

Article

Not peer-reviewed version

# Machine Learning Techniques for Effective Pathogen Detection based on Resonant Biosensors

[Guoguang Rong](#), Yankun Xu, [Mohamad Sawan](#) \*

Posted Date: 2 August 2023

doi: 10.20944/preprints202308.0121.v1

Keywords: Machine Learning; Support Vector Machine; Multilayer Perceptron; Photonic Biosensor; Signal Processing; Tamm Plasmon Polariton; Localized Surface Plasmon Resonance



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# Machine Learning Techniques for Effective Pathogen Detection Based on Resonant Biosensors

Guoguang Rong, Yankun Xu and Mohamad Sawan \*

CenBRAIN Neurotech Center of Excellence, School of Engineering, Westlake University, 600 Duncun Road, Xihu District, Hangzhou, Zhejiang 310030, P. R. China; rongguoguang@westlake.edu.cn (G.R.); xuyankun@westlake.edu.cn (Y.X.)

\* Correspondence: sawan@westlake.edu.cn

**Abstract:** We describe a machine learning (ML) approach to process the signals collected from Covid-19 optical-based detector. Multilayer Perceptron (MLP) and Support Vector Machine (SVM) were used to process both raw data and feature engineering data, and high performances for qualitative detection of the SARS-CoV-2 virus with concentration down to 1 TCID<sub>50</sub>/ml has been achieved. Valid detection experiments contain 486 negative and 108 positive samples; and control experiments, in which biosensors without antibody functionalization were used to detect SARS-CoV-2, contains 36 negative samples and 732 positive samples. Data distribution patterns of the valid and control detection dataset, based on T-distributed Stochastic Neighbor Embedding (t-SNE), was used to study the distinguishability between positive and negative samples, and explain the ML prediction performances. This work demonstrates that ML can be a generalized effective approach to process signals and dataset of biosensors dependent on resonant modes as biosensing mechanism.

**Keywords:** machine learning; Support Vector Machine; Multilayer Perceptron; photonic biosensor; signal processing; Tamm Plasmon Polariton; Localized Surface Plasmon Resonance

## 1. Introduction

The global COVID pandemic has caused huge impact on world health and economy [1]. The fast-spreading virus SARS-CoV-2 is the main culprit, and detection of the virus in human population is crucial for curbing the epidemic [2]. Traditional detection approaches are mainly Nucleic Acid Amplification Test (NAAT) [3] and antigen detection [4] techniques. Currently, the mainstream is quantitative Polymerase Chain Reaction (qPCR) [5] which is a kind of NAAT that has high sensitivity and specificity, but requires clean environment, bulky and expensive equipment, and trained personnel. Therefore, qPCR is not suitable for onsite, fast turnaround detection, or population scale screening, which are often required in pandemic control scenarios [6]. To complement qPCR, antigen detection based on lateral flow [7] has also been employed in home use and self-test. However, antigen detection is limited in detection sensitivity and specificity, hindering its efficacy in fighting a pandemic [8]. There still lacks rapid, accurate and low cost detection techniques that can be deployed onsite for population scale epidemic screening and/or surveillance [9], especially for regions of limited resources [10].

Biosensors have been proposed for detection of SARS-CoV-2 [11]. Biosensor technologies have the advantages of high sensitivity, good specificity, fast turnaround, ease of operation, low cost, and onsite deployment capability [12, 13]. We have previously proposed a photonic biosensor for fast onsite detection of SARS-CoV-2 with high sensitivity and specificity [14, 15]. The biosensor is based on nanoporous silicon material fabricated via CMOS-compatible silicon process, and nanophotonic working principles of Localized Surface Plasmon Resonance (LSPR) [16] and Tamm Plasmon Polariton (TPP) [17, 18]. The measurement of the biosensor is based on reflection spectroscopy [14].

We have also developed handheld and high throughput detection systems [19] that can collect the reflection spectrum of biosensors and process the spectral data to determine the detection results efficiently. The high throughput detection system is suitable for populations scale screening of infection, and the handheld detection system is for home use or self-test. The spectral data processing algorithm works by recognizing the characteristic resonant valleys in the reflection spectrum of the biosensor and determines the detection result by judging if there is spectral red shift of the characteristic resonant valleys. This is the often used and so called “find peaks” technique, with its name originating from the MATLAB function *find\_peaks()*. This technique can also be implemented on Field Programmable Gate Array (FPGA) for fast and efficient processing of signals from array of biosensors [20]. In addition, researchers have also proposed Interferogram Average over Wavelength (IAW) technique to process signals of optical biosensors that depend on spectral shift of characteristic resonant features, which can achieve sensitivity enhancement compared with spectral shift detection [21]. Detection of change in reflection intensity due to shift of spectral features in spectrum has also been used to detect biomolecules in real time [22]. However, both IAW and light intensity measurement techniques are subject to spectral amplitude fluctuations, and thus requires highly stable spectroscopy systems, such as stable light source and high signal-to-noise ratio spectrometer.

