analytical chemistry

pubs.acs.org/ac Article

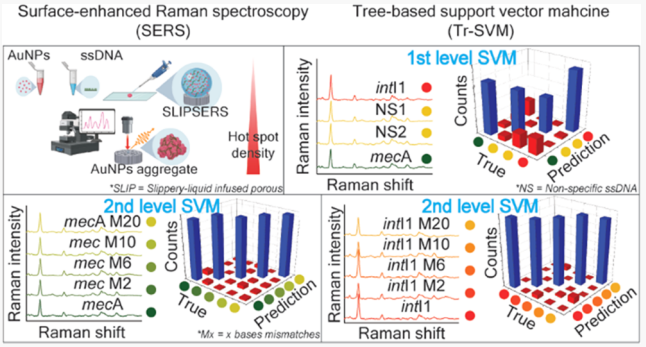
# Discriminatory Detection of ssDNA by Surface-Enhanced Raman Spectroscopy (SERS) and Tree-Based Support Vector Machine (Tr-SVM)

Seju Kang, Inyoung Kim, and Peter J. Vikesland*

Cite This: *Anal. Chem.* 2021, 93, 9319−9328 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** We report label-free detection of 86-base single-stranded DNA (ssDNA) gene segments by surface-enhanced Raman spectroscopy (SERS). The use of a slippery liquid infused porous (SLIP) membrane induced aggregation of 43 nm gold nanoparticles and ssDNA upon pin-free droplet evaporation. The combined SLIPSERS approach generates significant numbers of SERS hot-spots and enabled detection at the 100 nM level of *mec*A and *int*I1 gene segments—two genes of interest in the context of antibiotic resistance. Tree-based multiclass support vector machine (Tr-SVM) classifiers were built to discriminate SERS spectra of 12 different gene sequences obtained by SLIPSERS: *mec*A, *int*I1, as well as analogues of *mec*A and *int*I1, respectively, with 2−10 base mismatches, and two random sequences. The trained predictive Tr-SVM classifiers correctly identified each gene sequence with a prediction accuracy of ∼90%. This study illustrates a novel means for discriminatory label-free SERS detection of ssDNA enabled by Tr-SVM.




The rapid analysis of nucleic acids is of great interest for the characterization of microbial communities and biological function in molecular biology and genetics across numerous science, engineering, and industrial sectors.[1−7] Polymerase chain reaction (PCR)-based approaches are the most widely used techniques to amplify and detect DNA and RNA sequences.[8−10] Over the last two decades, next-generation sequencing (NGS) has rapidly developed as a tool for DNA profiling.[11−13] However, PCR and NGS both require not only centralized research facilities with professional personnel, but are also characterized by high costs and delayed analysis times. These entry barriers limit the widespread application of these tools. Hence, there is a need for alternative analytical approaches for the detection of nucleic acids that are more rapid, more accessible, and more cost effective.

Surface-enhanced Raman spectroscopy (SERS) is a promising candidate as an analytical biosensing tool for nucleic acid detection. Inelastic Raman scattering arising from vibrational modes within molecules provides unique molecular fingerprints.[14] The application of Raman scattering for sensing has historically been limited due to the intrinsic small Raman cross section of molecules. Following the discovery of SERS, the phenomenon whereby the Raman signal is enhanced by a factor of $10^5$−$10^6$ when an analyte is situated adjacent to the surface of a plasmonic metal substrate,[15] interest in Raman-based methods for chemical and biological analysis has rapidly increased. The SERS enhancement factor is determined by the electromagnetic resonance properties of a plasmonic roughened surface or plasmonic nanoparticles. The control of plasmonic gaps within the sub-10 nm regime is critical for SERS enhancement.[16]

Recently, researchers have used SERS to detect double-stranded nucleic acids as well as single-stranded DNA (ssDNA) oligonucleotides.[17−22] In particular, the Bell group has conducted experiments to develop label-free detection of sub-micromolar concentrations of DNA or RNA using Ag colloids with $MgSO_4$ as an aggregating agent.[23,24] In addition, the Vo-Dinh group has synthesized nanoprobes that induce high SERS signals following contact with target DNA.[25,26] They also developed a sandwich approach in which target DNA was hybridized to both metal nanoparticles and magnetic beads functionalized with complementary oligonucleotides for SERS enrichment.[27]

Herein, we applied the slippery liquid infused porous SERS (SLIPSERS) method developed by Wong and colleagues[28] for the label-free detection of ssDNA. A SLIP membrane was prepared by infusing the hydrophobic perfluorinated lubricant,

**Received:** October 29, 2020
**Accepted:** June 16, 2021
**Published:** July 1, 2021







https://doi.org/10.1021/acs.analchem.0c04576
*Anal. Chem.* 2021, 93, 9319−9328

Table 1. Gene Sequences for *mec*A, *int*I1, four mismatched analogues for *mec*A and *int*I1, and nonspecific (NS)1 and 2[a]

