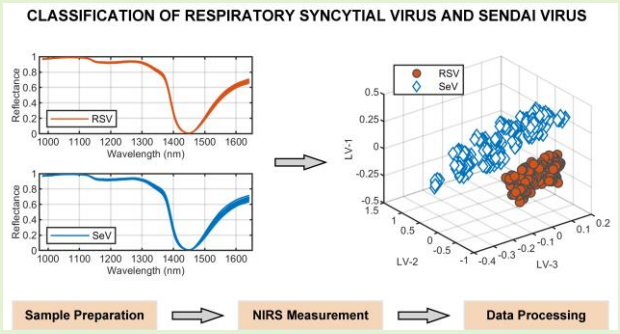
Classification of Respiratory Syncytial Virus and Sendai Virus Using Portable Near-infrared Spectroscopy and Chemometrics

Weiran Song, Hui Wang, Enayetur Rahman, Judit Barabas, Jiandong Huang, James McLaughlin, Chris Nugent, Paul Maguire

***Abstract*—There is evidence that it may be possible to detect viruses and viral infection optically using techniques such as Raman and infra-red (IR) spectroscopy and hence open the possibility of rapid identification of infected patients. However, high-resolution Raman and IR spectroscopy instruments are laboratory-based and require skilled operators. The use of low-cost portable or field-deployable instruments employing similar optical approaches would be highly advantageous. In this work, we use chemometrics applied to low-resolution near-infrared (NIR) reflectance/absorbance spectra to investigate the potential for simple low-cost virus detection suitable for widespread societal deployment. We present the combination of near-infrared spectroscopy and chemometrics to distinguish two respiratory viruses, *respiratory syncytial virus* (RSV), the principal cause of severe lower respiratory tract infections in infants worldwide, and *Sendai virus* (SeV), a prototypic paramyxovirus. Using a low-cost and portable spectrometer, three sets of RSV and SeV spectra, dispersed in phosphate-buffered saline (PBS) medium or Dulbecco's modified eagle medium (DMEM), were collected in long-term and short-term experiments. The spectra were pre-processed, and analysed by partial least squares discriminant analysis (PLS-DA) for virus type and concentration classification. Moreover, the virus type/concentration separability was visualized in a low-dimensional space through data projection. The highest virus type classification accuracy obtained in PBS and DMEM is 85.8% and 99.7%, respectively. The results demonstrate the feasibility of using portable NIR spectroscopy as a valuable tool for rapid, on-site and low-cost virus pre-screening for RSV and SeV with the further possibility of extending this to other respiratory viruses such as SARS-CoV-2.**



***Index Terms*— Chemometrics, classification, near-infrared spectroscopy, partial least squares discriminant analysis, respiratory syncytial virus, Sendia virus.**

I. Introduction

Viral infection can often present a serious risk to population health, and therefore early warning and continuous monitoring represent an important element of disease mitigation. The primary laboratory standard for the detection of viral infection relies on quantitative reverse transcription-polymerase chain reaction (RT-qPCR), which amplifies and detects RNA sequences. Although RT-qPCR can provide an accurate diagnosis, it is limited by cost and processing time, and is not suitable for real-time and on-site virus detection across a wide range of community and healthcare settings. Considerable effort is being devoted to exploring alternative detection approaches in point-of-care technologies that attempt to couple simplicity, speed and low cost with the required accuracy in terms of sensitivity and selectivity. Examples include electrochemical biosensors [1][2], immunoassay-based tests [3][4] and spectroscopic assays [5][6][7][8].

Optical and infrared spectroscopy represents powerful tools for the analysis of chemical species and the detection of trace inorganic, organic and biological components [9]. Combined with computational analysis, different spectroscopic techniques can provide spectral fingerprint information of biomarkers with the advantages of being rapid, non-destructive and requiring little or no sample preparation. Spectroscopic techniques such as surface-enhanced Raman spectroscopy (SERS), attenuated total

Manuscript received XXX X, XXXX; revised XXX X, XXXX; accepted XXX X, XXXX. Date of publication XXX X, XXXX; date of current version XXX X, XXXX. This work is funded by UK EPSRC under UKRI COVID-19 Call (EP/V026488/1, Virus Identification via Portable InfraRed Spectroscopy). The associate editor coordinating the review of this article and approving it for publication was XXXXX.(*Corresponding author: Hui Wang.*)

Weiran Song, Hui Wang, Enayetur Rahman, Jiandong Huang, James McLaughlin, Chris Nugent, Paul Maguire are with Ulster University, Jordanstown, BT37 0QB, UK (e-mail: song-w@ulster.ac.uk, h.wang@ulster.ac.uk, e.rahman@ulster.ac.uk, jd.huang@ulster.ac.uk, jad.mclaughlin@ulster.ac.uk, cd.nugent@ulster.ac.uk, pd.maguire@ulster.ac.uk).

