THE MICROFLUIDIC MICROWELL ARRAY INTEGRATING SURFACE ENHANCED RAMAN SCATTERING (SERS) PLATFORM ASSISTED WITH MACHINE LEARNING FOR BACTERIA STRAIN IDENTIFICATION

Po-Hsuan Chao[1] and Nien-Tsu Huang[1,2]*
[1]Graduate Institute of Biomedical Electronics and Bioinformatics and
[2]Department of Electrical Engineering, National Taiwan University, TAIWAN

ABSTRACT

We demonstrated a microfluidic microwell array integrating the SERS platform assisting with machine learning to enable bacteria strain identification. The SERS spectra of five *E. coli* strains were collected from 800 independent SERS measurements in one hour and processed by four Machine learning models. The confusion matrix results indicated that five *E. coli* strains treated with ampicillin could be successfully differentiated, even the same DH5α strain between wild type and plasmid-encoded antibiotic resistance. We believe this platform shows great potential to apply the microfluidic technique with ML models to enable rapid and accurate bacteria identification at the strain level.

KEYWORDS

Microfluidics, Surface-enhanced Raman scattering (SERS), Machine learning, bacteria strain identification

INTRODUCTION

Antimicrobial resistance (AMR) is currently a crucial public health issue due to the evolution of bacteria species and the misuse/overuse of antibiotics [1, 2]. To study AMR, it is important to identify specific bacteria species, even to the strain level. To perform bacteria identification, Surface-enhanced Raman spectroscopy (SERS), a label-free and sensitive optical detection method, has been recently utilized as a rapid and sensitive bacteria identification approach. Traditionally, the bacteria identification using SERS spectra analysis was only based on defining the characteristic Raman and SERS spectra peaks of specific molecules. With the assistance of machine learning (ML) methods, SERS-ML can characterize a whole spectra range that may contain more detailed spectra features and capture subtle spectra differences among various sample spectra [3, 4]. In recent years, various machine learning methods have been adopted for spectra-based pathogen strain identification. For example, the classification of Raman spectra of *B. mallei* and related species was demonstrated by the combination of two-level classifiers built by SVM and the data preprocessing by PCA, achieving > 95% prediction accuracy [5].

Although previous research utilized various ML methods to process complicated SERS spectra for bacteria identification, the major bottleneck is to collect a massive, independent, and reliable database to build the training model, which is usually labor-intensive and time-consuming [6]. Besides, specific bacteria subtypes may show similar SERS spectra but respond differently under antibiotic treatment. In such cases, a specific antibiotic stimulation to bacteria may be required before SERS measurement to reveal a subtle spectra difference. To realize this purpose, we developed a microfluidic microwell array integrating SERS substrate to enable a high-throughput bacteria SERS spectra collection. Then, we selected five *E. coli* strains treated with ampicillin, each collecting 800 independent SERS spectra and processing by four ML methods. The confusion matrix results indicated that five *E. coli* strains treated with ampicillin could be successfully differentiated by four ML methods, even the same DH5α strain type between wild type and plasmid-encoded antibiotic resistance. We believe this platform combined with the advantages of microfluidic, SERS, and ML models can enable rapid and accurate bacteria identification at the strain level.

RESULTS AND DISCUSSION

The microfluidic microarray device and operational procedures

The microfluidic microwell array integrating SERS platform is composed of (1) the top microchannel; (2) the bottom microwell array, and (3) the SERS-active substrate, which is a glass slide deposited with a sliver-island film. The operational procedure is shown in Figure 1A. Briefly, to collect the bacteria SERS spectra by mapping the microwell array, the microfluidic microwell-array chips were pre-vacuumed and stored before the sample injection. The sample solution contained 10^8 CFU/mL bacteria culture with or without antibiotic solution and was then loaded into the microfluidic chip through the inlet, filling up the whole channel and covering the microwell array. To properly encapsulate the bacteria inside the microcell array, the device was placed in 37 °C incubation for the 30-min on-chip bacteria incubation. When the bacteria were gradually sedimented to the bottom of the microwell, the washing process was conducted by flowing 30 μL of deionized water (DI) into the device at a 10 μL /min flow rate via the syringe pump. The DI would replace the culture medium in microwells by diffusion while the trapped bacteria remained inside the microwell, establishing a nutrient-insufficient environment for the bacteria. Then, the upper channel layer of the device was lifted, so the lower microwell-array part could become independent microwell chambers in which bacteria secreted metabolites in a nutrient-insufficient environment for 30 min.