| | 5′- sequences -3′ |
|---|---|
| *mec*A | TGGTGAAGTTGTAATCTGGAACTTGTTGAGCAGAGGTTCTTT |
| | TTTATCTTGGGTTAATTTATTATATTCTTCGTTACTCATGCCAT |
| *mec*A-M2 | TGGTGAAGT**G**GTAATCTGGAACTTGTTGAGCA**C**AGGTTCTTT |
| | TTTATCTTGGGTTAATTTATTATATTCTTCGTTACTCATGCCAT |
| *mec*A-M6 | TGGT**T**AAGTT**C**TAATCTG**A**AACTTGTTT**T**AGCA**C**AGGTTC**C**TTT |
| | TTATCTTGGGTTAATTTATTATATTCTTCGTTACTCATGCCAT |
| *mec*A-M10 | TGGT**TAC**GTT**C**TAATCT**CA**AACTT**C**TTT**T**AGCA**C**AGG**A**TC**C**TTT |
| | TTATCTTGGGTTAATTTATTATATTCTTCGTTACTCATGCCAT |
| *mec*A-M20 | T**C**GTT**TC**G**G**T**CC**AAG**C**T**CA**AACT**AC**TTT**G**G**T**A**C**AT**G**A**T**C**CG**TT |
| | TTATCTTGGGTTAATTTATTATATTCTTCGTTACTCATGCCAT |
| *int*I1 | GTGCACGGGCATGGTGGCTGAAGGACCAGGCCGAGGGCCGCA |
| | GCGGCGTTGCGCTTCCCGACGCCCTTGAGCGGAAGTATCCGCGC |
| *int*I1-M2 | GTGCACGGG**A**ATGGTGGCTGAAGGACCAGG**T**CGAGGGCCGCA |
| | GCGGCGTTGCGCTTCCCGACGCCCTTGAGCGGAAGTATCCGCGC |
| *int*I1-M6 | GT**A**CAC**A**GGCATT**G**TGGCTGAAG**A**ACCAG**A**CCGAGG**A**CCGCA |
| | GCGGCGTTGCGCTTCCCGACGCCCTTGAGCGGAAGTATCCGCGC |
| *int*I1-M10 | G**A**GCA**G**GC**G**A**A**TGGTGG**T**TGAA**C**GAC**T**AA**G**TCGAGGGC**T**GCA |
| | GCGGCGTTGCGCTTCCCGACGCCCTTGAGCGGAAGTATCCGCGC |
| *int*I1-M20 | G**A**CCATG**C**CC**G**TG**A**TG**AG**TCAATA**A**CA**A**TA**C**TG**A**C**G**G**A**T**G**G**A**G |
| | CGGCGTTGCGCTTCCCGACGCCCTTGAGCGGAAGTATCCGCGC |
| NS1 | ATGGCATGAGTAACGAAGAATATAATAAATTAACCCAAGATAA |
| | AAAAAGAACCTCTGCTCAACAAGTTCCAGATTACAACTTCACCA |
| NS2 | ACGAAGAAGAATATAATAAATTAACCCAGGAAGCGCAACGCC |
| | GCTGCGGCCCTCGGCCTGGTCCTTCAGCCACCATGCCCGTGCAC |

[a]Mismatched bases are **bolded**.

perfluoropolyether (PFPE), into a poly(tetrafluoroethylene) (PTFE)-lined porous membrane by simple centrifugation. When a water droplet containing plasmonic nanoparticles and ssDNA evaporates on the SLIP membrane, SERS hot-spots are generated upon sample aggregation without the coffee ring effect.[28−30] We synthesized gold nanoparticles (AuNPs) for the SERS application and tested two gene segments: *mec*A and the class 1 integron-integrase *int*I1. *mec*A is a representative antibiotic resistance gene (ARG) that encodes resistance to methicillin and other $\beta$-lactam antibiotics.[31,32] It is commonly found within methicillin-resistant *Staphylococcus aureus* (MSRA)—one of the most widespread antibiotic-resistant pathogens. *int*I1 is a constituent of integrons in plasmids that facilitate horizontal gene transfer between bacterial species through bacterial conjugation. Previous studies have shown that the concentration of *int*I1 correlates with the total concentration of ARGs.[33,34] Both *mec*A and *int*I1 are potential indicators of antibiotic resistance in clinical and environmental samples.

A SERS spectrum provides a unique molecular fingerprint allowing molecule identification; however, the similarities of SERS spectra between gene sequences remain a challenge for discrimination. While the Bell and Ren groups reported label-free SERS detection of DNA with single-base sensitivity,[35,36] their discriminations were limited to a direct comparison of peak ratios or to simple multivariate statistical approaches, such as the unsupervised learning tool principal component analysis (PCA), that cannot be used to evaluate prediction accuracy for high dimensional and correlated datasets. Although supervised classification tools such as parametric discriminant analysis can be used to evaluate prediction accuracy, they were built under strong distribution assumptions such as normality and a specific covariance structure that are not satisfied by SERS spectral data. More powerful classifiers are required to achieve a higher discriminatory capacity of SERS spectra.

We built a flexible discriminatory tool by combining a tree-based decision rule and a multiclass support vector machine (Tr-SVM) for the identification of gene sequences based on their SERS spectra. A tree-based decision rule groups correlated classes and provides multiple classifiers based on decision levels. Each classifier can be better optimized to the dataset of each group than a single classifier. SVM is a machine learning technique that differentiates multidimensional data by a separating hyperplane.[37,38] The optimal hyperplane maximizes the margin of the data between different classes. SVM has been shown to perform well in a variety of settings including omics and is often considered among the best "out of the box" classifiers. It also presents one of the most robust prediction methods.[39] SVM can digest high dimensional data containing several spectral features from SERS spectra. The flexible functionality of using different kernel functions (e.g., linear, polynomial, Gaussian) to define the hyperplane boundary enables discrimination of complex data structures. For instance, SVM was used to discriminate the Raman spectra of *Legionella* species from those of other aquatic bacteria.[40] It was found that the trained model correctly predicted nine *Legionella* bacteria at the species level with an accuracy of ∼86%.

Herein, we sought to combine SLIPSERS-enabled label-free detection of ssDNA with the development of a functional predictive model unifying a tree-based decision procedure and a multiclass SVM classifier. As shown, this combined approach provides the ability to discriminate SERS spectra arising from multiple gene sequences.