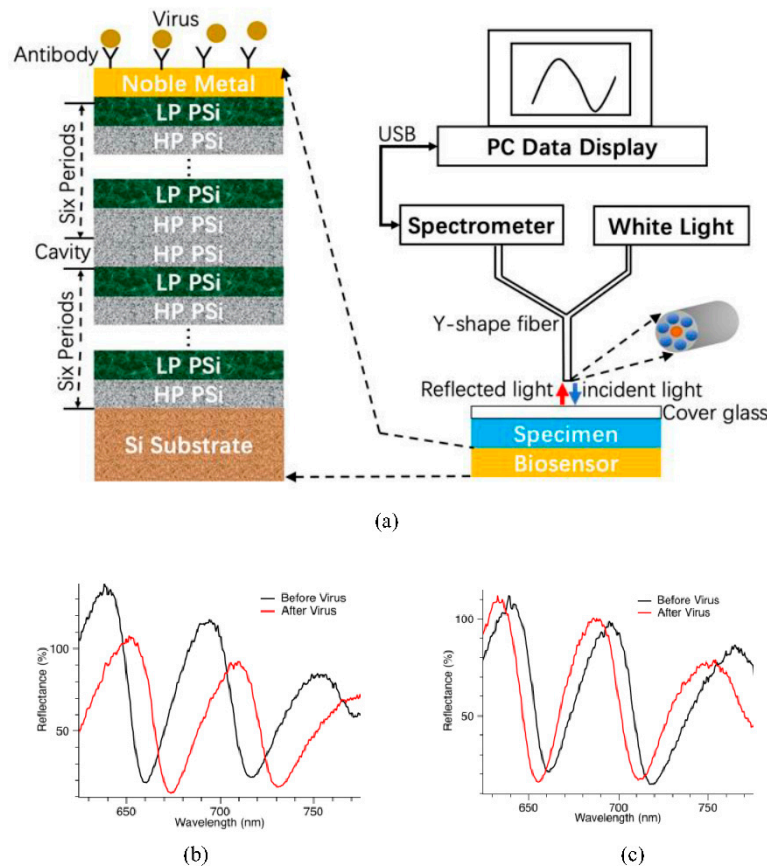
In this work, we demonstrate that it is advantageous to utilize artificial intelligence technology, more specifically machine learning (ML) algorithm to process the spectral data of the biosensor [23]. Instead of depending on programming, its algorithm is learnt from big volume of data [24]. Machine learning has been used for computer vision [25], face recognition [26], autonomous driving [27], auxiliary decision-making [28], brain-machine interface [29], and games [30]. It includes supervised learning, unsupervised learning, and reinforcement learning [31]. Supervised learning (SL) is an algorithm that learns from massive, labeled data and generates prediction models that can work to generate labels for new dataset. SL includes Support Vector Machine (SVM) [32], Multilayer Perceptron (MLP) [33], Linear Regression [34], Linear Discriminant Analysis [35], K-nearest Neighbor [36], Decision Tree [37], and Naïve Bayes [38]. In this work, we demonstrate that SVM and MLP can be used for processing of the photonic biosensor signal and dataset. Compared with previously proposed techniques, ML technique has the advantages of: 1) no need to find appropriate parameters of the algorithm, e.g. the *find\_peaks()* function, in a try-and-error way to guarantee accurate recognition of spectral feature; 2) no need to discriminate between redshift or blueshift which can be an extra issue in algorithm design; 3) not sensitive to spectral amplitude fluctuations so that requirements on stable and expensive hardware are relaxed; 4) generalizable to all kinds of sensors with salient features in response signal which serve as the basis of discriminating between positive and negative responses.

Data visualization approach can help us to understand the distribution of dataset and find out the distinguishability of the dataset. T-distributed Stochastic Neighbor Embedding (t-SNE) is a prevalent approach to map high-dimensional data to low-dimensional embedding [39]. In this contribution, we also implemented t-SNE approach on the SARS-CoV-2 detection dataset to clarify the distinguishability of the biosensor dataset so that a better understanding of the data processing and ML prediction performances can be obtained.