Judit Barabas is with Wellcome-Wolfson Institute for Experimental Medicine, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, BT7 1NN, UK (e-mail: Judit.Barabas@qub.ac.uk).

reflection Fourier-transform infrared (ATR-FTIR) spectroscopy and near-infrared spectroscopy (NIRS) have been reported as effective tools in virology studies [10]. Investigated tasks include the detection of influenza virus, respiratory syncytial virus (RSV), human immunodeficiency virus type-1 (HIV-1) and hepatitis C virus [10]. For influenza virus detection in nasal fluid, Sakudo et al. [11] combined visible and near-infrared spectroscopy (Vis-NIRS) and soft independent modelling of class analogy (SIMCA). The identification accuracies of non-influenza and influenza patients were 96.7% and 100%, respectively. Lim et al. [12] proposed a label-free method to distinguish different types of influenza viruses grown in cells using SERS and principal component analysis (PCA). The fingerprint related to the virus can be effectively extracted from the spectral peaks based on the PCA loadings. Similarly, for RSV identification, spectra measured by SERS can be unambiguously separated using PCA and hierarchical cluster analysis (HCA) [13]. In fact, in addition, these techniques can present the important wavelengths related to the biomarkers associated with HIV-1 and hepatitis B viral infection [14][15].

Recently the use of attenuated ATR-FTIR, surface-enhanced infrared absorption spectroscopy (SEIRA) and Raman spectroscopy have been investigated to detect SARS-CoV-2 infections. Barauna et al. [6] obtained ATR-FTIR spectra of contrived saliva samples spiked with inactivated γ-irradiated SARS-CoV-2 virus particles. A genetic algorithm-linear discriminant analysis (GA-LDA) model was constructed based on 100 training samples and achieved 95% sensitivity and 89% specificity in validation (61 negatives and 20 positives). Yao et al. [5] used SEIRA spectroscopy and PCA to capture the differences between infected and control samples, yielding 89.47% sensitivity and 87.5% specificity. Carlomagno et al. [7] combined Raman spectroscopy and convolutional neural network (CNN) to analyse patient saliva, and the accuracy was 89–92%. In addition, the infrared based saliva screening test based on partial least squares discriminant analysis (PLS-DA) achieved the same level of accuracy (93% sensitivity and 82% specificity) [8].

Although these spectroscopic techniques may achieve the accuracy required for viral detection, high-resolution laboratory-based instruments and skilled personnel are required. Early warning and routine community-based monitoring across the wide range of healthcare systems cannot readily avail of laboratory spectroscopy facilities to the required degree. Relatively low-cost portable NIR systems have become available. However, they suffer from much higher noise levels and lower resolution than standard systems [16]. Therefore, the application of machine learning techniques to interrogate these low-quality spectra and compensate for noise, variability and low resolution becomes essential, if effective portable low-cost systems are to be developed, which would offer the additional advantage of inherent user simplicity since the skilled decision process can be automatically transferred from the local collection site to remote centralized high-performance computation.

Over the last decade, the miniaturisation and on-site portability of benchtop spectrometers have shown great prospects in many fields, such as food safety [17], materials analysis [18] and pharmacology [19]. The use of portable NIRS spectra from both solid organic surfaces and liquids coupled with custom machine-learning algorithms to investigate the classification/regression of target components under different conditions has previously been demonstrated [20][21]. However, the application of portable spectrometers to virology has only recently been studied. Jian et al. [22] developed a spectrometer based on sunlight and smartphone to detect avian influenza virus H7N9 and porcine circovirus type 2 antibodies. The coefficients of determination were 0.959 and 0.969, respectively, which were comparable to those of a commercial microplate reader. A major challenge with the detection of trace elements in solution is the large matrix component present in the signal, particularly for aqueous solutions. Typically, the signal-to-noise ratio (i.e., the ratio of analyte signal to that of the matrix) in this scenario is < 0.01. Due to the narrow range of wavelengths and low spectral resolutions, the data obtained from the miniaturized spectrometers is considered low quality. This, coupled with fluctuating experimental parameters, poses a serious challenge to the accuracy and reliability of the qualitative and quantitative analysis. To tackle this issue, data pre-processing is essential to remove unwanted variations, such as instrumental and experimental artefacts [23]. Moreover, the use of chemometrics and machine learning methods efficiently improves the accuracy and reliability of data analysis [24][25]. Therefore, the application of chemometrics and machine learning is a promising method that can improve the accuracy of processing low-quality spectra obtained from portable spectroscopy systems.

In this work, we investigate the feasibility of using portable NIRS combined with chemometrics to classify respiratory syncytial virus (RSV) and Sendai virus (SeV). RSV is the principal cause of severe lower respiratory tract infections in young infants worldwide, and SeV is a prototypic paramyxovirus. Three sets of spectra, dispersed in phosphate-buffered saline (PBS) medium or Dulbecco's modified eagle medium (DMEM), were collected in long-term and short-term experiments. Although diagnostic screening requires a binary positive/negative output, the viral load of a positive case can vary significantly, and therefore a range of different concentrations was measured, for each virus. The PLS-DA was used to model the relationship between the NIR spectra and the virus types/concentrations, and the virus type/concentration separability was visualised in a low-dimensional space through data projection. Furthermore, the important variables of the RSV and SeV classification were identified using competitive adaptive reweighted sampling (CARS). The experimental results demonstrate that the portable NIRS system combined with chemometrics can provide a potential solution for simple, rapid and low-cost pre-screening of RSV and SeV classification.