https://doi.org/10.1021/acs.analchem.0c04576
Anal. Chem. 2021, 93, 9319−9328

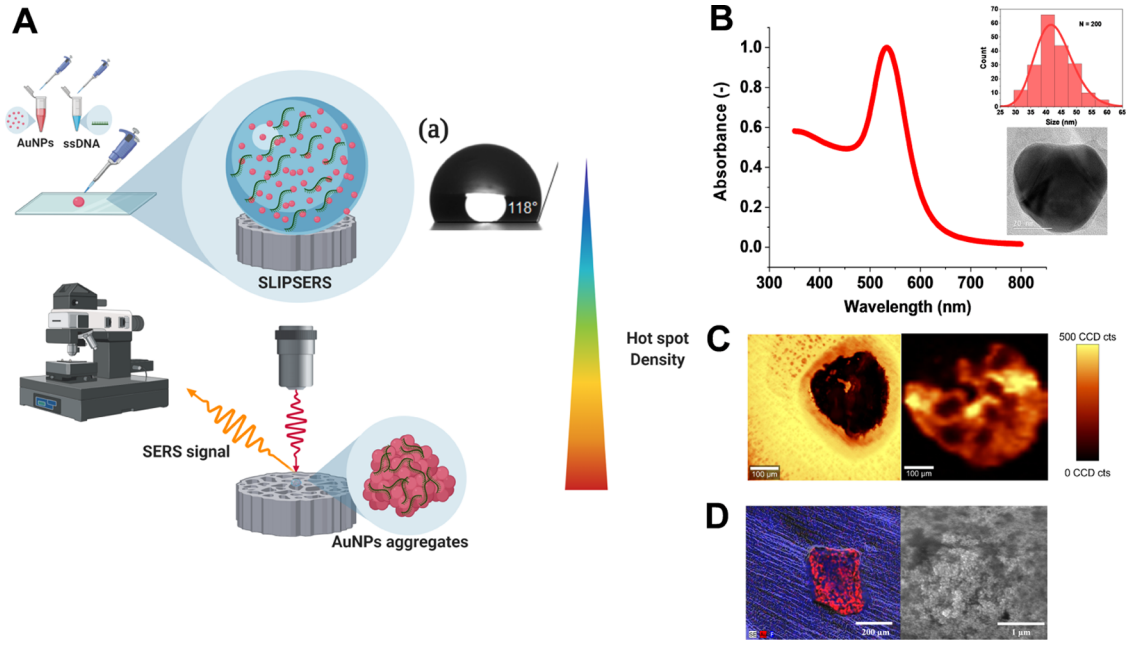



**Figure 1.** (A) Schematic illustration of the preparation of the SLIPSERS substrate and SERS spectra collection using Raman spectroscopy. The inset (a) shows the image of the droplet on the membrane with the contact angle of 118°. (B) Ultraviolet−visible (UV−vis) absorbance spectrum of as-synthesized AuNPs with the peak at 533 nm and insets show a TEM picture and size distribution of AuNPs ($N$ = 200) measured by the ImageJ software. (C) Microscopic image of the aggregate of AuNPs and ssDNA (left) and the SERS map at 77 $cm^{-1}$ of the same substrate (right). (D) SEM-EDS element mapping picture of the SLIPSERS substrate on the PTFE-lined membrane (left) and the SEM picture of aggregation of AuNPs (right).

## EXPERIMENTAL SECTION

**Materials.** Gold chloride trihydrate ($HAuCl_4{\cdot}3H_2O$), sodium citrate tribasic dihydrate ($Na_3Cit{\cdot}2H_2O$), and nuclease-free water were purchased from Sigma-Aldrich (St. Louis, MO). A PTFE-lined membrane with a 2 $\mu$m pore size was purchased from SKC Inc. (Eighty Four, PA). The perfluorinated liquid lubricant PFPE (Dupont Krytox GPL 103) was purchased from Zoro Tool Inc. (Buffalo Grove, IL). Prior to use, all glassware was cleaned with aqua regia—3:1 (v/v) hydrochloric acid (HCl)/nitric acid ($HNO_3$).

**Gene Sequence Design.** We tested two different ssDNA oligonucleotides: *mec*A and *int*I1. Sequences for *mec*A and *int*I1 were obtained from the National Center for Biotechnology Information (NCBI) database using the Basic Local Alignment Search Tool (BLAST). Briefly, forward and reverse primers matching each type of ssDNA were identified from previous studies.[41,42] BLAST then identified the best-fit sequences with complementary genes to two primers. Using the identified sequences for *mec*A and *int*I1, we selected 86-base-length gene segments for each. Second, for the development of the discriminatory and predictive model with high base sensitivity, we tested four analogues of the *mec*A and *int*I1 gene segments with 2, 6, 10, and 20 base mismatches. These are annotated as *mec*A or *int*I1-M2, -M6, -M10, and -M20. In addition, we tested two randomly designed sequences that we refer to as nonspecific 1 and 2 (NS1 and NS2). The sequences are provided in Table 1. Further, to investigate the effect of sequence differences with the same base composition and a single-base mismatch on the SERS spectra, we also tested an *int*I1 gene segment with two base position alternations and an *int*I1 gene segment with one base mismatch (*int*I1-M1).

All oligonucleotides were purchased from Integrated DNA Technologies Corp. (Coralville, IA) with the request of polyacrylamide gel electrophoresis (PAGE) purification. Following arrival in the solid state, the oligonucleotides were rehydrated using nuclease-free water and stored at −20 °C prior to use. The initial concentration of ssDNA was quantified using a Qubit fluorometer.

**Synthesis and Characterization of the SLIPSERS Substrate.** We synthesized citrate-coated AuNPs using a seed-mediated growth method reported previously.[43] To synthesize the seed solution, 1 mM $HAuCl_4{\cdot}3H_2O$ was heated to boil and 3.88 mM $Na_3Cit{\cdot}2H_2O$ was added as a reducing agent. After 30 min of stirring, the solution was cooled and filtered through a 0.22 $\mu$m PTFE filter. Then, a 100 mL aliquot of 0.254 mM $HAuCl_4{\cdot}3H_2O$ was heated to boil followed by the addition of 0.85 mL of seed solution and 0.44 mL of 38.8 mM $Na_3Cit{\cdot}2H_2O$. The mixture was refluxed for 30 min and cooled to room temperature. As-synthesized AuNPs were characterized using a Cary 5000 UV−Vis−NIR spectrophotometer, a JEOL 2100 transmission electron microscope (TEM), and a Zetasizer Nano ZS dynamic light scattering (DLS) instrument. The final product was stored at 4 °C prior to use.