## 2. Materials and Methods

### 2.1. Biosensor Working Principal and Measurement Setup

As shown in Figure 1(a), the biosensor is basically a porous silicon microcavity consisting of two Bragg reflectors and one resonant cavity [14, 40].



**Figure 1.** Photonic biosensor : (a) Structure and its reflection spectroscopy measurement; b) Typical example of redshift showing resonant valleys in reflection spectra; (c) blueshift of resonant valleys in reflection spectra.

One Bragg reflector is six periods of alternating low porosity (LP) and high porosity (HP) porous silicon (PSi) thin films of thickness equal to quarter resonant wavelength. On top of the porous silicon is deposited noble metal thin film. Because of the nanoporous structure of the porous silicon material, the conformally deposited noble metal thin film is also porous. When light is incident on the surface of the biosensor, some of its energy will couple into Localized Surface Plasmon Resonance (LSPR) supported [40] by the nanostructures of the noble metal thin film. In addition, some of its energy will also couple into Tamm Plasmon Polariton (TPP) supported by the interface between the top Bragg reflector and the noble metal thin film [41]. Therefore, LSPR and TPP are excited by the incident light simultaneously, coupled with each other, and form strong field confinement around the noble metal thin film. If specific antibodies are immobilized beforehand on the surface of the noble metal, they can capture SARS-CoV-2 virus specifically. Such binding events will cause addition of biomaterials around the noble metal thin film, and the added biomaterial will interact strongly with the LSPR and TPP coupled field. This is the working mechanism of the biosensor for sensitive detection of virus. As shown in Figure 1(a), in order to measure the signals of the biosensor, reflection spectroscopy is used. White light source provides incident light, which goes through the Y-shape fiber and shine vertically onto the biosensor surface. The light reflected from the biosensor surface is collected by the Y-shape fiber and goes into the spectrometer for data analysis. The Y-shape fiber consists of six circumferential fibers guiding incident light, and one central fiber guiding reflected light.

Figures 1(b) and 1(c) show representative reflection spectra of the biosensor. They have characteristic resonant valleys in the spectral range of 600-800 nm in wavelength. If there are viruses binding with antibodies on the biosensor surface, the binding events will cause shift of the spectral features to longer wavelength, which is also called "redshift". For example, Figure 1(b) shows such a case where virus binds with antibodies, redshift occurs and the detection result is determined to be

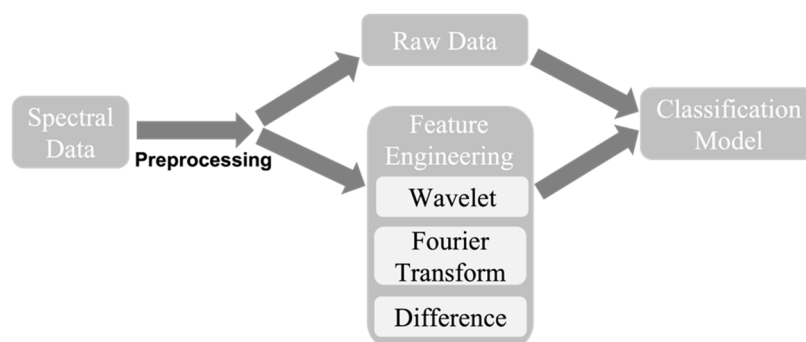
positive. On the other hand, if there is no virus binding with antibodies on the biosensor surface, there will be no shift of the spectral features, i.e., almost overlapping spectra for both before and after binding reaction. In the third case, there could also appear shift of the resonant features to shorter wavelength, which is also called “blueshift”. In such case, the detection result is determined to be negative. Figure 1(c) shows an example wherein there appears blueshift. In summary, the principle of the biosensor is based on interactions between biomaterials and photonic energy, and the detection result is determined based on shift of spectral features in optical spectrum collected from reflection spectroscopy measurement.

