Decoding Viral Mixtures: SERS and Deep Learning Unraveling Complex Pathogens

Yanjun Yang[1], Jiaheng Cui[1], Dan Luo[2], Jackelyn Murray[3], Les Jones[3], Xianyan Chen[4], Ralph A. Tripp[3], Yiping Zhao[5]

1. School of Electrical and Computer Engineering, University of Georgia, Athens, GA 30602, USA
2. Department of Statistics, The University of Georgia, Athens, GA, USA 30602
3. Department of Infectious Diseases, The University of Georgia, Athens, GA 30602
4. Department of Epidemiology and Biostatistics , University of Georgia, Athens, GA 30602, USA
5. Department of Physics and Astronomy, University of Georgia, Athens, GA 30602, USA

Introduction

Various respiratory viruses, including those causing influenza-like illnesses, have the potential to lead to epidemics. Notably, during the COVID-19 pandemic, instances of SARS-CoV-2 infections co-occurring with influenza, RSV, and/or adenoviruses were observed. Recognizing and identifying such co-infections is crucial for tailoring targeted treatment strategies, mitigating the risk of misdiagnosis, and gaining insights into the disease's progression.

We propose to use Surface-Enhanced Raman Spectroscopy (SERS) as a platform with its fingerprint SERS peaks, to differentiate and quantify mixed viruses in co-infected specimens. By leveraging deep machine learning to help differentiate and quantify SERS spectra of potential viruses in patient specimens, we aims to create a database of SERS spectra to build a deep learning model to simultaneously differentiate and quantify different virus species in a biological mixture such as saliva.

Objectives

- Construct a SERS spectral database of virus mixtures with different concentrations from thirteen virus species by collecting SERS spectra from AgNR@SiO2 SERS substrates.
- Build a CNN-based deep learning model (MixNet) to predict both the virus species and concentrations from single viruses, two-virus mixtures and three-virus mixtures.

Detection and database strategy

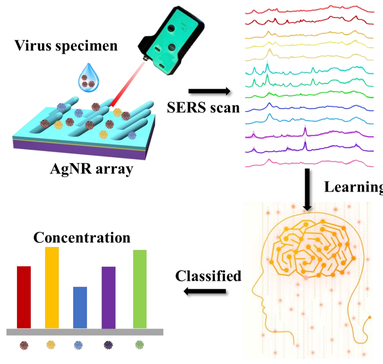

Fig. 1. Schematic illustration of deep learning-based virus mixture differentiation: specimen preparation and SERS measurements to obtain SERS spectra, as well as classification and quantification using deep learning models.

Construction of SERS spectral database:

- **Individual viruses:**
  - 13 respiratory viruses (SARS-CoV-2, SARS-CoV-2 B1, CoV-OC43, CoV-NL63, CoV-229E, Flu B, H1N1, H3N2, HNPV-A, HMPV-B, RSV-A2, RSV-B1, and Ad5).
  - The viruses were diluted to concentrations ranging from 10^2 to 10^5 PFU/mL.
- **Two-virus mixtures:**
  - Viruses with unique SERS peaks: 7 sets of mixtures (CoV-NL63 & RSV-A2, CoV-NL63 & RSV-B1, CoV-NL63 & H3N2, H1N1 & RSV-A2, H1N1 & RSV-B1, H3N2 & RSV-A2, and H3N2 & RSV-B1).
  - Highly similar viruses: 1 mixture set (CoV-NL63 & Flu B).
  - Virus A and Virus B were formulated into 11 concentrations ranging from 10^2 to 10^5 PFU/mL.
- **Three-virus mixtures:**
  - 4 sets of mixtures: CoV-NL63 & H1N1 & RSV-A2, CoV-NL63 & H1N1 & RSV-B1, CoV-NL63 & H3N2 & RSV-A2, and CoV-NL63 & H3N2 & RSVB1.
  - Virus A, Virus B, and Virus C were made into 7 or 8 different concentrations, spanning from 195 to 10^5 PFU/mL.

Understanding SERS spectra of virus mixtures

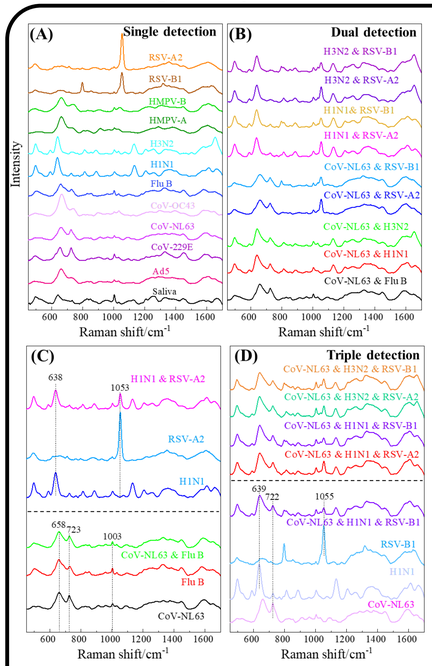

Fig. 2. (A) The representative SERS spectra of 13 single viruses with concentrations of 10^5 PFU/mL.
(B) 8 two-virus mixtures with both concentrations of 10^5 PFU/mL.