Figure 1A illustrates the preparation of the SLIPSERS substrate and a SERS measurement. To make the SLIP membrane, we followed the protocol of Wong et al. with minor changes.[28] A PTFE-lined porous membrane was immersed in PFPE and then subjected to centrifugation at 100 × $g$ for 1 min to remove excess fluids. The SLIP membrane was then transferred onto a glass slide. Then, 5 $\mu$L of as-synthesized 0.1 nM AuNPs with 0.2% glycerol and 100 nM ssDNA was pipetted and mixed thoroughly on the SLIP membrane. Prior to addition to the membrane, the samples were preheated to 95 °C for 30 s to ensure oligonucleotide extension.[36] Glycerol maintains the wet state of the three-dimensional (3D) aggregate matrix and prevents the quenching of SERS hotspots.[44,45] The contact angle of the droplet on the SLIP


https://doi.org/10.1021/acs.analchem.0c04576
Anal. Chem. 2021, 93, 9319−9328

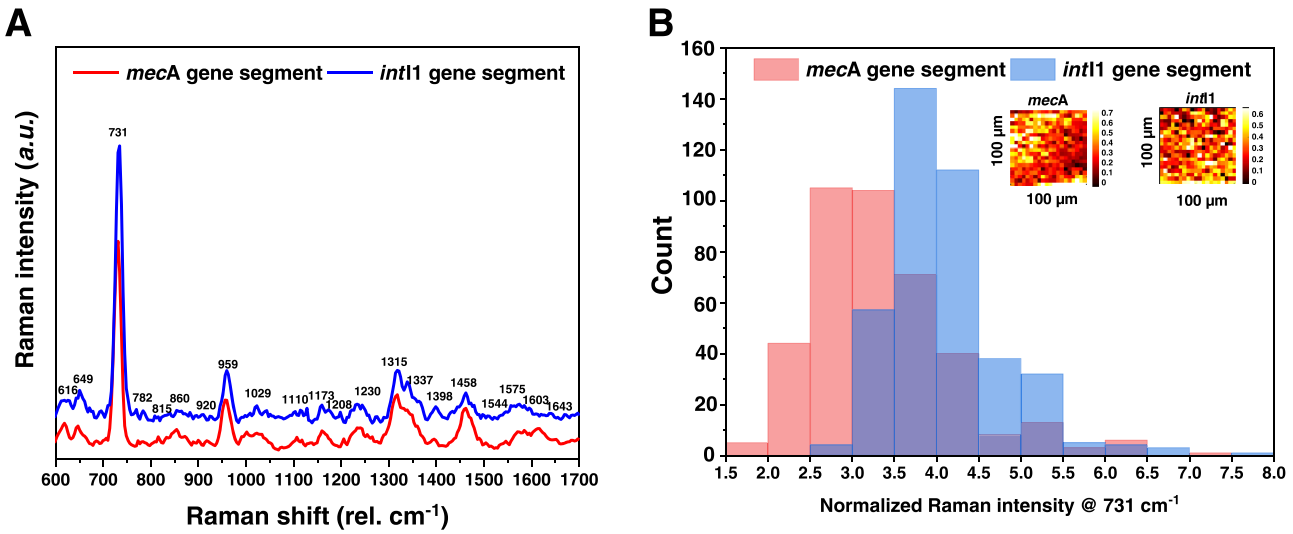



**Figure 2.** (A) Average of 400 SERS spectra for 100 nM *mec*A and *int*I1 gene segments in the range from 600 to 1700 $cm^{-1}$. (B) Histogram of the normalized Raman intensity at 731 $cm^{-1}$ for 100 nM *mec*A and *int*I1 gene segments. Insets show the spatial normalized intensity distribution across the scanning area.

membrane was measured using a ramé-hart Model 250 standard goniometer. The droplet was then evaporated at 60 °C. After evaporation, the 3D aggregate of AuNPs and the analytes (i.e., the SLIPSERS substrate) was a visible black dot. The SLIPSERS substrates were characterized by FEI environmental scanning electron microscopy (SEM) and Bruker energy dispersive spectroscopy (EDS).

**Instrumentation.** SERS spectra of the target ssDNA on the SLIPSERS substrate were collected using a WITec Alpha500R Raman spectrometer with a 785 nm laser and a 10× confocal microscope objective. A 300 gr/mm grating was used and the spectral center was set to 1500 $cm^{-1}$. Following microscope focusing, 20 × 20 ($X \times Y$) points were measured across a 100 × 100 $\mu m^2$ area using a 0.1 s integration time per point. SERS peaks were selected using automated peak labeling within the WITec Control Five (v.5.0) software with the minimum relative height set to 0.1. The baseline was corrected using an asymmetric least squares smoothing method.

**Evaluation of the Tree-Based Multiclass SVM (Tr-SVM) Classifiers.** We prepared 12 different gene sequences (i.e., *mec*A and *int*I1, four analogues with 2−20 base mismatches for each, and two nonspecific genes) to collect our SERS spectral dataset and evaluated their prediction powers using the Tr-SVM classifiers. Once the ssDNA was subjected to SLIPSERS, 400 (20 × 20) SERS spectra were collected per gene sequence using the area scan method. Within the spectra, 21 spectral features in the range of 600−1700 $cm^{-1}$ labeled by the WITEC Control Five (v.5.0) software were extracted after baseline correction. To account for point-to-point variability, the extracted features were normalized by the peak at 959 $cm^{-1}$ corresponding to the vibrational mode of deoxyribose.[17] A total of 4800 spectra from 12 gene sequences were used to develop the discriminatory SVM classifier. Prior to the development of the predictive model, we designed a tree-based classification system with two decision levels to reduce the complexity of discrimination. At the first decision level, considering the similarity in base composition between the 12 gene sequences, the *mec*A or *int*I1 gene segments and their four base mismatch analogues were considered as two groups (*mec*A and *int*I1 groups). The difference in base composition for the *mec*A and *int*I1 gene segments within each group ranged from 2.3 to 16.3 and 4.7 to 23.3%, respectively, while that for NS1 and NS2 varied from 60.5 to 67.4% and 32.6 to 74.4%, respectively. At this decision level, the SVM classifier was trained to discriminate four gene groupings (i.e., *mec*A and *int*I1 groups, NS1, and NS2). At the second decision level, two models were separately trained to discriminate the SERS spectra of five *mec*A and five *int*I1 species independently (*mec*A, *mec*A-M2, *mec*A-M6, *mec*A-M10, *mec*A-M20; *int*I1, *int*I1-M2, *int*I1-M6, *int*I1-M10).