## 2.2. Data Preprocessing

The dataset was obtained from detection experiments of inactivated SARS-CoV-2 in clinical swab specimens, with virus concentration as low as 1 TCID<sub>50</sub>/ml [14]. Figure 1 shows example spectra of the biosensor for positive and negative detection results. For positive result, there is spectral redshift; and for negative result, there is no spectral shift or there is spectral blueshift. The experimental data is collected by reflection spectroscopy with corresponding spectra for before and after applying specimens on biosensor surface. Each spectral data contains 2048 data points representing reflection intensities, with data-to-data spacing of 0.48 nm in the wavelength range of 200-1200 nm. We usually need to carry out preprocessing of the spectral data before data analysis, such as normalization and artifacts removal. Furthermore, normalization was implemented on each data sample for the purpose of training convenience. Spectral data of both before and after adding specimens are combined as a single sample, so that the size of the reformed sample is 2×2048, or 4096. Each detection experiment is regarded as a sample for either training or test purpose. After several outliers were removed to clean the dataset, there are 486 negative samples and 108 positive samples left in total for the classification model training and prediction test.

## 2.3. Feature Engineering

As shown in Figure 2, the input to the model could be the data of size 4096. This requires 4096 input neuron nodes which could be a computation burden. In addition to this raw data approach, the input could also be features extracted from the data. We propose feature engineering methods comprising three different approaches – wavelet transform, Fourier transform, and spectral difference. As for wavelet domain, we used the wavelet transform with scales of 30 and take average of each scale, which generates 30 features for each spectral curve, two curves (before and after virus) would generate 60 wavelet-based features. In terms of Fourier domain, we have found that most information appears in the low frequency range (< 50Hz), so that we took average of each 5Hz in order from 0 to 50Hz, so that 10 features for each spectral curve and 20 features for spectra pairs would be obtained in Fourier domain. For spectral difference, we utilized the difference between the spectral data of before and after binding reaction on biosensor, instead of two separate spectra. There are three features selected from spectral difference, they are mean, variance and sign change rate.



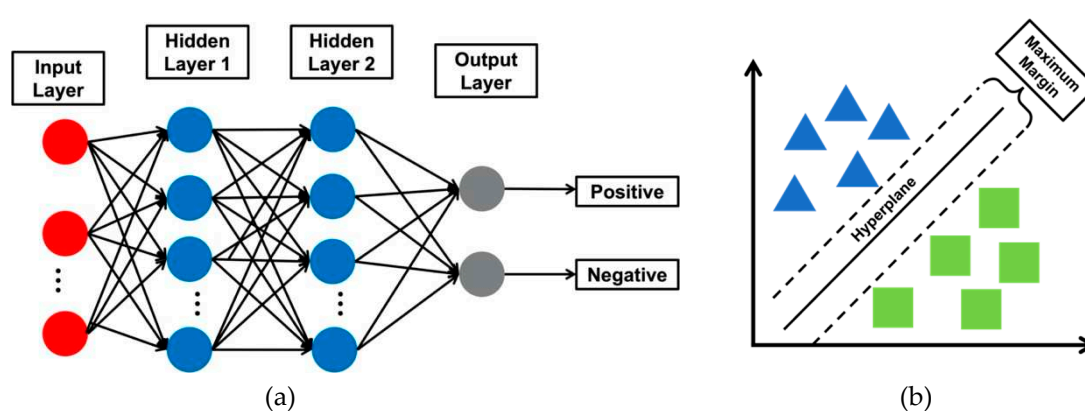
**Figure 2.** Simplified block diagram of the data processing procedure. Raw data and feature engineering methods are used in the experiments.



Eventually, for each training sample containing spectral data of before and after reaction, wavelet transform and Fourier transform domain features need spectra of both before and after reaction, and spectral difference features only need difference between spectra of before and after reaction. Therefore, there are 83 (60 wavelet domain + 20 Fourier domain + 3 spectral difference) features selected for the classification experiments.

#### 2.4. Classification Models

All samples were randomly shuffled and separated as 70% for training and 30% for test. This allocation ratio is a practical standard for benchmark performance. Multilayer Perceptron (MLP) and Support Vector Machine (SVM) models were used since they are usually considered as efficient ML models capable of achieving baseline performance. As shown in Figure 3, in terms of MLP model, two hidden layers with 100 and 50 neuron nodes are implemented, optimizer is stochastic gradient decent solver, learning rate and epoch is set as 0.1 and 30, respectively. As for SVM model, we set the gamma parameter of radial basis kernel function as 1.



**Figure 3.** (a) Simplified diagram for illustration of MLP architecture; and (b) SVM explanation.

#### 2.5. Control Experiments

For control experiments, we detected SARS-CoV-2 specimens with photonic biosensors which do not have specific antibodies immobilized on biosensor surface beforehand. There are in total 732 data samples of detecting SARS-CoV-2 virus specimens of various concentrations, and 36 data samples of detecting specimens containing no SARS-CoV-2 viruses. This new dataset is processed by the SVM and MLP models already trained as shown in Figure 3.