SERS mapping

The effectiveness of SERS mapping using the microwell-array device is the major key to generating a massively and uniform dataset for ML analysis. To confirm the SERS mapping quality, we fulfill the *E. coli* sample in the microwell array and used another empty microwell array filled with air as a comparison. The mapping array size was 20 x 20 microwell per scanning round. Figure 2A shows the SERS spectra mapping result of the bacteria encapsulated and empty microwell. The peak at 690 cm^{-1} measured from the empty microwell represented inherent characteristic signals of SERS-substrate spectra. On the other hand, the 733 cm^{-1} peak appeared in the average spectra from *E. coli* encapsulated microwells, which was not founded in empty microwells. To further analyze the spatial distribution of the microwell-array mapped SERS data, the heatmaps of 733 cm^{-1} and 690 cm^{-1} peak intensity from two samples were plotted (Figure 2C). The heatmap of *E. coli* SERS mapping data with randomly distributed intensity was observed, indicating that the spectra difference of independent samples could be captured through all microwell arrays. In addition, a few abnormal points in the heatmap implied the undesired air bubble present in a few microwells. To remove such unwanted SERS data, the 733/690 peak ratio was utilized as the quality check indicator. If the 733/690 peak ratio of *E. coli* generally is below 1.2, we suspect the microwell may not be fully fulfilled. In this case, we remove about 3% of the data and ensure a representative database for following ML analysis.

Broth Microdilution results of five *E. coli* strains

To demonstrate the feasibility of differentiating the same bacteria specie but different strain types, we chose five *E. coli* strains (ATCC 27662, BL21, BW25113, DH5α wild-type (DH5α-WT), and ampicillin-resistant DH5α (DH5α-ampR)) that responded differently to ampicillin treatments and did antimicrobial susceptibility test (AST) by broth microdilution (Figure 3A-F). According to the Clinical & Laboratory Standards Institute (CLSI) guideline, the ampicillin susceptibility of five *E. coli* could be classified into three categories based on the minimum inhibitory concentration (MIC) level: (1) susceptible (S): DH5α-WT and BL21 (MIC < 8 μg/mL); (2) intermediate (I): ATCC 27662 and BW25113 (MIC ~8 μg/mL), and (3) resistant(R): DH5α ampR (MIC > 32 μg/mL) [7]. Although the microdilution result can successfully differentiate five *E. coli* strains, it requires a long incubation time (>16 h), which cannot provide accurate and timely information for proper clinical treatment, such as correct antibiotic type or dosage.

SERS spectra analysis of five *E. coli* strains

To address the above problems, we aim to use SERS spectra of five *E. coli* strains to perform an accurate bacteria strain identification under a simpler experimental setup and shorter time. First, we used the microfluidic microwell array to encapsulate targeted bacteria and collect the corresponding SERS spectra (a total of 3869 spectra in five *E. coli* strains), shown in Figure 4A. The results showed that the SERS spectra of DH5α-WT and DH5α-ampR were quite similar. In addition, BL21 also shared similar spectra shapes to two DH5α strains. To better evaluate the spectra similarity, two linear dimension reduction algorithms, Principal Component Analysis (PCA) and t-distributed stochastic neighbor embedding (T-SNE), were applied to visualize the data distribution pattern. Figure 4B showed that DH5α-WT and DH5α-ampR largely overlapped in the PCA plot, indicating the variations between the two DH5α strains were relatively insignificant. Similar to the PCA plot, two DH5α spectra were also highly clustered in the T-SNE plot (Figure 4C). In addition, BW25113 and ATCC 27662 also share similar patterns but are not largely overlapped from the scatter plot. In summary, these results show that five *E. coli* strains have similar SERS spectra, which cannot be directly differentiated by PCA or T-SNE plots.