II. Materials and Methods

A. Sample Preparation

Three groups of RSV and SeV samples were prepared in a class II virus laboratory in Queen's University Belfast. The first two groups consist of virus samples measured in four sessions on different dates (2020-10-15, 2020-10-29, 2020-11-20, and 2020-12-10) in PBS and DMEM, respectively. The third group has virus samples measured in five sessions on the same date (2021-02-04), and the samples were measured in PBS. The three groups of samples are named as long-term PBS (L-PBS), long-term DMEM (L-DMEM) and short-term PBS (S-PBS) for simplicity. A total of 44, 44 and 54 virus samples were prepared for the three groups, respectively. The RSV and SeV ratios in

TABLE I
THE INFORMATION OF THE THREE GROUPS OF RSV AND SEV SAMPLES

	Samples	RSV/SeV	Spectra	Medium	Sessions	Date
L-PBS	44	16:28	440	PBS	4	2020/10/15, 2020/10/29, 2020/11/20, 2020/12/10
L-DMEM	44	16:28	440	DMEM	4	2020/10/15, 2020/10/29, 2020/11/20, 2020/12/10
S-PBS	54	19:35	540	PBS	5	2021/02/04

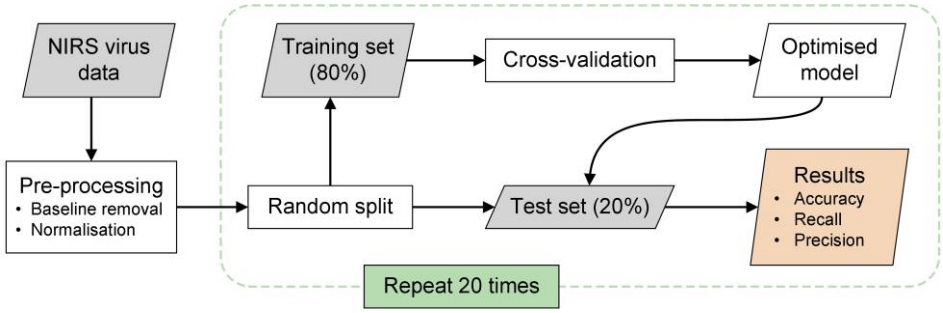


Fig. 1. Schematic diagram of NIRS virus data analysis.

the three groups were 16:28, 16:28 and 19:35 respectively. The information of the virus samples is summarised in Table I.

Virus stocks (RSV titre: $2.8\text{x}10^7$ $TCID_{50}$/ml; SeV titre: 10^8 pfu/ml) stored at –80°C, were thawed rapidly in a 37°C water bath. Ten-fold serial dilutions of stocks were made with PBS – the RSV stock was diluted 10^{-1} (Conc-1), 10^{-2} (Conc-2), 10^{-3} (Conc-3) and 10^{-4} (Conc-4), while the SeV stock was diluted 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5} (Conc-5), 10^{-6} (Conc-6), 10^{-7} (Conc-7) and 10^{-8} (Conc-8). For each dilution, a drop (50 µL) of virus sample was pipetted onto the sample well of a microscope glass slide containing a fixed-depth recess and covered with a coverslip for measurement. After each measurement, the microscope slide was washed in sterile water and dried with ethanol.

B. NIRS Measurement

The measurement system consisted of a NIR spectrometer, a light source, and a sample holder. The spectrometer was NIRQuest512 by Ocean Optics, which has a wavelength range of 901.06–1721.24 nm, and a resolution of 1.65 nm. The light source was HL-2000-HP-FHSA by Ocean Insight, which provides light output in 360–2400 nm. The typical nominal bulb power is 20W and the typical optical output power is 8.4 mW. The sample holder consists of a reflectance standard, a parafilm covered microscope glass slide with a fixed depth (120 µm) sample well of area 15 mm^2, and a slide holder. An Ocean Insight Spectralon diffuse reflectance standard with a broad, diffuse reflectance over the wavelength range of 800–2400 nm was used.

For each virus sample, 10 NIR spectra of 512 variables were acquired in reflectance mode using the OceanView software. The Spectralon reflectivity is 85%–98% across the wavelength range which was approximated to 100% reflectance. The reflectance of the test sample was determined as a relative value compared to this reflectance standard. To minimise the boundary effect, only the 51st to 412th variables (wavelengths range: 981.71–1641.9 nm) were retained for data analysis.

C. Data Processing

The data analysis scheme used in this study is shown in Fig. 1. The raw spectral data were first pre-processed and randomly divided into training set and test set according to the ratio of 4:1. All spectra of each virus sample were grouped together for training or testing so that no virus' spectra were used in both training and testing. The testing result should therefore be reliable. 5-fold cross-validation was performed on the training set to optimise the model parameter. Then the optimised model was used to classify the test set. Classification performance was assessed by accuracy, recall, precision and F1 measures [26]. The above process was repeated 20 times, and the performance over the 20 repetitions was averaged.

In this study, standard normal variate (SNV) normalization and Savitzky-Golay (SG) derivative were employed to improve classification performance. For a given spectrum, SNV subtracts the mean value of the spectrum from each variable. The obtained values are divided by the standard deviation of the spectrum. SG derivative is a commonly used pre-processing technique in spectroscopy, which includes smoothing and differentiation after the polynomial fitting [27]. PLS is a standard chemometrics method to tackle high-dimensional and high-collinear issues. It searches for linear combinations of independent variables called latent variables (LVs) that maximize the covariance between the independent variables and the response. For classification purposes, the categorical response can be transformed into numerical responses using dummy matrix coding [28].