#### 2.6. Dataset Distinguishability Analysis

Nowadays, data visualization approach can help understand the distribution of the dataset and intuitively investigate whether the dataset is distinguishable or not. T-distributed Stochastic Neighbor Embedding (t-SNE) is a tool to visualize high-dimensional data. It converts similarities between data points to joint probabilities and tries to minimize the Kullback-Leibler divergence [42] between the joint probabilities of the low-dimensional embedding and the high-dimensional data.

We implemented t-SNE tool on specimen detection dataset to interpret the distinguishability of the dataset. The data distribution patterns can help interpret performances of models on the dataset. Both raw dataset and features extracted from raw data are considered of their distinguishability. We also investigated whether extracted features have distributions different from that of raw data.

### 3. Results and Discussion

In terms of the experiments, we used SVM and MLP models to test the raw data processing and feature engineering method. Two performance metrics are considered in the experiments: sensitivity (SEN) and specificity (SPE) which are defined as

$$SEN = \frac{TP}{TP+FN} \quad (1)$$

$$SPE = \frac{TN}{TN+FP} \quad (2)$$

where TP, FN, TN, FP stand for true positive, false negative, true negative and false positive, respectively.

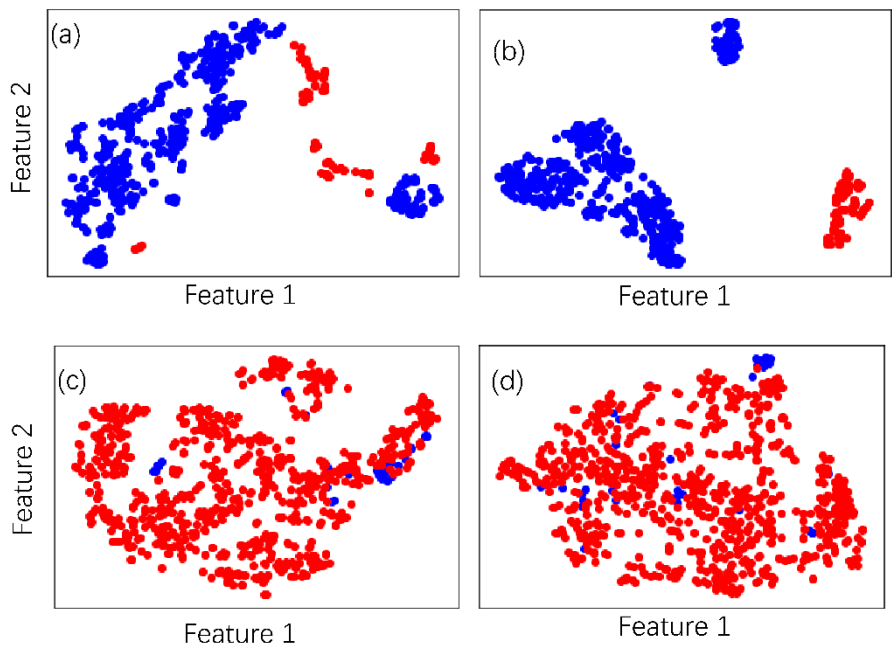
Table 1 shows the performances of ML model predictions. We can see that perfect performances are achieved for both raw data and feature engineering methods, combined with either SVM or MLP model. The last row in Table 1 shows the performance of the models in processing the control experiment dataset. The performance is very poor, and this is due to the fact that the biosensors have not been functionalized with specific antibodies and thus, cannot detect SARS-CoV-2 virus effectively.

**Table 1.** Performance of Raw Data and Feature Engineering Processing Methods with Two Machine Learning Models.

Method	Raw Data				Feature Engineering			
Model	SVM		MLP		SVM		MLP	
Parameter	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE
Performance on Control Detection Data	100%	0%	100%	0%	0%	83%	0%	90%
Performance on Valid Detection Data	100%							

SVM: Support Vector Machine      MLP: Multilayer Perceptron      SEN: Sensitivity      SPE: Specificity

Figure 4 (a) shows data distribution of raw dataset in 2D space by t-SNE data visualization approach. We can see that the positive and negative samples from dataset of valid detection experiments are clustered without any overlapping. Thus, the valid experimental dataset is distinguishable. Figure 4 (b) shows data distribution of features extracted from the dataset of Figure 4 (a). The extracted features change the data distribution, while maintaining the distinguishability because the samples are separated into different clusters. Figure 4(c) shows the data distribution of dataset obtained from control experiments wherein biosensors are not functionalized with specific antibodies. Negative samples are overlapping with positive samples, and the dataset is indistinguishable according to the visualization results. Figure 4 (d) shows data distribution of features extracted from the dataset of Figure 4 (c). The distribution of features' dataset is still mixed up, so that feature engineering cannot help the dataset to be classified effectively. These dataset distribution results could serve to interpret the performance comparisons demonstrated in Table 1.