Next, we applied ampicillin as the stimulus to treat bacteria and collected the corresponding SERS spectra. In the microfluidic device operation procedure, the bacteria were sedimented in the microwell and incubated with antibiotics for 30 min. Recent studies related to antibiotic inhibition mechanisms have demonstrated that even after 30 min antibiotic treatment, lethality-relevant global metabolic perturbation would occur in bacteria cells, such as increased respiratory activities and more potentially affected metabolic pathways. Many metabolites may contribute to the whole SERS spectra shape, which changes as the antibiotic-induced early state metabolic perturbation occurs. Here, we used 16 μg/mL ampicillin as the targeted antibiotic stimulation level since this concentration can divide *E. coli* into non-resistant (S, I) or resistant (R) types. Once 30-min antibiotic treatment was done, a proper washing, isolation process, and in-situ SERS measurement were processed. Figure 4D shows the ampicillin-treated SERS spectra of five *E. coli* strains. The following PCA and t-SNE plots are shown in Figures 4E and 4F, respectively. Although the PCA scatter plot still indicated an overlapped pattern, the T-SNE plot suggested these antibiotic-treated spectra were more distinguishable compared to the T-SNE plot without antibiotic treatment (Figure 4C).

The confusion matrix results of four ML models

Finally, we analyzed a total of 3869 SERS spectra in five *E. coli* strains using four ML methods (random forest (RF), supporting vector machine (SVM), k nearest neighbor (KNN), and convolutional neural network (CNN)) for bacteria strain identification. For each ML task, the training and testing datasets were prepared by randomly splitting the total dataset at the ratio of 9:1. Figure 5 shows the confusion matrix results of five *E. coli* strains with and without antibiotic treatment. Although, without antibiotic treatment, the confusion matrix results of four ML methods all showed >95% overall accuracy, the misclassification rate between DH5α-WT and DH5α-ampR with at least 8%. Instead, with 30-min antibiotic treatment, the overall accuracy of four ML methods can all reach >97%. Furthermore, the misclassification rate between DH5α-WT and DH5α-ampR was less than 5%. The results show the antibiotic-treated bacteria SERS spectra can improve the bacteria strain identification.

CONCLUSIONS

We demonstrated a microfluidic microwell array

integrating the SERS platform assisting with machine learning to enable an accurate bacteria strain identification. As a proof-of-concept, we chose five *E. coli* strains treated with and without ampicillin. The preliminary result showed that 800 independent SERS measurements could be collected in one hour and processed by four ML models. The confusion matrix results indicated that five *E. coli* strains could be successfully differentiated, which cannot be done by conventional unsupervised PCA and T-SNE plots. We envision this platform can enable rapid and accurate bacteria identification at the strain level.

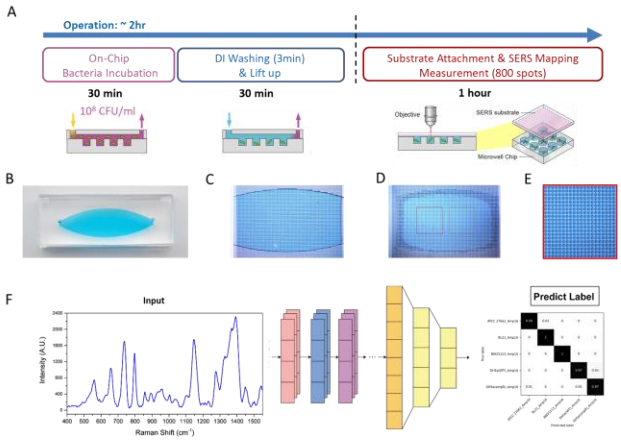


Figure 1: The microfluidic microwell array integrating the SERS platform (A) The operation Procedure. (B-E) The photo of the microwell array device filled with food dye covered (B) with; (C) without the top microchannel;(D) replaced the top microchannel with the glass slide. microchannel lifted. (E) The large view of a 20 x 20 array of the square area in (D). (F) The schematic of bacteria SERS spectra discrimination by ML model.