Further, competitive adaptive reweighted sampling (CARS) was used to select important variables based on the absolute values of PLS regression coefficients [29]. It adopts Monte Carlo simulations to select some sets of variables, and important variables are determined using the exponentially decreasing function (EDF) based enforced variable selection and adaptive reweighted sampling based competitive variable selection. The separability of virus concentrations was investigated

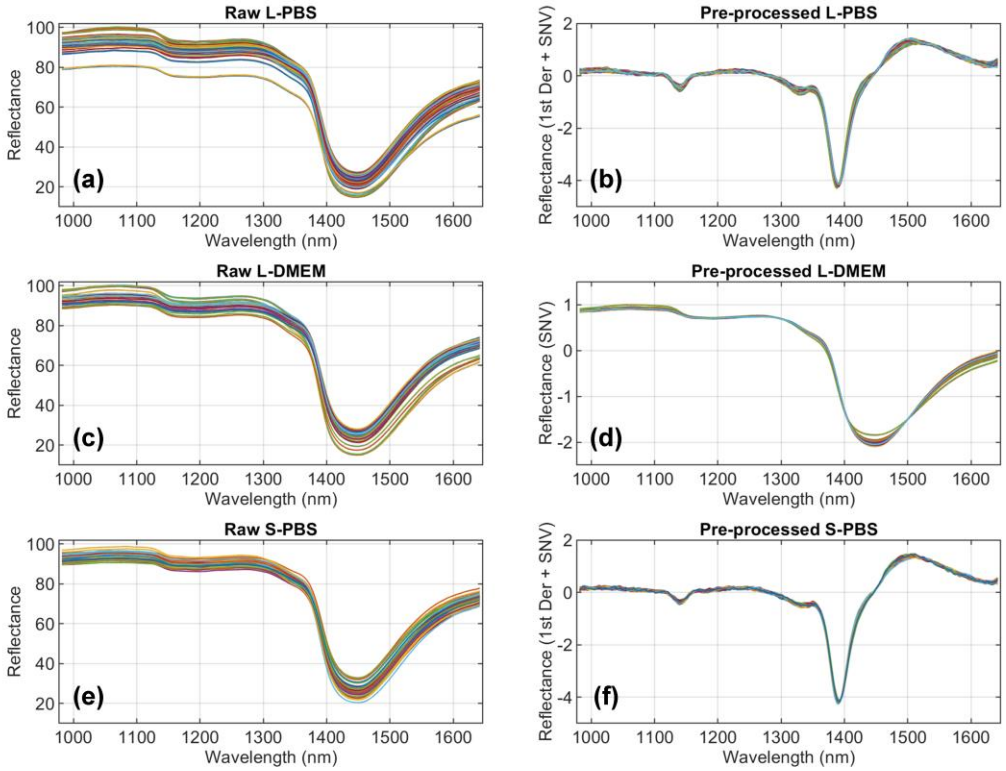


Fig. 2. The raw and pre-processed NIR spectra of RSV and SeV samples. The L-PBS and S-PBS spectra were pre-processed by SG first-order derivative and SNV, and the L-DMEM spectra were pre-processed by SNV.

TABLE II
THE AVERAGE TRAINING ACCURACY (%) ON RAW AND PRE-PROCESSED NIR VIRUS DATA

	L-PBS	L-DMEM	S-PBS
Raw	72.1	83.9	95.5
1st Der	72.4	82.4	99.2
2st Der	71.5	82.2	99.1
SNV	72.3	**87.9**	96.1
MSC	73.1	79.8	96.2
Smth	71.9	83.7	95.3
1st Der + SNV	**74.8**	87.1	**99.3**
1st Der + MSC	74.2	82.8	99.3
SNV + Smth	72.1	87.8	96.1

using t-distributed stochastic neighbour embedding (t-SNE) [30], which transforms the high-dimensional spectra into two-dimensional space to visualise their distribution. The spectra pre-processing, variable selection and data classification were implemented in MATLAB R2018b environment (The MathWorks Inc., Natick, MA, USA).

III. RESULTS AND DISCUSSION

A. Classification of RSV and SeV

Fig. 2 shows the raw and pre-processed NIR spectra measured in PBS and DMEM, where the pre-processed spectra are tightly constricted compared to the raw spectra. Due to low-cost measurement and less controlled environments, the RSV and SeV NIR spectra show notable similarities in region 981.7–1641.9 nm. Table II compares the training accuracy of different pre-processing methods, including first-order derivative (1st Der), second-order derivative (2nd Der), SNV, multiplicative signal correction (MSC) and smoothing (Smth). The average accuracy of L-PBS, L-DMEM and S-PBS data is increased by 2.7%, 4% and 3.8%, respectively, when data pre-processing is applied. More specifically, the L-PBS and S-PBS data are pre-processed by the SG first-order derivative (second-order polynomial) and SNV, with the moving window of SG derivatives being 7 and 5 points, respectively. The L-DMEM data are pre-processed by SNV normalization. Moreover, the average optimal number of LVs in the L-PBS and S-PBS data is reduced from 6.5 to 5.9 and 7.9 to 4.7, respectively, indicating that pre-processing (SG derivatives and SNV normalisation) can improve the simplicity of the PLS-DA model.