**Figure 4.** The t-SNE data visualization results of experimental SARS-CoV-2 detection dataset, red and blue represent positive and negative samples respectively: (a) Raw dataset of valid detection experiment; (b) Feature engineering dataset of valid detection experiment; (c) Raw dataset of control detection experiment; (d) Features engineering dataset of control detection experiment.

Table 2 demonstrates the advantages of ML data processing technique when compared with other techniques. It can be seen that the general advtantages of ML are valid, in addition to eased hardware requirement.

**Table 2.** Comparison of machine learning technique with other signal processing techniques.

<div>Factor Technique</div>	Need data filtering and denoising	Need to take care of shift direction	Need stable light source and low noise spectroscopy system	Needed researcher work
Find peaks and calculate spectral shift	Yes	Yes	No	Algorithm design and test
Interferogram average over wavelength	Yes	No	Yes	Algorithm design and test
Intensity interrogation	Yes	No	Yes	Algorithm design and test
Machine learning	Yes	No	No	Model training from data

To verify the efficacy of the ML data processing technique for biosensors, detection experiments of inactivated SARS-CoV-2 in vaccination sites of Hangzhou Center for Disease Control and Prevention (CDC) were carried out and the detection results are compared with the gold standard-reverse transcription qPCR technique. The envrionmental specimens were collected from various locations in different vaccination sites, delivered to Hangzhou CDC within 4 hours,, and were simultaneously analyzed by both techniques. Table 3 shows that biosensors, together with ML data



processing, generate detection results that are consistent with qPCR results. Note that qPCR provides semi-quantitative results dependent on the Ct value [5], while ML processing of biosensor data only provides qualitative results. This comparative study demonstrates that the ML technique is an effective tool for biosensor signal and data processing,

**Table 3.** Comparison of detection results of inactivated SARS-CoV-2 in vaccination sites of Hangzhou CDC, by both qPCR technique and biosensor with ML technique.

Specimen Collection Location		qPCR Result	Biosensor with ML Result
Vaccination Site 1	Operation Desktop	Weak positive	Positive
Vaccination Site 1	Vaccination Station	Strong positive	Positive
Vaccination Site 2	Operation Desktop	Weak positive	Positive
Vaccination Site 2	Vaccination Station	Weak positive	Positive
Vaccination Site 2	Ventilation Plate	Strong positive	Positive
Vaccination Site 2	Innocation Table Handle	Weak positive	Positive
Vaccination Site 4	Keyboard and Mouse	Negative	Negative
Vaccination Site 5	Pen and White Board	Strong positive	Positive
Vaccination Site 55	Innocation Table Handle	Negative	Negative
No. 4 ans No. 5 Room	Door Handle and Innocation Desk Switch	Negative	Negative
Other	Hemostatic Swab	Weak positive	Positive
Other	Cleaner’s Hand	Negative	Negative

4. Conclusion

In this work, machine learning techniques have been used to process the signals and dataset of photonic biosensors. Both SVM and MLP have been used to process raw data and future engineering data, and perfect results have been obtained to distinguish between negative and positive detections. Control experiments have also been carried out wherein biosensors not functionalized with specific antibodies are used to detect SARS-CoV-2 virus. Both SVM and MLP models trained with valid experimental data cannot distinguish between negative and positive detections in control experiments. To demonstrate the distinguishability of the raw data and the future engineering data for both valid experiments and control experiments, we implemented t-SNE data visualization approach. Results show that the valid experimental dataset is distinguishable, and the control experimental dataset is indistinguishable according to both raw data and features engineering methods. The results are consistent with the data processing performances of machine learning techniques achieved for valid experimental dataset and control experimental dataset. Future research will focus on ML techniques for determination of quantitative detection results so that the quantity of target biospecies in specimen can be obtained. ML can be a powerful tool in processing signals and dataset of biosensors for which there are salient features in the response signals of such biosensors, such as optical, electrochemical, thermal, and mechanical biosensors.

