AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

Automated Detection of Non-Structural Protein 1 in Saliva from Raman Spectrum with Linear Discriminant Analysis

[1]F.M. Twon Tawi, [2]Khuan Y. Lee, [3]W. Mansor and [4]A.R. M. Radzol

[1]*Jabatan Kejuruteraan Elektrik, Politeknik Seberang Perai, 13500 Permatang Pauh, Malaysia*
[2]*Faculty of Electrical Engineering, Computational Physiologic Detection RIG, Brain & Neuro Science Communities of Research UniversitiTeknologi MARA, 40450 Shah Alam, Selangor DE, Malaysia*
[3]*Faculty of Electrical Engineering, Computational Physiologic Detection RIG, Brain &Neuro Science Communities of Research UniversitiTeknologi MARA, 40450 Shah Alam, Selangor DE, Malaysia*
[4]*Faculty of Electrical Engineering, UniversitiTeknologi MARA, 40450 Shah Alam, Selangor DE, Malaysia*

ARTICLE INFO

Article history:
Received 30 September 2014
Received in revised form
17 November 2014
Accepted 25 November 2014
Available online 13 December 2014

Keywords:
Surface Enhanced Raman Spectroscopy(SERS),
Non-Structural Protein 1 (NS1),
Linear Discriminant Linear Analysis (LDA)

ABSTRACT

Background: Non-Structural protein 1 (NS1)is considered as one of the biomarkers for early detection of disease transmitted by mosquitoes such as Japanese encephalitis (JE), Murray Valley encephalitis (MVE), West Nile encephalitis (WNE), dengue fever (DF), and yellow fever, which has terminal consequences. Previous studies found that NS1 is Raman active. Raman spectroscopy is a technique used to study the vibration and rotational modes of molecule. It produces unique spectrum, or molecular fingerprint, for each and every molecule from inelastic scattering of light. Nano-technology innovates Raman spectroscopy to produce an enhanced technique, the Surface Enhanced Raman Spectroscopy (SERS). SERS amplifies Raman signal from molecule many thousand folds with the use of nano-size noble metal known as substrate. Linear Discriminant Analysis (LDA) is a supervised statistical method suitable for classification of two or more classes of data. In our study, LDA is applied to detect the presence of NS1 in saliva from Raman spectra. Raman spectra of saliva and NS1 adulterated saliva samples are first analyzed using Surface Enhanced Raman Spectroscopy (SERS) with gold substrate. NS1 adulterated saliva samples at different concentrations were prepared. The region of interest (ROI) of these SERS spectra then served as inputs to the LDA algorithm. The performance [Accuracy Sensitivity Specificity] attained with the LDA classifier is reported to be [85.7% 93.75% 100%], with the (training:testing) dataset ratio as (80%:20%) of the spectra data. The execution time for a matrix of [80x16] input data is less than 15 seconds.

To Cite This Article: F.M. Twon Tawi, Khuan Y. Lee, W. Mansor, A.R.M. Radzol., Automated Detection of Non-Structural Protein 1 in Saliva from Raman Spectrum with Linear Discriminant Analysis, ***Aust. J. Basic & Appl. Sci.,*** *8(24): 27-32, 2014*

INTRODUCTION

Diseases such as Japanese encephalitis (JE), Murray Valley encephalitis (MVE), West Nile encephalitis (WNE), dengue fever (DF), and Yellow fever (YF)(Lindenbach & Rice, 2003) are caused by viruses transmitted to human by mosquitoes, with the risk of loss of life. These viruses are represented as a single strand of ribonucleic acid (RNA) of *flavivirus* from *flaviviridae* family(Maeda & Maeda, 2013). Upon infection, the RNA of the virus is translated and duplicated producing new viral strands. It is believed that the replication process of viral RNA is regulated by NS1(Alcon *et al*., 2002). This makes NS1 a viable molecular marker for diseases mentioned above.