Based on the above optimised PLS-DA models, the classification accuracies of the L-PBS, L-DMEM and S-PBS test sets are 76.4%, 85.8% and 99.7%, respectively. Such results suggest the feasibility of portable NIRS combined with chemometrics to classify RSV and SeV. Table III shows the confusion matrix containing information about true (column) and predicted (row) classes over 20 repetitions. For example, in the L-PBS test set, 18.5 out of 26.5 (18.5+8) RSV predictions are correct and the remaining 8 RSV predictions are incorrect. So, the recall of RSV is 69.8% (18.5/26.5). For the precision of RSV, 18.5 out of 31.7 (18.5+13.2) true RSV cases are correctly predicted and the remaining 13.2 true RSV cases are incorrectly predicted, so the precision of RSV is 58.4% (18.5/31.7). In comparison, the

TABLE III
CONFUSION MATRIX FOR RSV AND SEV CLASSIFICATION USING PLS-DA

	Virus	RSV	SeV	Recall (%)	Precision (%)	F1 score
L-PBS	RSV	**18.5**	8.0	69.8	58.4	63.6
	SeV	13.2	**50.3**	79.2	86.3	82.6
L-DMEM	RSV	**27.6**	8.0	77.6	85.2	81.2
	SeV	4.8	**49.7**	91.2	86.2	88.6
S-PBS	RSV	**44.9**	0.1	99.8	99.6	99.7
	SeV	0.2	**64.8**	99.7	99.8	99.8

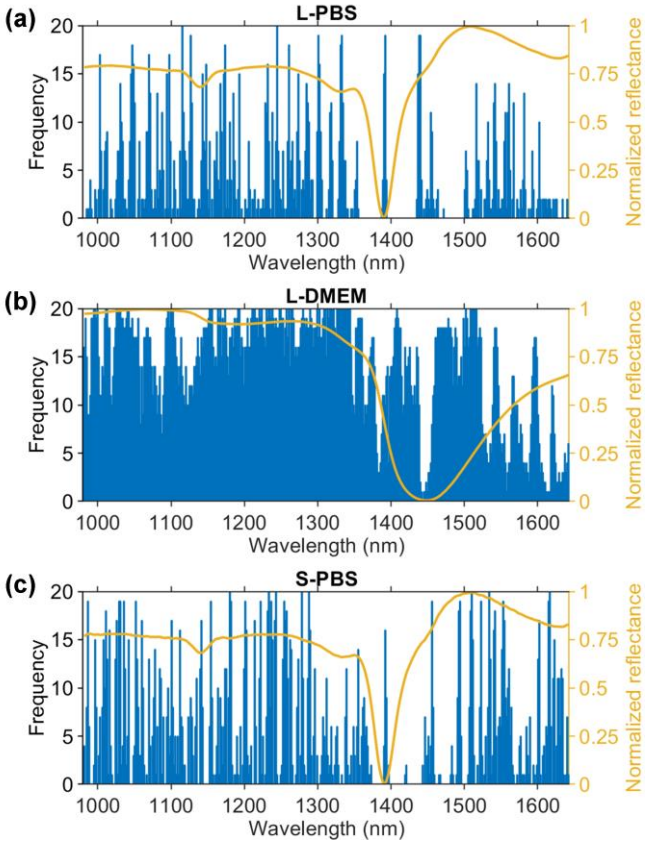


Fig. 3. The selection of important wavelengths using CARS. The frequency represents the number of times the wavelength is selected in 20 repetition tests. The average spectrum after pre-processing is shown in yellow.

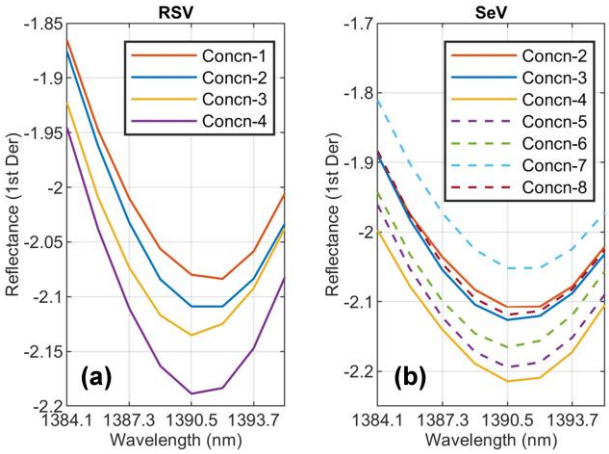


Fig. 4. The average spectra of RSV (a) and SeV (b) samples with different concentrations in the wavelength range of 1384.1–1395.3 nm.

recall and precision of RSV are significantly improved in the L-DMEM test set, reaching 77.6% and 85.2%, respectively. Moreover, the recall of SeV increases from 79.2% to 91.2%. Notably, due to the less controlled sampling environment and the limited number of training samples, the applicability domain of the model may not cover all of the test data. Therefore, the random splitting of data into training and test sets can sometimes lead to unsatisfactory test results on the long-term data. For the short-term data measured in PBS, the recall, precision and F1 score are close to 100% for each class.