To discriminate the 12 gene sequences, a multiclass SVM classifier with a binary class tool was trained using the classifier learner application in Mathworks MATLAB/SIMULINK (v.R2019b). To illustrate, for three or more classes, multiple binary classifiers were further evaluated for every possible pair of classes and the most voted class was predicted. When training was completed, the performance of the SVM classifier was evaluated in terms of the true positive rate, true negative rate, false-positive rate, and false-negative rate. These results were reported and visualized using a confusion matrix. With these rates, the SVM classifier was exhaustively evaluated in terms of accuracy, sensitivity, and specificity. Accuracy indicates how the model correctly identifies and excludes a given condition. Sensitivity and specificity evaluate how the classifier correctly predicts the positive and rules out the negative result, respectively. Ten-fold cross-validation was conducted to validate the trained model. A total of 4800 spectra were divided into 10 subsets of equal size. Only one of the subsets was subjected to the test and the rest were used to train the classifier. The procedure was iterated until every subset was used for testing. Instead of leave-one-out cross-validation (LOOCV), 10-fold cross-validation has a computational advantage over LOOCV while providing more accurate estimates of the test error rate because it uses less correlated trained sets. In addition, the iterative estimation of the prediction on unseen data by cross-validation reflects how well the model generalizes to unseen data.


https://doi.org/10.1021/acs.analchem.0c04576
Anal. Chem. 2021, 93, 9319−9328

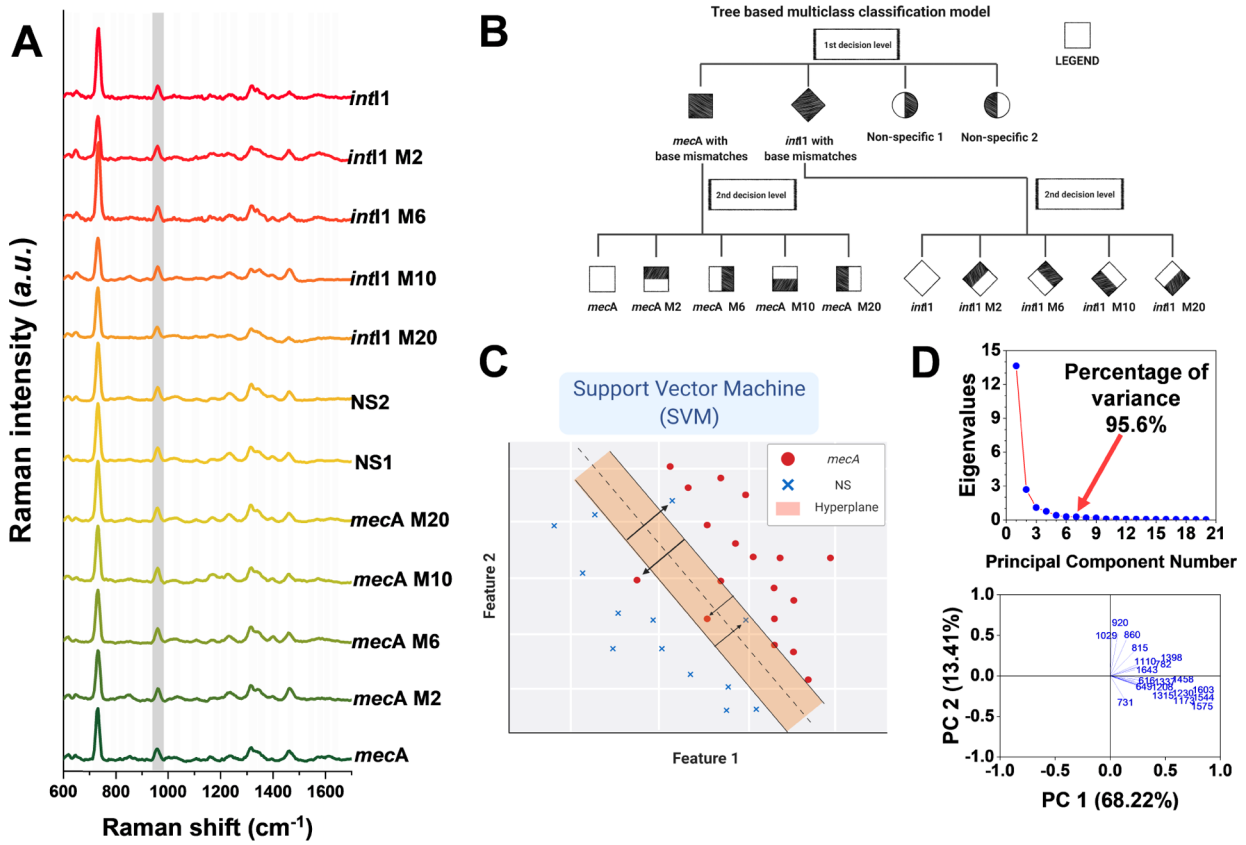



**Figure 3.** (A) SERS spectra for 12 different gene sequences normalized by the 959 $cm^{-1}$ peak (dark gray region) stacked vertically. The distinct peaks in the gray region were identified as components of ssDNA. (B) Tree-based multiclass classification model at first (*mec*A gene segment and its analogues vs *intI*1 gene segment and its analogues vs nonspecific 1 vs nonspecific 2) and second decision levels (*mec*A gene segment vs its analogues; *intI*1 gene segment vs its analogues). (C) Schematic illustration of SVM with the hyperplane to differentiate two gene sequences. (D) Eigenvalues of principal components with the number of principal components (top) and the PCA plot for the selected features (bottom).