NS1 protein is one of the encoded non-structural proteins in the *flavivirus* genome. It circulates in the blood serum of DF (one of the *flavivirus* infected diseases) patients for the first few days after onset of symptoms of DF(Alcon, *et al*., 2002; Datta & Wattal, 2010). That explains it being acknowledged as a biomarker for early diagnosis of DF infection(Kassim, Izati, TgRogayah, Apandi, & Saat, 2011; Zainah *et al*., 2009). In fact, there exist currently a few commercial blood serum based NS1 testkits with disposable strips for early screening(Kassim, Izati, TgRogayah, Apandi, & Saat, 2011; Zainah *et al*., 2009), which are rarely used by physicians due to their cost. Enzyme-linked immunosorbent assay (ELISA) on serum Immunoglobulin M (IgM)and/orImmunoglobulin G (IgG) has long been established and practiced as the diagnosis for DF, according to guidelines for dengue by the World Health Organization(WHO)(Hadinegoro, 2012). A recent attempt to replace the serum based to salivary based test for DF NS1-ELISA has also been reported(Anders *et*

Corresponding Author: Faizal Bin Mohamad TwonTawi, JKE, Politeknik Seberang Perai, Jalan Permatang Pauh, 13500, Permatang Pauh, Pulau Pinang Malaysia.
Ph: +06012460366, E-mail: fmtt@psp.edu.my

al., 2012). However, the sensitivity attained, 64.7%, is low, as a result of lower concentration NS1 molecules present in saliva than blood. As such, detection of NS1 in saliva requires a more sensitive technique, for a more reliable detection.

Surface Enhanced Raman Spectroscopy (SERS) was accidentally stumbled upon by Martin Fleischmann from Department of Chemistry, University of Southampton(Fleischmann, Hendra, & McQuillan, 1974). He discovered that the intensity of Raman spectrum multiplies by many thousands folds, when the molecule of analyte is made to absorb onto noble metal, known as substrate. The impact of this effect mainly depends on the choice of substrate, amongst other minor elements. This technique claims capable of detection up to a single molecule, given a good match of substrate and analyte(Katrin Kneipp *et al*., 1997; Nie & Emory, 1997). With the additional advantage of simple sample preparation method for analysis, this technique is fast gaining the interest of researchers who make use of it in numerous applications, food technology (Cheng *et al*., 2010), drug screening(Ryder, 2005), photographic sciences(K. Kneipp, 1990) and biomolecules(Combs *et al*., 2011; Datta & Wattal, 2010).

Another appealing feature of SERS is that its spectral signals are amenable to signal processing for automated classification and detection. Of feature extraction methods for statistical pattern recognition, Linear Discriminant Analysis (LDA) is found popularly used with SERS spectra. The LDA algorithm is formulated by Sir Ronald Aylmer Fisher(Fisher, 1936). It selects features that are most effective for class separability but not for class representation. It separates between data classes using projection lines, which can be linear or quadratic. It is a preferred technique when information on the data class is available. In the field of biomedical engineering, its application on Raman spectra has been found with detection of cancer such as lung cancer(Li, Yang, & Lin, 2012; Yan *et al*., 2010), gastric cancer(Feng *et al*., 2011) and colorectal cancer(Lin *et al*., 2011); prostatic adenocarcinoma cell in cell line study(Crow *et al*., 2005); micro-organism types in microbiology(Mobili *et al*., 2010). Yet, the use of LDA to discriminate between saliva samples with and without NS1 from their Raman spectra has not been explored.

Our work here is the first attempt to detect for NS1 molecule in saliva from their SERS spectra. LDA is used to classify between saliva samples with and without NS1 after SERS analysis. Section II explains the background and theories on LDA. Section III details the sample preparation method and Raman analysis. Section IV presents and discusses results from classification by LDA.

1. Linear Discriminant Analysis:

Linear Discriminant Analysis is a supervised algorithm with the assumption that the covariance matrix for each class is identical. It can be used to analyze scatter matrix data and categorize between two or more groups of data after an optimal projection line is found (Mansor, Syam, Rejab, & Syam b, 2012), using procedure depicted in the following flowchart.