Spectral peak overlapping and shifting caused by low-cost measurement can drastically degrade data quality, resulting in weak spectral fingerprints. As a data-driven approach, we use an ensemble variable selection strategy to identify spectral fingerprints [31][32]. Fig. 3 shows the selection of important wavelengths, based on the CARS method, that contributes to the classification of RSV and SeV spectra. A subset of variables is first selected based on the training samples in each random data split. Using the subset of variables, a PLS-DA model is constructed and used to predict the labels of test samples. The wavelengths with high frequencies are useful variables and will appear in most of the models. Specifically, in the L-PBS and S-PBS data, 8 wavelengths obtain frequencies exceeding 18 (1126.8 nm, 1244.7 nm, 1300.9 nm, 1332.9 nm, 1392.1 nm, 1438.4 nm and 1440 nm) and equal to 20 (1180.2 nm, 1233.4 nm, 1243.1 nm, 1278.4 nm, 1288.1 nm, 1510 nm, 1533.9 nm and 1616.5 nm), respectively. In the DMEM data, 69 of the 412 wavelengths reach 20 frequencies. Furthermore, the wavelengths of 1030.8 nm, 1069.9 nm, 1231.8 nm, 1273.6 nm and 1541.8 nm are high frequencies (>12) among the three datasets. Using the CARS method, the average number of variables of L-PBS, L-DMEM and S-PBS data after 20 repetition runs is reduced from 412 to 77.5, 291.6, and 105.3, respectively. The average test accuracy on the three datasets remains at the same level, which is 78.7%, 83.6% and 96.7%, respectively.

B. Virus Concentration Detection

In this section, we study the distinction of RSV/SeV samples with different concentrations by performing dimensionality reduction and data classification on the short-term data (S-PBS).

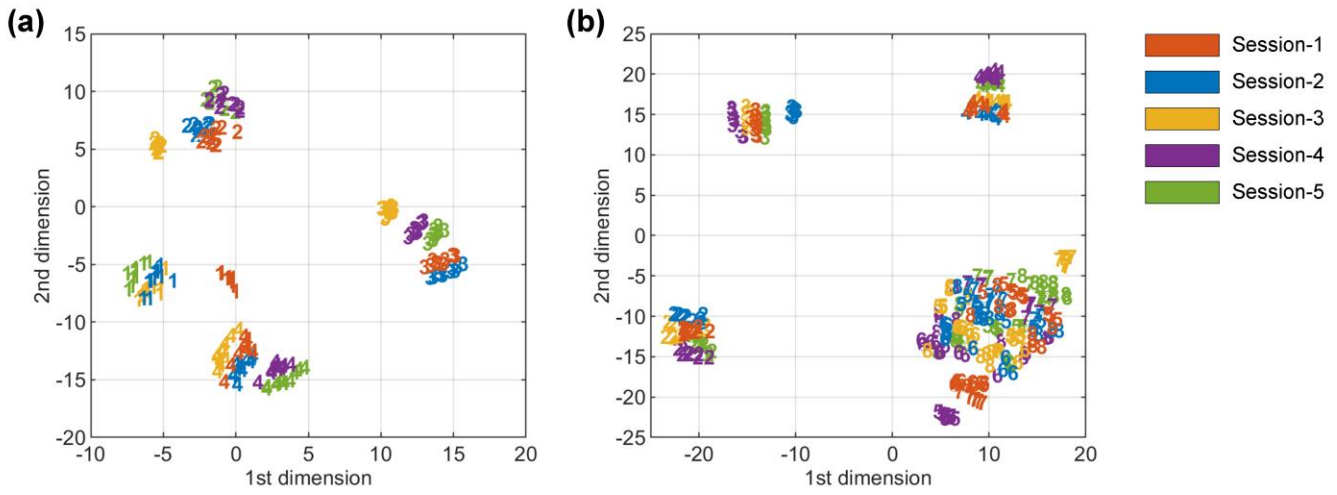


Fig. 5. The t-SNE visualisation of RSV (a) and SeV (b) spectra with different concentrations in S-PBS data. The increases in numbers from 1 to 4 and 2 to 8 represent a decrease in virus concentration.

TABLE IV
CONFUSION MATRIX OF THE CONCENTRATION CLASSIFICATION OF RSV SAMPLES

	Conc-1	Conc-2	Conc-3	Conc-4	Recall (%)	Precision (%)	F1 score
Conc-1	**10.05**	0	0	0.45	95.7	100	97.8
Conc-2	0	**10.5**	0	0	100	100	100
Conc-3	0	0	**11**	0	100	100	100
Conc-4	0	0	0	**8**	100	94.7	97.3

TABLE V
CONFUSION MATRIX OF THE CONCENTRATION CLASSIFICATION OF SEV SAMPLES

	Conc-2	Conc-3	Conc-4	Conc-5	Conc-6	Conc-7	Conc-8	Recall (%)	Precision (%)	F1 score
Conc-2	**10.5**	0	0	0	0	0	0	100	100	100
Conc-3	0	**9.95**	0	0	0	0.05	0	99.5	99.5	99.5
Conc-4	0	0	**10**	0	0	0	0	100	100	100
Conc-5	0	0.05	0	**4.65**	1.05	2.6	1.15	48.9	63.3	55.2
Conc-6	0	0	0	0.1	**9.85**	0.5	0.05	93.8	82.4	87.7
Conc-7	0	0	0	1.5	0.85	**6.75**	0.4	71.1	63.7	67.2
Conc-8	0	0	0	1.1	0.2	0.7	**8**	80	83.3	81.6

Fig. 4 shows the average spectra of different concentrations in a wavelength interval (1384.1–1395.3 nm). As the concentration increases, the derivatised reflectance increases in its value (Concn-1, 2, 3 and 4). However, this concordance between concentration and reflectance does not exist for low-concentration (Concn-5, 6, 7 and 8) spectra.