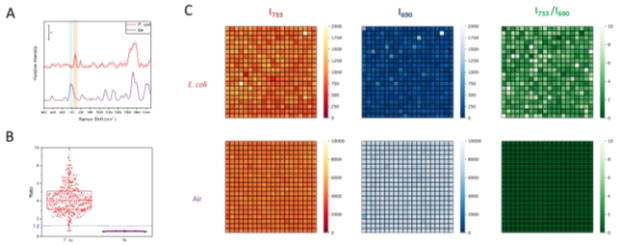
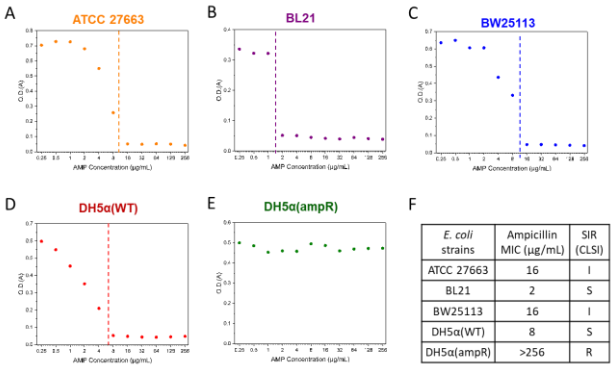


Figure 2: The SERS mapping results of E. coli, DI water, and air in a 20 x 20 microwell array. (A) The average SERS spectra of E. coli DH5α WT (red line) and air (purple line). The light blue and brown regions mark the peaks at 733 cm^{-1} and 690 cm^{-1}. (B) The 733/690 peak ratio of two samples. (C) The SERS heatmaps of the 733 cm^{-1} peak (left), 690 cm^{-1} peak (middle), and 733/690 peak ratio (right) of the three samples. Each pixel represents one microwell.



E. coli strains	Ampicillin MIC (μg/mL)	SIR (CLSI)
ATCC 27663	16	I
BL21	2	S
BW25113	16	I
DH5α(WT)	8	S
DH5α(ampR)	>256	R

Figure 3. The AST results of ampicillin-treated five E. coli strains using broth microdilution. (A) ATCC 27662, (B) BL21, (C) BW25113, (D) DH5α-WT, and (E) DH5α-ampR. The dash lines indicate the long-term-inhibited concentration. (F) The corresponding ampicillin MIC and SIR (Susceptible/Intermediate/ Resistant) are categorized by the CLSI guidelines.

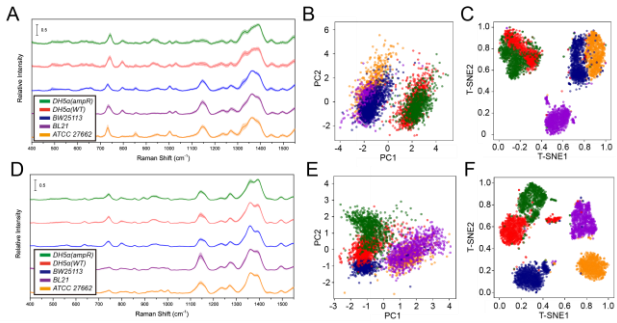


Figure 4. SERS spectra, PCA, and T-SNE plots of five E. coli strains: DH5α-ampR (green), DH5α-WT (red), BW25113 (blue), BL21 (purple), and ATCC 27662 (orange) (A-C) without and (D-F) with 30-min ampicillin-treatment, respectively.