## RESULTS AND DISCUSSION

**Development of the SLIPSERS Substrate.** The as-synthesized AuNPs exhibited a distinct plasmonic property with a characteristic localized surface plasmon resonance (LSPR) peak at ~533 nm.[46,47] The AuNPs were spherical and had a normally distributed, TEM determined mean diameter of 42.9 ± 6.2 nm (Figure 1B). The electrophoretic mobility (EM) and ζ-potential of AuNPs suspended in 50 μM $Na_3Cit$ solution were −3.1 (±0.1, $n$ = 3) × $10^{-8}$ $m^2$/Vs (= −39.1 mV (±0.8)) and were consistent with other citrate-coated AuNPs.[48]

To produce a SLIPSERS substrate, 10 μL of a droplet containing AuNPs and a given ssDNA analyte was deposited on the SLIP membrane. The contact angle of the droplet was measured by a goniometer to be 118° at room temperature (Figure 1A,a). The extreme hydrophobicity of the SLIP membrane enables formation of a pin-free liquid contact line.[28] The low surface tension of the droplet enhances the mixing of AuNPs and gene sequences during droplet evaporation and results in the formation of dense SERS hot-spots. After complete evaporation, the 3D aggregates of AuNPs and ssDNA analytes were observed by light microscopy to have dimensions of ~200 × 200 $\mu m^2$ and arbitrary shapes (Figure 1C, left). A SERS map produced based upon the intensity of the band at 77 $cm^{-1}$ shows the ubiquity of SERS hot-spots across the substrate (Figure 1C, right). Previous studies have shown that the density of SERS hot-spots is reflected by the intensity of the pseudo-band at 77 $cm^{-1}$.[49−52] SEM-EDS images indicate closely packed AuNPs with a porous structure on top of the PTFE-lined membrane (Figure 1D).

**SERS Detection of ssDNA.** We observed several distinct peaks between 600 and 1700 $cm^{-1}$ in the SERS spectra of the *intI*1 and *mec*A gene segments (Figure 2A). While these spectra are similar in terms of the locations of the peaks at 731, 959, 1315, and 1458 $cm^{-1}$, the relative peak intensities differ. ssDNA consists of a phosphate backbone (deoxyribose and phosphate groups) and four base nucleotides: adenine (A), cytosine (C), guanine (G), and thymine (T). Twenty-one separate spectral features were selected via automated peak labeling within the WITec Control Five (v.5.0) software. Most of the featured peaks arise from the aromatic ring breathing mode of the base nucleotides: 616 (A),[36] 649 (G),[36,53] 731 (A),[53] 782 (C,T),[54] 1029 (C),[24] 1173 (C,G),[55] 1208 (A,T),[56] 1230 (A,T),[54] 1315 (G),[55] 1337 (A,G),[56] 1544 (G),[57] 1575 (A,G),[35,58] 1603 (C),[57] and 1643 (C,T)[35] $cm^{-1}$. Additionally, the 815/860,[59,60] 1110,[35] and 1230[54] $cm^{-1}$ peaks arise from the symmetric bend, symmetric stretch, and asymmetric stretch modes of $PO_2^{-1}$. The 959[53] and 1458[55] $cm^{-1}$ peaks reflect deoxyribose ring symmetric stretch and bend. Some minor peaks such as the C−N stretch (920 $cm^{-1}$) and $CH_2$ deformation (1398 $cm^{-1}$) were also detected.[59,61]

While some portions of the SERS spectra in our study are similar to those in previous studies on label-free ssDNA detection (e.g., the strongest peak from the adenine breathing mode at 731 $cm^{-1}$), some peaks did not match as well.[17,23,24,35,36,62] This can be attributed to the different


https://doi.org/10.1021/acs.analchem.0c04576
Anal. Chem. 2021, 93, 9319−9328

conformational states of ssDNA to AuNP that are dependent on sample preparation. For instance, the spectra in our study exhibit a high-intensity band at 959 $cm^{-1}$ that arises from deoxyribose. Xu et al. did not report a peak at 959 $cm^{-1}$ and instead report a strong peak at 1087 $cm^{-1}$ arising from the $PO_2^{-1}$ stretch.[35] We suggest that such differences in the SERS profiles reflect how ssDNA and the plasmonic substrates interact and that localized SERS hot-spots enhance signals from different locations within ssDNA. Xu et al. used Ag colloid with $MgSO_4$ as an aggregating agent. $Mg^{2+}$ has a charge shielding effect that increases the interaction between $PO_2^{-1}$ and AgNPs. Our method, pin-free evaporation-induced AuNP aggregation, should not enhance such charged interactions so it is more likely that charge-neutral deoxyribose interacts with AuNP than does $PO_2^{-1}$. Unlike $MgSO_4$-induced aggregation, SLIPSERS primarily induces nonspecific sorption of ssDNA. This tendency is reflected in the weaker peak intensities of the $PO_2^{-1}$-related peaks than those from deoxyribose. To test this hypothesis, we compared the SERS profiles of ssDNA prepared by different approaches (i.e., SLIPSERS, $MgSO_4$-induced aggregation, and thiol-linked ssDNA; Figure S1). As expected, we found that there was a significant difference in the SERS spectra across the different sample preparation approaches. The replicability of our sample analysis approach was tested by the preparation of duplicate samples. As illustrated in Figure S2, there was no discernable difference in the collected spectra, thus illustrating the replicability of our approach.

Figure 2B shows the histogram of the normalized peak at 731 $cm^{-1}$ for both *mec*A and *int*I1 gene segments across the scan area. The inset illustrates that the intensities were evenly distributed across the scan area. The coefficients of variation (CV) of the normalized intensity at 731 $cm^{-1}$ were 24.9 and 16.4% for *mec*A and *int*I1, respectively. It has been reported that CVs for many commercial SERS substrates are >45%.[63] Unless they are rigorously designed by high-cost fabrication techniques, signal reproducibility is often a challenge in SERS due to the heterogeneous aggregation of nanoparticles within different SERS substrates.[64] The low CVs for the SLIPSERS samples indicate that the method has sufficiently high spatial reproducibility. Furthermore, the results imply that the conformational and adsorption states of ssDNA to the Au surface are reproducible across different SLIPSERS substrates.