If two mixtures are identical in one component and similar in the other (e.g., H3N2 & RSV B1, H3N2 & RSV-A2), the spectra of the mixtures will be similar.

(C) Representative SERS spectra of H1N1, RSV-A2, and their mixtures; representative SERS spectra of CoV-NL63, Flu B, and their mixtures.
(D) 4 three-virus mixtures with all components having concentrations of 10^5 PFU/mL; representative SERS spectra of CoV-NL63, H1N1, RSV-B1, and their mixtures.

The unique peaks of the component viruses will appear in the mixture spectra, no matter if the components are similar (CoV-NL63 & Flu B) or not (H1N1 & RSV-A2).

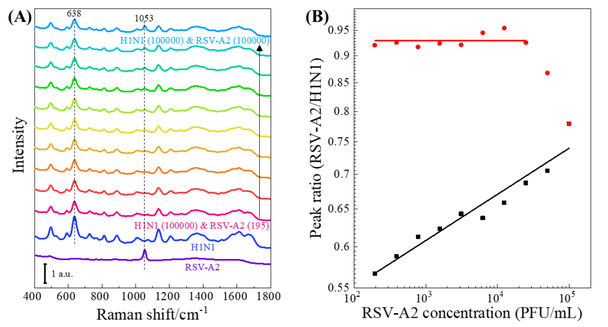

Fig. 3. (A) Average SERS spectra of H1N1, RSV-A2, and two mixtures of H1N1 and RSV A2, the value in the parenthesis is the concentration of the virus with the unit PFU/mL; (B) The peak intensity ratio for virus mixtures with fixed H1N1 concentration and varied RSV-A2 concentration (black data points), as well as virus mixtures with $C_{H1N1} = C_{RSV\text{-}A2}$ and varied RSV-A2 concentration (red data points).

It is expected that the relative peak intensities unique to the two viruses changes with the relative concentration of the viruses. **Figure 3A** plots the normalized average spectra of H1N1 & RSV-A2 for different RSV-A2 concentrations. Though all these three characteristic peaks were observed in the spectra of the mixtures, their relative peak intensities from the same spectrum vary due to the change of the concentration ratios.

Figure 3B shows a semi-log plot of the peak ratio I_{1503}/I_{638} versus the log $[C_{RSV\text{-}A2}]$ and a linear relationship appeared. For the mixture with $C_{H1N1} = C_{RSV\text{-}A2}$, I_{1503}/I_{638} does not change with $C_{RSV\text{-}A2}$ in a certain concentration region as shown in red data points.

The deep learning model - MixNet

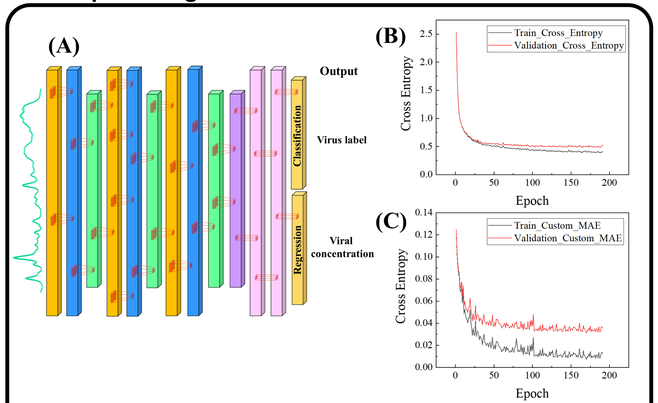

Fig. 4. (A) The architecture of deep learning model "MixNet" for multi-class classification and regression. (B) and (C) Loss curves for classification (Cross entropy) and regression (MAE) during model training and validation.