**Acknowledgements:** This research was funded by the Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (grant number 2020R01005), Westlake University (grant number 10318A992001), Tencent Foundation (grant number XHTX202003001), and Zhejiang Key R&D Program (grant number 2021C03002).

## References

1. Tao, Y.D., et al., *A Survey on Data-driven COVID-19 and Future Pandemic Management*. Acm Computing Surveys, 2023. **55**(7).
2. Tenali, N. and G.R.M. Babu, *A Systematic Literature Review and Future Perspectives for Handling Big Data Analytics in COVID-19 Diagnosis*. New Generation Computing, 2023.
3. Duncan, D.B., et al., *Performance of saliva compared with nasopharyngeal swab for diagnosis of COVID-19 by NAAT in cross-sectional studies: a systematic review and meta-analysis*. Clinical Biochemistry, 2022. **109**: p. 117-117.
4. Tng, D.J.H., et al., *Amplified parallel antigen rapid test for point-of-care salivary detection of SARS-CoV-2 with improved sensitivity (vol 189, 14, 2021)*. Microchimica Acta, 2023. **190**(2).
5. Dutta, D., et al., *COVID-19 Diagnosis: A Comprehensive Review of the RT-qPCR Method for Detection of SARS-CoV-2*. Diagnostics, 2022. **12**(6).
6. Larremore, D.B., et al., *Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening*. Science Advances, 2021. **7**(1).
7. He, J., et al., *Rapid detection of SARS-CoV-2: The gradual boom of lateral flow immunoassay*. Frontiers in Bioengineering and Biotechnology, 2023. **10**.
8. Al-Hashimi, O.T.M., et al., *The sensitivity and specificity of COVID-19 rapid anti-gene test in comparison to RT-PCR test as a gold standard test*. Journal of Clinical Laboratory Analysis, 2023.
9. Liu, K.S., et al., *Laboratory detection of SARS-CoV-2: A review of the current literature and future perspectives*. Heliyon, 2022. **8**(10).
10. Chong, Y.P., et al., *SARS-CoV-2 Testing Strategies in the Diagnosis and Management of COVID-19 Patients in Low-Income Countries: A Scoping Review*. Molecular Diagnosis & Therapy, 2023.
11. El-Sherif, D.M., et al., *New approach in SARS-CoV-2 surveillance using biosensor technology: a review*. Environmental Science and Pollution Research, 2022. **29**(2): p. 1677-1695.
12. Abid, S.A., et al., *Biosensors as a future diagnostic approach for COVID-19*. Life Sciences, 2021. **273**.
13. Wei, H.S., et al., *Research progress of biosensors for detection of SARS-CoV-2 variants based on ACE2*. Talanta, 2023. **251**.
14. Rong, G.G., et al., *A high-throughput fully automatic biosensing platform for efficient COVID-19 detection*. Biosensors & Bioelectronics, 2023. **220**.
15. Rong, G.G., et al., *A Closed-Loop Approach to Fight Coronavirus: Early Detection and Subsequent Treatment*. Biosensors-Basel, 2022. **12**(10).
16. Takemura, K., *Surface Plasmon Resonance (SPR)- and Localized SPR (LSPR)-Based Virus Sensing Systems: Optical Vibration of Nano- and Micro-Metallic Materials for the Development of Next-Generation Virus Detection Technology*. Biosensors-Basel, 2021. **11**(8).
17. Zhu, Y.G., W.L. Hu, and Y.T. Fang, *Direct excitation of the Tamm plasmon-polaritons on a dielectric Bragg reflector coated with a metal film*. Opto-Electronics Review, 2013. **21**(3): p. 338-343.
18. Liu, C.D., M.D. Kong, and B.C. Li, *Tamm plasmon-polariton with negative group velocity induced by a negative index meta-material capping layer at metal-Bragg reflector interface*. Optics Express, 2014. **22**(9): p. 11376-11383.
19. Rong, G., Sawan, M., *Rapid Onsite Detection of SARS-CoV-2 with Novel Optical Biosensors and Detection Systems of Handheld and High Throughput Design*. Talanta, 2023. **submitted**.
20. Cao, Y.J., et al., *Efficient Optical Pattern Detection for Microcavity Sensors Based Lab-on-a-Chip*. Ieee Sensors Journal, 2012. **12**(6): p. 2121-2128.
21. Mariani, S., et al., *10 000-Fold Improvement in Protein Detection Using Nanostructured Porous Silicon Interferometric Aptasensors*. Acs Sensors, 2016. **1**(12): p. 1471-1479.
22. Wu, C., et al., *Physical analysis of the response properties of porous silicon microcavity biosensor*. Physica E-Low-Dimensional Systems & Nanostructures, 2012. **44**(7-8): p. 1787-1791.
23. Kotsiantis, S.B., I.D. Zaharakis, and P.E. Pintelas, *Machine learning: a review of classification and combining techniques*. Artificial Intelligence Review, 2006. **26**(3): p. 159-190.
24. Kliegr, T., S. Bahník, and J. Furnkranz, *A review of possible effects of cognitive biases on interpretation of rule-based machine learning models*. Artificial Intelligence, 2021. **295**.
25. Metri-Ojeda, J., et al., *Rapid screening of mayonnaise quality using computer vision and machine learning*. Journal of Food Measurement and Characterization, 2023.
26. Mughaid, A., et al., *A novel machine learning and face recognition technique for fake accounts detection system on cyber social networks*. Multimedia Tools and Applications, 2023.
27. Xu, Y.S., et al., *Machine Learning-Driven APPs Recommendation for Energy Optimization in Green Communication and Networking for Connected and Autonomous Vehicles*. Ieee Transactions on Green Communications and Networking, 2022. **6**(3): p. 1543-1552.
28. Dai, Z.H., R.H. Wang, and J.H. Guan, *Auxiliary Decision-Making System for Steel Plate Cold Straightening Based on Multi-Machine Learning Competition Strategies*. Applied Sciences-Basel, 2022. **12**(22).
29. Fidencio, A.X., C. Klaes, and I. Iossifidis, *Error-Related Potentials in Reinforcement Learning-Based Brain-Machine Interfaces*. Frontiers in Human Neuroscience, 2022. **16**.