Further, the spectra are projected into a two-dimensional space by using t-SNE to present the separability of different concentrations, as shown in Fig. 5. The numbers from 1 to 7 and the 5 colours represent different concentrations and sampling sessions, respectively. The projected RSV data exhibit clear concentration separation, except for the Concn-1 spectra obtained in the first session. The SeV data in low-dimensional space have 4 distinct clusters, including Concn-2, 3, 4 and the rest of the concentrations. This suggests that the detection limit of short-term measurement will not be lower than Concn-4. In addition, the data of the same high concentration are grouped regardless of the different sessions, demonstrating good repeatability of high-concentration spectra in short-term measurement.

20 repetitive tests are performed on 19 RSV (190 spectra) and 35 SeV (350 spectra) samples. The average testing results of PLS-DA on different concentration samples are listed as confusion matrices in Table IV and Table V. The total concentration classification accuracy of RSV and SeV samples is 98.9% and 85.2%, respectively. High precisions and recalls (≥94.7%) are achieved in Conc-1, 2, 3 and 4 samples, and much lower precisions and recalls (between 63.3% to 83.3% in most cases) are achieved in lower concentration samples.

IV. Conclusions

This work presents the use of portable NIRS and PLS-DA to classify RSV and SeV spectra. Three sets of samples prepared in PBS and DMEM were measured in the long term and short term. The raw data were first pre-processed by SNV normalisation and SG derivative to improve the classification results. Then PLS-DA models were trained to build the relationship between the spectra and the virus labels/concentrations. The high-dimensional spectral data were projected into low-dimensional space to visually study the separability of different classes and concentrations. Moreover, important wavelengths that contribute to the classification of RSV and SeV spectra were identified using the CARS variable selection method.

The accuracy of virus (RSV/SeV) type classification based on 20 repetition testing were 76.4%, 85.8% and 99.7% for L-PBS, L-DMEN and S-PBS datasets, respectively. A clear separation between the two classes was obtained in the latent space when the virus samples were measured in PBS. Furthermore, in short-term measurement, the results of detecting different concentrations of RSV and SeV samples were 98.9% and 85.2%, respectively. The detection limit was around Concn-4 from the t-SNE visualisation and confusion matrix. The long-term and short-term experimental results demonstrate the effectiveness and reliability of using portable NIRS combined with chemometrics to classify RSV and SeV. Therefore, this simple, fast and low-cost approach can be potentially used as a pre-screening tool for real-time and on-site detection of similar viruses. Our future work will include identification of spectral markers, virus biomarkers, and further testing of the experimental protocol on rhinovirus, influenza virus and SARS-CoV-2.

References

[1] M. Alafeef, K. Dighe, P. Moitra, and D. Pan, "Rapid, Ultrasensitive, and Quantitative Detection of SARS-CoV-2 Using Antisense Oligonucleotides Directed Electrochemical Biosensor Chip," *ACS Nano*, vol. 14, no. 12, pp. 17028–17045, Dec. 2020.

[2] A. Parihar, P. Ranjan, S. K. Sanghi, A. K. Srivastava, and R. Khan, "Point-of-Care Biosensor-Based Diagnosis of COVID-19 Holds Promise to Combat Current and Future Pandemics," *ACS Appl. Bio Mater.*, vol. 3, no. 11, pp. 7326–7343, Nov. 2020.

[3] K. Ejima *et al.*, "Time variation in the probability of failing to detect a case of polymerase chain reaction testing for SARS-CoV-2 as estimated from a viral dynamics model," *J. R. Soc. Interface*, vol. 18, no. 177, p. 20200947, Apr. 2021.

[4] S. Yadav *et al.*, "SERS Based Lateral Flow Immunoassay for Point-of-Care Detection of SARS-CoV-2 in Clinical Samples," *ACS Appl. Bio Mater.*, vol. 4, no. 4, pp. 2974–2995, Apr. 2021.

[5] Z. Yao *et al.*, "Rapid detection of SARS-CoV-2 viral nucleic acids based on surface enhanced infrared absorption spectroscopy," *Nanoscale*, vol. 13, no. 22, pp. 10133–10142, 2021.

[6] V. G. Barauna *et al.*, "Ultrarapid On-Site Detection of SARS-CoV-2 Infection Using Simple ATR-FTIR Spectroscopy and an Analysis Algorithm: High Sensitivity and Specificity," *Anal. Chem.*, vol. 93, no. 5, pp. 2950–2958, 2021.

[7] C. Carlomagno *et al.*, "COVID-19 salivary Raman fingerprint: innovative approach for the detection of current and past SARS-CoV-2 infections," *Sci. Rep.*, vol. 11, p. 4943, Dec. 2021, doi: 10.1038/s41598-021-84565-3.

[8] B. R. Wood *et al.*, "Infrared based saliva screening test for COVID-19," *Angew. Chemie Int. Ed.*, p. anie.202104453, May 2021.

[9] A. Gredilla, S. Fdez-Ortiz de Vallejuelo, N. Elejoste, A. de Diego, and J. M. Madariaga, "Non-destructive Spectroscopy combined with chemometrics as a tool for Green Chemical Analysis of environmental samples: A review," *TrAC - Trends Anal. Chem.*, vol. 76, pp. 30–39, 2016.

[10] M. C. D. Santos, C. L. M. Morais, Y. M. Nascimento, J. M. G. Araujo, and K. M. G. Lima, "Spectroscopy with computational analysis in virological studies: A decade (2006–2016)," *TrAC - Trends Anal. Chem.*, vol. 97, pp. 244–256, 2017.