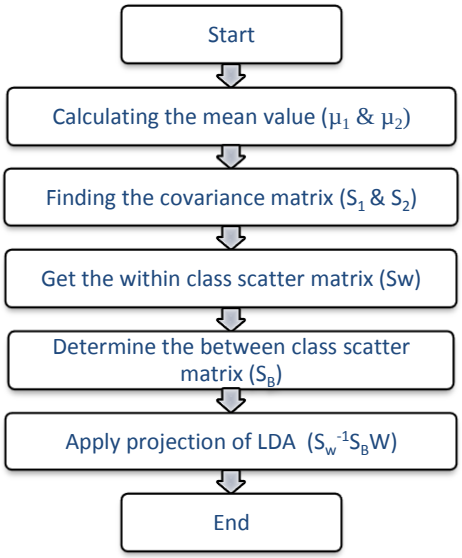


Fig. 1: Linear Discriminant Analysis Algorithm.

Say, there are m-dimensional data to be classified with the LDA algorithm. N1 equals $\{z^1,z^2,\ldots z^N\}$ and N2 equals $\{z^1,z^2,\ldots z^N\}$ are assigned to group ω_1 and ω_2 respectively. The first step is to calculate the mean of classes for both groups, using equation (1) and (2).

$$\mu_1 = \frac{1}{N_1}\sum_{x\in\omega_1} z \quad (1)$$

$$\mu_2 = \frac{1}{N_1}\sum_{x\in\omega_2} z \quad (2)$$

Next, the covariance matrix for both of the groups is computed using the following equations,

$$S_1 = \sum_{x\in\omega_1}(z-\mu_1)(z-\mu_1)^T \quad (3)$$

$$S_2 = \sum_{x\in\omega_2}(z-\mu_2)(z-\mu_2)^T \quad (4)$$

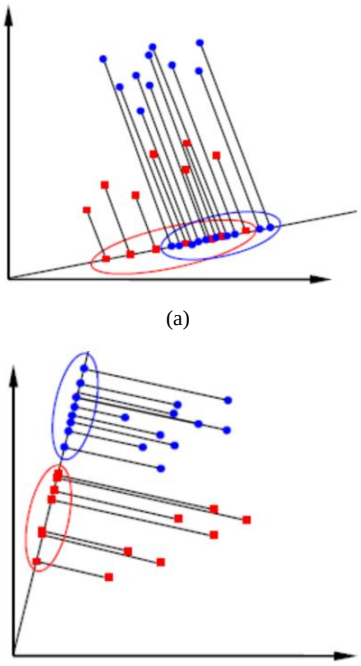
Then, equation (5) and (6) are applied to obtain the class scatter matrix (S_w) and between-class scatter matrix (S_b) respectively. The measure of difference between class, which is encoded in the between-class scatter matrix, carries the meaning of normalization by a measure of the within-class scatter matrix.

$$S_w = S_1 + S_2 \quad (5)$$

$$S_B = (\mu_1-\mu_2)(\mu_1-\mu_2)^T \quad (6)$$

The final step projects data onto the chosen projection line by solving the generalized eigenvalue problem, represented by equation (7). If the solution yields two values for eigenvalue λ, the maximum of the two is selected to obtain the optimal projection line as illustrated in Figure 2.

$$S_w^{-1}S_B w = \lambda w \quad (7)$$



(a)

(b)

Fig. 2: LDA algorithm (a) Data mixed as projected onto a projection line, (b) Data separate as projected onto an optimal projection line.

3.Methodology:

Non-structural glycoprotein 1 (NS1) used in this study is purchased from Abcam. It is received in a liquid form at 90% purity and 1000ppm concentration. It is diluted to 100ppm, 50ppm, 25ppm, 20ppm and 10ppm using phosphate buffer saline (PBS) of pH7.The NS1 adulterated saliva samples are prepared by mixing saliva from healthy normal subjects with NS1 at different concentrations. The NS1 adulterated saliva sample is prepared by mixing 30µL of diluted NS1 with 30µL of saliva supernatant. 10µL of the mixture is pipetted onto gold coated slide and left dried prior to Raman analysis to obtain the SERS spectra.