**Development of the Discriminatory Tool Based on SVM.** SERS spectra of different gene sequences of the same length are expected to differ due to the varying sequences of bases and the associated changes in oligonucleotide conformation. The Bell group investigated the effect of base position and content on the SERS spectra of ssDNA, concluding that SERS only provides information on a simple sum of the signals from ssDNA components but not the primary sequence of bases.[24] Li et al. recently reported that ssDNA can potentially form random coils or various hairpin conformations thus affecting the SERS spectrum.[53] The formation of Watson−Crick hydrogen bonds upon folding of ssDNA induces A−T and G−C pair formation on the surface of the substrate, thus resulting in the enhanced intensity of peaks from A and G. Such localized enhancement of certain peaks makes the SERS spectra of ssDNA highly complex. Comparing one or two peaks and/or their ratios within the collected spectra is insufficient if one wants to elucidate the possible formation of hairpins or coils of ssDNA and their consequences on the spectra. To rigorously differentiate SERS spectra for different gene sequences, we collectively extracted all labeled peaks from the spectra and subjected them to the SVM technique. The SERS spectra of 12 different gene segments are vertically stacked in Figure 3A. All gene segments had 86 bases and only varied in total base composition and sequence. As expected, due to the different sequences and total base contents, we found slightly different peak intensities in the SERS spectra but consistent peak positions. For comparative purposes, we tested *int*I1 segments with lengths of 43, 65, 86, 107, and 129 nts. These segments had highly similar SERS profiles that could not be differentiated by eye (Figure S2). This consistency indicates that the differences observed in Figure 3A primarily arise due to differences in base composition and sequence.

A key challenge in the development of a predictive model based upon spectral data is to standardize how spectral features are defined such that they reflect all possible differences resulting from gene sequence variation. If all spectral features, including baseline noise, are used during model learning, the developed model can be unnecessarily biased such that it does not solely reflect gene classification. Over the spectral range of 600−1700 $cm^{-1}$, at a resolution of ∼4 $cm^{-1}$, a total of 302 spectral features could be extracted. PCA of these 302 spectral features resulted in the selection of >200 principal components (PCs; Figure S3). This large number of PCs implies that intensities close to the baseline (i.e., spectral noise) can be expected to be involved in model training and generate feature clusters prone to overfitting.[65] Additionally, different background correction methods can generate significant differences in noise-like intensities that bias the resulting models. We suggest that instead of processing all data points within the SERS spectrum that computer-aided selection of distinct peaks defined by standards such as the relative minimum height helps avoid potential model biases. In Figure 3A, the peak at 959 $cm^{-1}$ (dark gray), arising from the deoxyribose vibration, was used to normalize the remaining 20 peaks (gray). These normalized spectral features were extracted to develop our predictive models. Peak selection is dependent on the defined minimum relative height for automated peak labeling within the WITec Control Five (v.5.0) software.

Multiclass SVM classifiers were built using the normalized 20 spectral features. To predict 12 gene sequences, we built Tr-SVM classifiers at the first and second tree-decision levels using a tree-based classification system (Figure 3B). At the first tree-decision level, *mec*A or *int*I1 gene segments and their four base mismatch analogues (i.e., *mec*A-M2, *mec*A-M6, *mec*A-M10, *mec*A-M20 or *int*I1-M2, *int*I1-M6, *int*I1-M10, *int*I1-M20) were separated into two groups and one SVM classifier was built for the combined *mec*A group, the combined *int*I1 group, NS1, and NS2. At the second tree-decision level, two SVM classifiers were built for *mec*A or *int*I1 gene segments and their analogues. Given the dataset, SVM finds the optimal hyperplane that maximizes the margin between classes (Figure 3C). The kernel function was applied to project the data to a high dimensional space where they can be separated. SVM was chosen over other classification techniques because it not only includes many kernel functions, providing flexibility in the determination of the decision boundary (i.e., hyperplane) between classes, but it is also effective in high dimensional data matrices. We tested both polynomial (i.e., linear, quadratic, and cubic) and different scale Gaussian (i.e., fine, medium, and coarse) kernel functions to find the best-score classification algorithm. In addition, we also compared the SVM classifiers to parametric-based discriminant analyses, including linear


https://doi.org/10.1021/acs.analchem.0c04576
Anal. Chem. 2021, 93, 9319−9328

Table 2. Prediction Results via the PCA-SVMs with Highest Accuracy for the First and Second Decision Levels

| first decision level | | | | | second decision level | | | | |
|---|---|---|---|---|---|---|---|---|---|
| kernel function | samples | accuracy (%) | sensitivity (%) | specificity (%) | kernel function | samples | accuracy (%) | sensitivity (%) | specificity (%) |
| fine Gaussian | *mecA* | 88.7 | 89.0 | 90.4 | quadratic | *mecA* | 90.8 | 88.0 | 97.8 |
| | *intI1* | | 93.0 | 94.9 | | *mecA*-M2 | | 93.3 | 95.9 |
| | NS1 | | 83.8 | 99.0 | | *mecA*-M6 | | 89.3 | 98.2 |
| | NS2 | | 68.5 | 98.0 | | *mecA*-M10 | | 91.8 | 98.6 |
| | | | | | | *mecA*-M20 | | 91.5 | 97.9 |
| | | | | | quadratic | *intI1* | 94.4 | 96.3 | 98.5 |
| | | | | | | *intI1*-M2 | | 93.3 | 97.1 |
| | | | | | | *intI1*-M6 | | 94.5 | 99.0 |
| | | | | | | *intI1*-M10 | | 96.8 | 99.7 |
| | | | | | | *intI1*-M20 | | 91.3 | 98.7 |