30. Brown, J.A., et al., *A Machine Learning System for Supporting Advanced Knowledge Discovery from Chess Game Data*. 2017 16th Ieee International Conference on Machine Learning and Applications (Icmla), 2017: p. 649-654.
31. Hu, J.L., et al., *A hierarchical learning system incorporating with supervised, unsupervised and reinforcement learning*. Advances in Neural Networks - Isnn 2007, Pt 1, Proceedings, 2007. **4491**: p. 403-+.
32. Kumar, B., O.P. Vyas, and R. Vyas, *A comprehensive review on the variants of support vector machines*. Modern Physics Letters B, 2019. **33**(25).
33. Champati, B.B., et al., *Application of a Multilayer Perceptron Artificial Neural Network for the Prediction and Optimization of the Andrographolide Content in Andrographis paniculata*. Molecules, 2022. **27**(9).
34. Fang, X. and M. Ghosh, *High-dimensional properties for empirical priors in linear regression with unknown error variance*. Statistical Papers, 2023.
35. Li, M. and B.Z. Yuan, *2D-LDA: A statistical linear discriminant analysis for image matrix*. Pattern Recognition Letters, 2005. **26**(5): p. 527-532.
36. Valero-Mas, J.J., et al., *Multilabel Prototype Generation for data reduction in K-Nearest Neighbour classification*. Pattern Recognition, 2023. **135**.
37. Mohanty, S.K., et al., *Decision tree approach for fault detection in a TCSC compensated line during power swing*. International Journal of Electrical Power & Energy Systems, 2023. **146**.
38. Kim, T. and J.S. Lee, *Maximizing AUC to learn weighted naive Bayes for imbalanced data classification*. Expert Systems with Applications, 2023. **217**.
39. Meniaillov, I., S. Krivtsov, and T. Chumachenko, *Dimensionality Reduction of Diabetes Mellitus Patient Data Using the T-Distributed Stochastic Neighbor Embedding*. Smart Technologies in Urban Engineering, Stue-2022, 2023. **536**: p. 86-95.
40. Wu, B., et al., *A Nanoscale Porous Silicon Microcavity Biosensor for Novel Label-Free Tuberculosis Antigen-Antibody Detection*. Nano, 2012. **7**(6).
41. Kaliteevski, M., et al., *Tamm plasmon-polaritons: Possible electromagnetic states at the interface of a metal and a dielectric Bragg mirror*. Physical Review B, 2007. **76**(16).
42. van Erven, T. and P. Harremoës, *Renyi Divergence and Kullback-Leibler Divergence*. Ieee Transactions on Information Theory, 2014. **60**(7): p. 3797-3820.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.