[11] A. Sakudo, K. Baba, and K. Ikuta, "Discrimination of influenza virus-infected nasal fluids by Vis-NIR spectroscopy," *Clin. Chim. Acta*, vol. 414, pp. 130–134, 2012.

[12] J. Y. Lim *et al.*, "Identification of Newly Emerging Influenza Viruses by Detecting the Virally Infected Cells Based on Surface Enhanced Raman Spectroscopy and Principal Component Analysis," *Anal. Chem.*, vol. 91, no. 9, pp. 5677–5684, 2019.

[13] R. A. Dluhy, S. Shanmukh, L. Jones, Y. P. Zhao, J. D. Driskell, and R. A. Tripp, "Identification and classification of respiratory syncytial virus (RSV) strains by surface-enhanced Raman spectroscopy and multivariate statistical techniques," *Anal. Bioanal. Chem.*, vol. 390, no. 6, pp. 1551–1555, 2008.

[14] J. H. Lee, B. C. Kim, B. K. Oh, and J. W. Choi, "Rapid and sensitive determination of HIV-1 virus based on surface enhanced raman spectroscopy," *J. Biomed. Nanotechnol.*, vol. 11, no. 12, pp. 2223–2230, 2015.

[15] S. Khan *et al.*, "Analysis of hepatitis B virus infection in blood sera using Raman spectroscopy and machine learning," *Photodiagnosis Photodyn. Ther.*, vol. 23, pp. 89–93, 2018.

[16] N. Liu, H. A. Parra, A. Pustjens, K. Hettinga, P. Mongondry, and S. M. van Ruth, "Evaluation of portable near-infrared spectroscopy for organic milk authentication," *Talanta*, vol. 184, pp. 128–135, 2018.

[17] A. Kartakoullis, J. Comaposada, A. Cruz-Carrión, X. Serra, and P. Gou, "Feasibility study of smartphone-based Near Infrared Spectroscopy (NIRS) for salted minced meat composition diagnostics at different temperatures," *Food Chem.*, vol. 278, pp. 314–321, 2019.

[18] H. S. Rashed, P. Mishra, A. Nordon, D. S. Palmer, and M. J. Baker, "A comparative investigation of two handheld near-ir spectrometers for direct forensic examination of fibres in-situ," *Vib. Spectrosc.*, vol. 113, p. 103205, Mar. 2021.

[19] A. Guillemain, K. Dégardin, and Y. Roggo, "Performance of NIR handheld spectrometers for the detection of counterfeit tablets," *Talanta*, vol. 165, pp. 632–640, Apr. 2017.

[20] W. Song, H. Wang, P. Maguire, and O. Nibouche, "Nearest clusters based partial least squares discriminant analysis for the classification of spectral data," *Anal. Chim. Acta*, vol. 1009, pp. 27–38, 2018.

[21] W. Song, Z. Song, J. Vincent, H. Wang, and Z. Wang, "Quantification of extra virgin olive oil adulteration using smartphone videos," *Talanta*, vol. 216, p. 120920, Aug. 2020.

[22] D. Jian *et al.*, "Sunlight based handheld smartphone spectrometer," *Biosens. Bioelectron.*, vol. 143, p. 111632, 2019.

[23] J. Engel *et al.*, "Breaking with trends in pre-processing?," *TrAC - Trends Anal. Chem.*, vol. 50, pp. 96–106, 2013.

[24] J. F. Lopes, L. Ludwig, D. F. Barbin, M. V. E. Grossmann, and S. Barbon, "Computer Vision Classification of Barley Flour Based on Spatial Pyramid Partition Ensemble," *Sensors*, vol. 19, no. 13, p. 2953, Jul. 2019.

[25] W. Song *et al.*, "Spectral knowledge-based regression for laser-induced breakdown spectroscopy quantitative analysis," *Expert Syst. Appl.*, vol. 205, p. 117756, Nov. 2022.

[26] M. M. Oliveira, B. V. Cerqueira, S. Barbon, and D. F. Barbin, "Classification of fermented cocoa beans (cut test) using computer vision," *J. Food Compos. Anal.*, vol. 97, p. 103771, Apr. 2021.

[27] Å. Rinnan, F. van den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC - Trends Anal. Chem.*, vol. 28, no. 10, pp. 1201–1222, 2009.

[28] M. Barker and W. Rayens, "Partial least squares for discrimination," *J. Chemom.*, vol. 17, no. 3, pp. 166–173, 2003.

[29] H. Li, Y. Liang, Q. Xu, and D. Cao, "Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration," *Anal. Chim. Acta*, vol. 648, no. 1, pp. 77–84, 2009.

[30] V. der M. L and H. G, "Visualizing Data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

[31] W. Song *et al.*, "Validated ensemble variable selection of laser-induced breakdown spectroscopy data for coal property analysis," *J. Anal. At. Spectrom.*, vol. 36, no. 1, pp. 111–119, 2021.

[32] Q. Ouyang, L. Wang, M. Zareef, Q. Chen, Z. Guo, and H. Li, "A feasibility of nondestructive rapid detection of total volatile basic nitrogen content in frozen pork based on portable near-infrared spectroscopy," *Microchem. J.*, vol. 157, p. 105020, 2020.