Classification using deep learning model

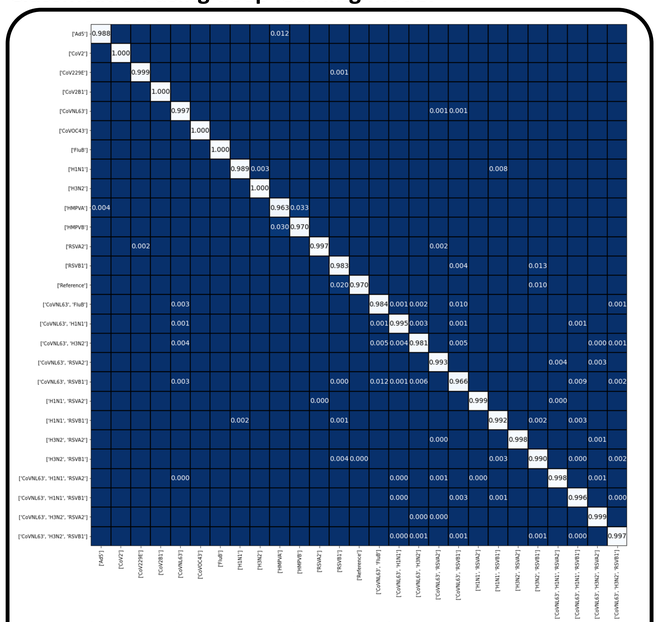

Fig. 5. Confusion matrix of the MixNet model for mixture classification. Entries in the matrix represent the percentage of test spectra that were predicted by the MixNet model as class (first row) given a ground truth of class (first column); entries along the diagonal represent the accuracies for each class.

The accuracy in the test set is 0.89.

A virus is prone to be misclassified to another virus when they have similar spectral shape, such as HMPVA and HMPVB. A mixture is prone to be misclassified to a mixture with similar components, such as H1N1 & RSV-A2 mixture and H3N2 & RSV-A2 mixture, or with one of its components, such as H1N1 & RSV-A2 mixture and RSV-A2.

Regression using deep learning model

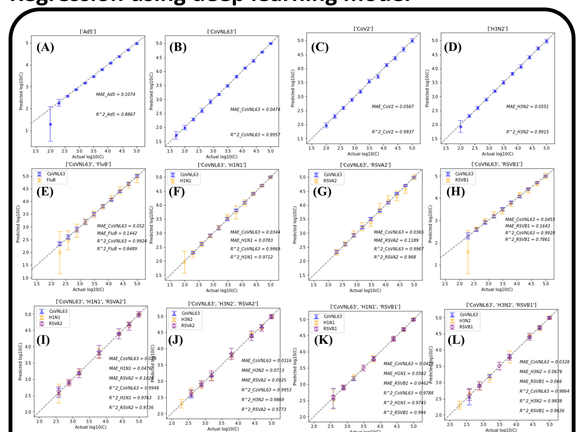

Fig. 6. Regression results of deep learning model for detection of thirteen viruses in saliva for single viruses: (A) Ad5, (B) CoV-NL63, (C) CoV-2, (D) H3N2; two-virus mixtures: (E) CoV-NL63 & Flu B, (F) CoV-NL63 & H1N1, (G) CoV-NL63 & RSV-A2, (H) CoV-NL63 & RSV-B1; three-virus mixtures: (I) CoV-NL63 & H1N1 & RSV-A2, (J) CoV-NL63 & H3N2 & RSV-A2, (K) CoV-NL63 & H1N1 & RSV-B1, (L) CoV-NL63 & H3N2 & RSV-A2. The x-axis is the logarithm of actual concentration of testing spectra, and y-axis is the logarithm of the predicted concentration of testing spectra. The dash line represents the predicted concentration C_{pre} is identical to actual concentration C_{act}, i.e., perfect prediction.

Examples of regression results are plotted in **Figure 6** in log-log scale. The predicted concentrations for virus mixtures all follow the linear relationship $\log(C_{pre}) = \log(C_{act})$, with small MAEs and $R^2 > 0.8$.

In general, higher quantification accuracies can be observed from single virus compared to those with virus mixtures. This is understandable since larger variation shall be presented due to the similarity in virus spectra, especially in mixture specimens.

In addition, for specimens with low concentrations (≤195 PFU/mL), either of single viral specimens or mixtures, there is a large difference between C_{pre} and C_{act}, which is due to the interference of background medium as well as the noise.

Conclusions

We have created a label-free diagnostic platform that utilizes SERS and deep learning to detect mixtures of respiratory virus species quickly and accurately. The platform can detect 13 single virus species, 8 dual virus species, and 4 triple virus species. We also developed a deep learning model called MixNet, which can classify and quantify virus mixtures. The model can predict not only the types of viruses in the mixtures with 89% accuracy but also the absolute concentration of each virus in the mixture. These results demonstrate the effectiveness of the SERS + deep learning approach in diagnosing complex infectious specimens.

Acknowledgement

Yanjun Yang, and Yiping Zhao were partially supported by the National Science Foundation under the contract #ECCS-1808271. Jiaheng Cui, Xianyan Chen, and Yiping Zhao are funded by USDA NIFA Grant number 2023-67015-39237.