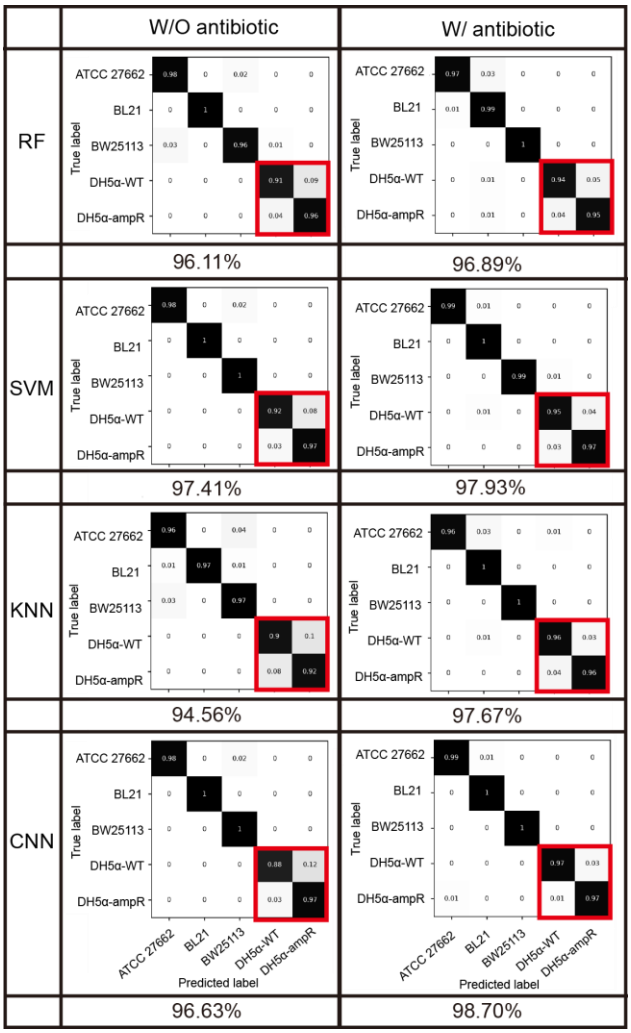


Figure 5. The confusion matrix results of RF, SVM, KNN, and CNN models without and with 30-min ampicillin-treated SERS spectra.

REFERENCES

[1] W.H. Organization, Antimicrobial resistance: global report on surveillance: World Health Organization; 2014.

[2] R.M. Jarvis, R. Goodacre, Characterisation and identification of bacteria using SERS, Chemical Society Reviews, 37(2008) 931-6.

[3] R. Pilot, R. Signorini, C. Durante, L. Orian, M. Bhamidipati, L. Fabris, A review on surface-enhanced Raman scattering, Biosensors, 9(2019) 57.

[4] W. Hu, S. Ye, Y. Zhang, T. Li, G. Zhang, Y. Luo, et al., Machine learning protocol for surface-enhanced raman spectroscopy, The journal of physical chemistry letters, 10(2019) 6026-31.

[5] A.A. Moawad, A. Silge, T. Bocklitz, K. Fischer, P. Rösch, U. Roesler, et al., A machine learning-based Raman spectroscopic assay for the identification of Burkholderia mallei and related species, Molecules, 24(2019) 4516.

[6] C.-S. Ho, N. Jean, C.A. Hogan, L. Blackmon, S.S. Jeffrey, M. Holodniy, et al., Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning, Nature communications, 10(2019) 1-8.

[7] C.a.L.S. Institute., Performance standards for antimicrobial susceptibility testing, 30th ed. CLSI supplement M100., Clinical and Laboratory Standards Institute, Wayne, PA.2020.

CONTACT

*N.T.Huang, tel: +886-958-32862; nthuang@ntu.edu.tw