The SERS spectra were acquired using a Raman Station™ 400 Dispersive Raman Spectrometer from Perkin Elmer which employs a high sensitivity open electrode CCD detector with excitation source near infrared 785nm laser and minimum laser spot size of 100 microns. The Raman Spectrometer was operated at approximately 100mW, 100% of full power and as exposure time of 20 seconds, unless specified otherwise. Cosmic ray removal mode with median filter and baseline correction mode was enabled for cleaner spectra. The measurement was obtained with a (2 by 4) mapping and repeated twice at every concentration. In total, 40 spectra from NS1 adulterated saliva samples at five different concentrations were obtained, while 8 spectra were obtained from saliva of normal healthy subjects. Each spectrum contained 902 features of Raman shift. All the spectra at the same concentrations are averaged before classification.

RESULTS AND DISCUSSIONS

This Section first discusses results from Raman analysis of spectra of saliva and NS1 adulterated saliva samples. This is then ensued with data preparation for input to LDA classifier. The performance of the LDA classifier is examined at different ratios of (training:testing) datasets.

A. Raman Spectra of Saliva and NS1 Adulterated Saliva:

Figure 3 shows the Raman spectra of saliva (red) and NS1 adulterated saliva sample (black) at 10ppm. The Raman spectra for NS1 adulterated saliva sample at other concentrations look similar, only differ in fingerprint peak magnitude. The molecular fingerprint of NS1 (Radzol, Lee, & Mansor, 2013b) can be observed to stand out from the background.

Since saliva constitutes more than 90% of water, it does not exhibit any significant peak in the spectral region of 200cm^{-1} to 2000cm^{-1}. The broad peaks exhibited on the saliva spectra are found to originate from the blank gold coated slide, the substrate. The spectrum of blank gold coated slide as background signal for removal so as not to interfere with the spectra of NS1 has also been discussed(Radzol, Lee, Mansor, &Tawi, 2013).

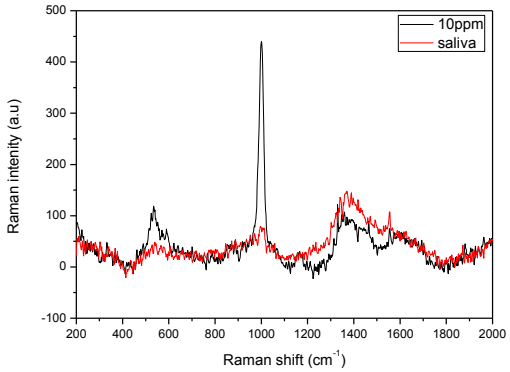


Fig. 2: Raman spectra of saliva (red) and NS1-saliva mixture (black).

The prominent peak at around 1000cm^{-1} corresponds to the breathing mode of benzene ring structure of aromatic amino acid contained in NS1(Radzol, Lee, *et al*., 2013b). The intensity of this peak is linearly proportional to the concentration of NS1 in the adulterated mixtures(Radzol, Lee, & Mansor, 2013a). As a consequence, the task of looking for the fingerprint peak becomes challenging as concentration of NS1 becomes very low, when the fingerprint peak would be engulfed by noise. In addition, peak intensity is also liable to contamination due to surface plasmon resonance effect, as the NS1 molecules are absorbed onto the surface of substrate.

B. Linear Discriminant Analysis:

The raw spectra contained 902 features for Raman shift between 200cm^{-1} to 2000cm^{-1}. For efficiency use of computation resources, the raw spectra is narrowed to a region of interest (ROI), centring about the molecular fingerprint peak at 1000cm^{-1}. The ROI is marked from 990cm^{-1} to 1020cm^{-1} containing only 16 features, a reduction of 882 features, or 98% from the raw spectra. The ROI matrix of [80x16] is then randomized before being served as inputs to LDA.

