecular composition. The chemical composition of cells always changes because of the production of new cells as a result of any abnormality. So Raman spectroscopy is the best option to monitor these changes in real time non-invasively, thus enables the correct detection of the disease at early stage of the disease [10–13].

In the recent years, Raman spectroscopy combined with the diagnostic algorithm gained wide popularity in the field of medical diagnostics [14–17]. For this purpose, different types of multivariate statistical algorithm such as principal component analysis (PCA) [18], linear discriminate analysis (LDA) [19] and support vector machine (SVM) [20, 21] have been used for classification between and normal and malignant cells/tissues. This article, demonstrate the use of Raman spectroscopy and SVM for classification of dengue infected and normal healthy sera. Several different kernels are used for transforming the input space to higher dimensional feature space and thus enabling to analyze and classify the Raman Spectra. SVM is considered as an effective classification technique in the field of machine learning because it not only classifies the pattern in the data but also optimize the decision boundary [22, 23].

2. Materials and methods

2.1 Sample collection and preparation

The overall procedure of sample collection and serum extraction is same as previously mentioned in [24]. In total, Raman spectra of 84 samples of different ages and genders have been used in this study. Out of 84 samples, 31 were dengue positive and 53 were negative according to IgM findings of hospital. The blood samples of all patients have directly been acquired from the Dengue ward of Holy Family Hospital, Rawalpindi, Pakistan in autumn 2015. A quantity of about 3 ml was collected from each patient admitted in the hospital on the basis of the common dengue symptoms prior to any medication at different days. For the extraction of serum, all samples have been centrifuged at 3500 rpm for 10 minutes using Hittich Centrifuge D-7200. The extracted serum, from all blood samples, have been aliquoted in centrifuge tubes and stored at −16°C till use. The overall experimental procedure in this study has been carried out after obtaining written permission from the ethical committee of Rawalpindi medical college (RMC). The standard safety rules have been followed at each step from sample collection till acquisition of Raman spectra [25].

2.2 Raman spectrum acquisition

A quantity of about 15µl of each sample was initially put on the glass slide and dried at room temperature for about two hours. Raman spectrum for all samples have been acquired with Raman spectrometer (µRamboss DONGWOO OPRTON, South Korea). All Raman spectrometers have three main parts i.e. excitation source, sampling apparatus and detector.
Laser beam from diode laser emitting at 532 nm has been used as an excitation source. A microscope objective having a numerical aperture of 0.7 and magnification of 100X has been used both for focusing the laser light onto the sample and collection of Raman scattering light in backscattering configuration. A sketch of experimental setup is given in Fig. 1. Since Raman signal is normally very weak as compared to Rayleigh scattering, therefore an acquisition time of 15 seconds has been used for recording each spectrum. The spectrum from the sera samples have been recorded in the spectral range of 600 cm\(^{-1}\) to 1700 cm\(^{-1}\), as it contained the most useful information.

![Sketch of Raman setup](image)

**Fig. 1. A sketch of Raman setup.**

### 2.3 Data analysis and processing

Raman spectrum of biological samples is normally very complex and rich of biochemical information. Since in biological samples, there exist different types of macromolecules such as lipids, proteins, nucleic acids etc. The Raman spectrum of each of these molecules consists of numerous peaks. The visual assignment of any particular peaks to a specific biomolecule usually produces imprecision in the final result, because most of the time different biomolecules contribute to the same peak. In order to overcome this limitation of visual analysis, statistical methods are mostly used for the interpretation of Raman data of biological samples. With the statistical approach one can extract useful information from the data set by high lighting the similarities and differences. In this study, we are using different kernels of SVM for the analysis of dengue infection in the human serum samples, which has not been investigated, previously for the analysis of dengue infected human sera.

### 2.4 Support vector machine (SVM)

SVM is a powerful supervised learning algorithm that has many applications in the field of biophotonics, pattern recognition, and classification [21, 22, 26]. Initially, it was developed...
for two class classifications but one can also apply it to problems involving multiple classes by using one-versus-one and one-versus-all strategies. In case of binary problem, the basic aim of SVM is to define a boundary between clusters of data and to maximize the distance of the boundary line (or separating hyperplane in case of multiple dimensions) from data points lying closest to it. These closest data points, which lie on both sides of the line or hyperplane, are termed as support vectors. This leads to good generalization capability of the classifier that can potentially produce better results on unseen samples. In case of data which is linearly inseparable, mathematical functions (also called kernel functions) are used to transform the data to a higher dimensional space such that it may become linearly separable in the new space. For a linearly separable problem, the equation of a linear SVM can be written as [27]:

$$f(x) = \sum_{i=0}^{N} \alpha_i y_i x_i^T x + \beta_0$$

where $x_i$ is the instant with label $y_i$, $\alpha$ is Lagrange multiplier and $\beta_0$ is bias. For non-linearly separable problem, the above equation can be modified for kernel SVM as:

$$f(x) = \sum_{i=0}^{N} \alpha_i y_i K(x_i, x) x + \beta_0$$

Here $N$ represents the number of support vectors, whereas $K(x_i, x)$ is the kernel function.

3. Results and discussion

3.1 Raman spectral analysis

Figure 2 shows Raman spectra of dengue infected and normal human blood sera. For demonstration purpose, spectra of normal sera are shown in red color whereas the spectra of dengue infected sera are shown in light green color. In the spectrums of the normal blood serum, three prominent peaks exist at 1003, 1156 and 1516 cm$^{-1}$. All these peaks are highly reproducible exactly at the same Raman shift. In case of dengue infection these three intense peaks get suppressed as clearly visible in Fig. 2. Apart from the suppression of these peaks, additional Raman peaks appeared at 750, 850, 1450, 1660 cm$^{-1}$. The detail assignments of all these peaks are previously given in [24, 28].

![Fig. 2. Raman spectra of normal (red) and dengue infected (green) human blood sera.](image-url)
3.2 Model development

As mentioned earlier that dengue fever is a very complex disease. The visual inspection of Raman spectra of the blood serum samples of the suspected individual some times results in misdiagnosis of the disease. There is an extreme need for a diagnostic algorithm that correctly differentiates between infected samples from the normal one. For this purpose, SVM based classification system has been developed. The model was trained with immunoglobulin M (IgM) anti-bodies results of the hospital. In hospital three tests i.e. nonstructural protein 1 (NS1), IgG and IgM are performed in routine for all suspected patients. After an infection, different types of anti-bodies are produced at different time period in the human blood. In our case, all the samples have been collected at the time of admission before starting any medication, so initially there is more likelihood for IgM rather than IgG.

In primary dengue infection IgM antibodies are produced first and become detectable in about 80% patients at day 5 of the onset. In contrast to IgM, IgG anti-bodies are produced after 10 days in primary infection and exist for long time. In case of secondary infection, IgG anti-bodies are detectable even on the day first, because of their presence due to previous infection [9]. So the most obvious reason for considering IgM findings as gold standard is that the IgM results are comparatively more reliable.

For the statistical analysis of Raman spectra of human blood sera an SVM model was developed. As mentioned before the model was trained on Elisa capture IgM results of the hospital. The model takes the whole Raman spectrum and selects discernable features from the spectrum. Later on the model uses those features for predicting unknown samples. The developed model has been evaluated by using 10-fold cross validation approach. It basically divided the whole data set into10-subsets. Each time the model is trained on 9 subsets and tested on the remaining one. The overall process is repeated 10-times, to predict all the samples stepwise. The beauty of this method is that it does not care about how the data set are divided, because each data must come k-1 times in the training set and once in test set.
3.3 Model evaluation

The performance of an algorithm has been evaluated by computing confusion matrix under different conditions. Confusion matrix is basically NxN matrix, where each column represents an event in the predicted class, whereas each row represents an event in the actual class. The confusion matrix has been generated with three different kernel functions i.e. Gaussian radial basis, polynomial and linear. Initially, Gaussian radial basis kernel with the scaling factor of 1 and 2 has been applied. Later on the model has been tested with polynomial kernel function of order 1 and 2 and linear kernel, respectively.

Figure 3 shows the plot of first two principal components i.e. (PC1) and (PC2), which shows that the two classes are barely separable. However, by increasing the number of principal components to 5, an increase in the classification accuracy was observed. Since visualization in more than two dimensions becomes difficult, therefore, the separating boundary of a polynomial kernel of order 2 for one of the 10 folds using the first 2 PCA transformed features are used as shown in Fig. 4. For the purpose of clarity, the training and testing results are plotted separately where Fig. 4(a) shows the training samples and Fig. 4(b) shows test samples. It can be observed that the poly-SVM is able to separate the unseen (test) samples quite effectively by generating a nonlinear boundary between the infected and normal samples.

The overall results for the model with different kernel functions are given in Table 1. The best performance has been obtained with the polynomial kernel function of order 1. The performance of a model is usually evaluated in terms of accuracy, precision, sensitivity and specificity. Sensitivity correctly sorts out all patients with the disease, whereas specificity correctly identifies all patients who don’t have that disease [29]. A laboratory test with high specificity and sensitivity is usually desired, but rarely both of these conditions are met at the same time. The aforementioned four parameters for the current SVM model with polynomial kernel function of order 1 have been found 85%, 90%, 73% and 93%, respectively.

![Image](fig4.png)

**Fig. 4.** Polynomial order 2 separating boundaries for one of the 10 folds using the first 2 PCA transformed features; (a) training set (b) testing set.
Table 1. Evaluation of SVM combined with Raman spectroscopy for analysis of dengue infection

|                  | Polynomial Kernel | RBF Kernel | Linear Kernel |
|------------------|-------------------|------------|---------------|
|                  | Order-1 | Order-2 | rbf sigma-1 | rbf sigma-2 |               |
| True positive    | 28      | 24      | 25          | 26          | 27            |
| False positive   | 3       | 7       | 6           | 5           | 4             |
| True negative    | 43      | 44      | 43          | 43          | 42            |
| False negative   | 10      | 9       | 10          | 10          | 11            |
| Accuracy         | 85%     | 81%     | 81%         | 82%         | 82%           |
| Precision        | 90%     | 77%     | 81%         | 84%         | 87%           |
| Sensitivity      | 73%     | 72.7%   | 71%         | 72%         | 71%           |
| Specificity      | 93%     | 86%     | 87%         | 89%         | 91%           |
| f-measure        | 81%     | 75%     | 76%         | 78%         | 78%           |

4. Conclusions

This study demonstrates the use of Raman spectroscopy combined with SVM technique for the classification of the spectral data acquired from the sera of dengue suspected patients. Raman spectroscopy coupled with statistical tools has great potential to contribute significantly in the diagnosis and research of dengue fever in an effective way. There is also a great likelihood to use Raman spectroscopy combined with one of the existing methods for initial screening in order to increase the diagnostics efficiency. The results obtained are quite promising and interesting. The research work in our laboratory is still in progress striving for increasing sensitivity as well as specificity.

Acknowledgement

We are thankful to staff Nurse Uzma Shazia, Dengue Ward, Holy Family Hospital Rawalpind, and Muhammad Irfan and Mrs. Fatima Batool Scientific Assistants at Agri-Biophotonics division at National Institute for Lasers and optronics (NILOP) for assisting the present research work.