LDA separates the ROI matrix of data into training and testing datasets. The performance of LDA classifier at different ratios of (training:testing) dataset is evaluated in terms of sensitivity, specificity and accuracy. Their performance is summarized in Table 1.

With a specificity of 100% for all the ratios of (training:testing) datasets, it means the LDA classifier is capable of recognising all the samples without NS1. None of the only saliva sample is mistaken as adulterated with NS1.

It is found that the highest accuracy and sensitivity (93.75% and 87.50% respectively) are attained when 80% of the data is used for training and 20% for testing. It is interesting to find that the Pareto principle of 80-20 rule applies here too.

Table 1: Result after applying the LDA algorithm.

	Train (%)	*Test (%)*	*Train (%)*	*Test (%)*	*Train (%)*	*Test (%)*	*Train (%)*	*Test(%)*
	60	40	*70*	30	80	20	90	10
Accuracy	95.83	75.00	*92.86*	83.33	95.31	93.75	90.28	87.50
Sensitivity	91.30	52.94	*85.19*	69.23	90.63	87.50	81.58	50.00
Specificity	100.00	100.00	*100.00*	100.00	100.00	100.00	100.00	100.00

4. Conclusion:

This study investigates the performance of LDA classifier in detecting NS1 in saliva samples from Raman spectra. It is found that the best performance LDA classifier is 85.7% of accuracy, 93.75% of sensitivity and specificity of 100%. This is achieved by limiting the analysis to a region of interest, centring about the molecular fingerprint of NS1 at 1000cm^{-1}, giving an input feature matrix of [80X16], at a (training:testing) dataset ratio of (80%:20%). The execution time in total is less than 15 seconds.

REFERENCES

Alcon, S., A. Talarmin, M. Debruyne, A. Falconar, V. Deubel and M. Flamand, 2002. 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. Journal of Clinical Microbiology, 40(2): 376-381.

Anders, K.L., N.M. Nguyet, N.T.H. Quyen, T.V. Ngoc, T.V. Tram, T.T. Gan and C.P. Simmons, 2012. An evaluation of dried blood spots and oral swabs as alternative specimens for the diagnosis of dengue and screening for past dengue virus exposure. American Journal of Tropical Medicine and Hygiene, 87(1): 165-170.

Cheng, Y., Y. Dong, J. Wu, X. Yang, H. Bai, H. Zheng, M. Li, 2010. Screening melamine adulterant in milk powder with laser Raman spectrometry. Journal of Food Composition and Analysis, 23(2): 199-202. doi: 10.1016/j.jfca.2009.08.006.

Combs, Z.A., S. Chang, T. Clark, S. Singamaneni, K.D. Anderson and V.V. Tsukruk, 2011. Label-free raman mapping of surface distribution of protein A and IgG biomolecules. Langmuir, 27(6): 3198-3205.

Crow, P., B. Barrass, C. Kendall, M. Hart-Prieto, M. Wright, R. Persad, N. Stone, 2005. The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines. British Journal of Cancer, 92(12): 2166-2170.

Datta, S. and C. Wattal, 2010. Dengue NS1 antigen detection: A useful tool in early diagnosis of dengue virus infection. Indian Journal of Medical Microbiology, 28(2): 107-110.

Feng, S., R. Chen, J. Lin, J. Pan, Y. Wu, Y. Li and H. Zeng, 2011. Gastric cancer detection based on blood plasma surface-enhanced Raman spectroscopy excited by polarized laser light. Biosensors and Bioelectronics, 26(7): 3167-3174. doi: 10.1016/j.bios.2010.12.020

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7(2): 179-188. doi: 10.1111/j.1469-1809.1936.tb02137.x.

Fleischmann, M., P.J. Hendra and A.J. McQuillan, 1974. Raman spectra of pyridine adsorbed at a silver electrode. Chemical Physics Letters, 26(2): 163-166.

Hadinegoro, S.R.S., 2012. The revised WHO dengue case classification: Does the system need to be modified? Paediatrics and International Child Health, 32(SUPP1): 33-38.

Kassim, F.M., M.N. Izati, T. TgRogayah, Y.M. Apandi and Z. Saat, 2011. Use of dengue NS1 antigen for early diagnosis of dengue virus infection. Southeast Asian Journal of Tropical Medicine and Public Health, 42(3): 562-569.

Kneipp, K., 1990. Application of SERS in photographic sciences. Journal of Molecular Structure, 218: 357-362. doi: 10.1016/0022-2860(90)80293-s.

Kneipp, K., Y. Wang, H. Kneipp, L.T. Perelman, I. Itzkan, R.R. Dasari and M.S. Feld, 1997. Single Molecule Detection Using Surface-Enhanced Raman Scattering (SERS). Physical Review Letters, 78(9): 1667-1670.

Li, X., T. Yang and J. Lin, 2012. Spectral analysis of human saliva for detection of lung cancer using surface-enhanced Raman spectroscopy. Journal of Biomedical Optics, 17(3): 0370031-0370035. doi: 10.1117/1.jbo.17.3.037003

Lin, D., S. Feng, J. Pan, Y. Chen, J. Lin, G. Chen and R. Chen, 2011. Colorectal cancer detection by gold nanoparticle based surface-enhanced Raman spectroscopy of blood serum and statistical analysis. Optics Express, 19(14): 13565-13577.

Lindenbach, B.D. and C.M. Rice, 2003. Molecular biology of flaviviruses, 59: 23-61.

Maeda, A. and J. Maeda, 2013. Review of diagnostic plaque reduction neutralization tests for flavivirus infection. Veterinary Journal, 195(1): 33-40.

Mobili, P., A. Londero, G. De Antoni, A. Gómez-Zavaglia, C. Araujo-Andrade, H. Ávila-Donoso and C. Frausto-Reyes, 2010. Multivariate analysis of Raman spectra applied to microbiology: Discrimination of microorganisms at the species level. Revista Mexicana de Fisica, 56(5): 378-385.

Nie, S. and S.R. Emory, 1997. Probing Single Molecules and Single Nanoparticles by Surface-Enhanced Raman Scattering. Science, 275(5303): 1102-1106. doi: 10.1126/science.275.5303.1102.

Radzol, A.R.M., Lee, Y. Khuan and W. Mansor, 2013a. Nonstructural protein 1 characteristic peak from NS1-saliva mixture with Surface-Enhanced Raman spectroscopy.

Radzol, A.R.M., Lee, Y. Khuan and W. Mansor, 2013b. Raman molecular fingerprint of non-structural protein 1 in phosphate buffer saline with gold substrate.

Radzol, A.R.M., Lee, Y. Khuan, W. Mansor and F.M.T. Tawi, 2013. Automated extraction of principal components of non-structural protein 1 from SERS spectrum.

Ryder, A.G., 2005. Surface enhanced Raman scattering for narcotic detection and applications to chemical biology. Current Opinion in Chemical Biology, 9(5): 489-493.

Twon Tawi, F.M., Lee, Y. Khuan, W. Mansor and A.R.M. Radzol, 2013b. Automatic Non-Structural Protien 1 recognition based on LDA classifier. Paper presented at the Proceedings - 2013 IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2013.

Yan, W., S. Shuang, Q. Dian, C. Anyu, C. Zijian, Y. Yulu and L. Chunwei, 2010. Preliminary study on early detection technology of lung cancer based on surface-enhanced Raman spectroscopy. Paper presented at the Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on.

Zainah, S., A.H.A. Wahab, M. Mariam, M.K. Fauziah, A.H. Khairul, I. Roslina and K.B. Chua, 2009. Performance of a commercial rapid dengue NS1 antigen immunochromatography test with reference to dengue NS1 antigen-capture ELISA. Journal of Virological Methods, 155(2): 157-160. doi: 10.1016/j.jviromet.2008.10.016.