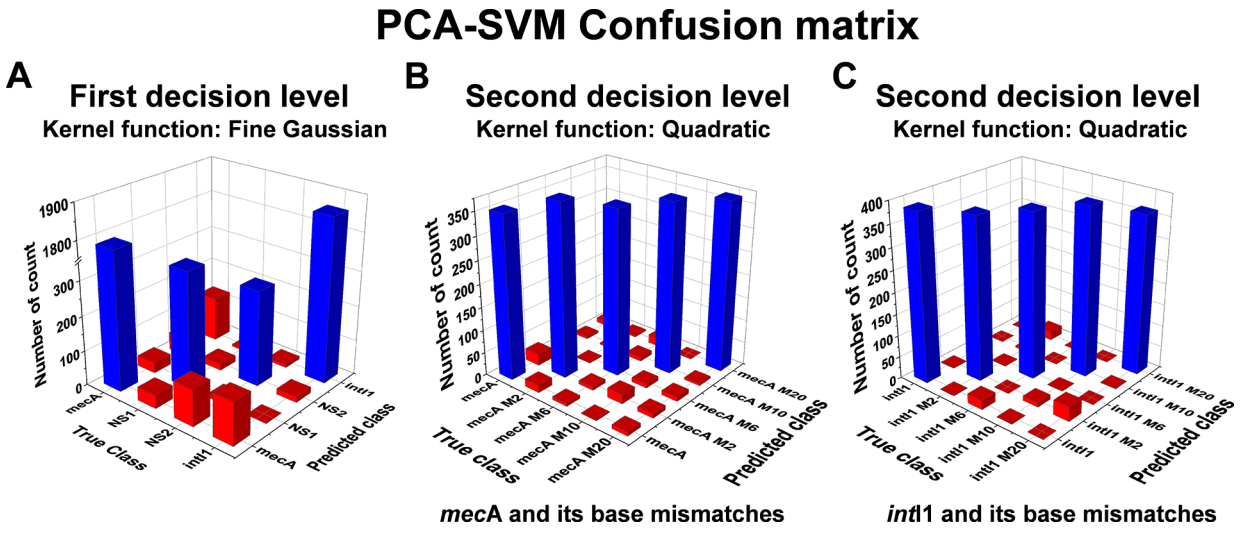



Figure 4. PCA-SVM confusion matrix results for first (left) and second (middle and right) decision levels. *X*, *Y* axes show true and predicted class, while the *Z* axis shows the number of counts for each case. Blue and red colored columns indicate the correct and wrong prediction counts, respectively.

discriminant analysis (LDA) and quadratic discriminant analysis (QDA), to investigate the efficacy of SVM for SERS discrimination of ssDNA.

PCA was preconducted to reduce the dataset dimensionality. A prior study has illustrated the effect of a pre-PCA step on the accuracy of different SVM models for microarray data for colon cancer.[66] It was reported that PCA increased model performance in terms of accuracy and running time. Figure 3D shows the results of the PCA analysis of 20 spectral features. Most of the spectral features were closely clustered and seven significant principal components (PCs) were determined with an explained variance of ∼95%.

The predictive results of the Tr-SVM classifiers with the highest accuracy at both the tree-decision levels are summarized in Table 2. To validate the predictive models, a 10-fold cross-validation was conducted and confusion matrices were generated based on the validation results (Figure 4). The fine Gaussian SVM classifier successfully discriminated the gene sequences with an overall accuracy of 88.7% at the first tree-decision level. From 4800 SERS spectra, 4257 were correctly predicted by the model. At the second tree-decision level, for both *mecA* and *intI1*, the quadratic SVM classifiers showed the highest prediction accuracy of 90.8 and 94.4% with 1816 and 1888 spectra correctly identified from 2000. The predictive models showed high sensitivity and specificity at both tree-decision levels. Sensitivity and specificity relate to the classifier's ability to correctly identify the true positives as positives and the true negatives as negatives. At the first tree-decision level, the maximum sensitivity was 93.0% for the *intI1* group, while the lowest sensitivity was 68.5% for NS2. All specificities were >90% for the *intI1* and *mecA* groups, NS1, and NS2. At the second tree-decision level, the maximum sensitivities were 93.3 and 96.8% for *intI1*-M2 and *mecA*-M10. The lowest sensitivities were 88.0 and 91.3% for *mecA* and *intI1*-M20. Both second tree-decision levels had a high specificity of >95% for all classes. We also trained the predictive model without a tree-based classification system, in other words, the classification of 12 gene sequences with one model (Figure S4). The quadratic SVM classifier showed the highest overall accuracy of 80.1% among other kernel functions. This simplified model had lower accuracy, thus illustrating that the multidecision tree-based system improved the overall classification performance of SVM. Using one predictive model for a large number of classes limits the flexibility of the hyperplane. Instead, individual predictive models for a limited number of classes can allow the hyperplanes to be better optimized for data clusters such as *mecA* and *intI1* gene segments and their analogues. Owing to this benefit, at the second decision level, the models showed high sensitivity and specificity of >90% despite their similarity in base composition.


https://doi.org/10.1021/acs.analchem.0c04576
Anal. Chem. 2021, 93, 9319−9328

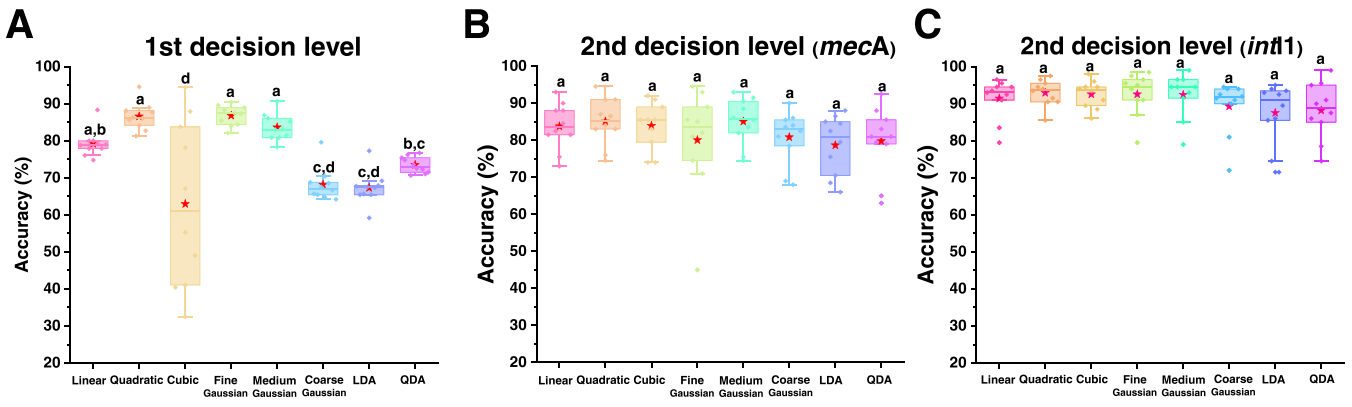



**Figure 5.** Comparison of the accuracy of the different trained multiclass SVM classifiers (linear, quadratic, cubic kernel functions and fine, medium, coarse Gaussian kernel functions) and LDA and QDA with PCA-enabled feature selection. Each boxplot has 10 points from the 10-fold cross-validation results, upper and lower quartiles, the outlier whisker, median, and mean (red-colored star) at first and second decision levels.

We investigated the base discrimination limit of our method. For the *int*I1 group, we added one base mismatch (*int*I1-M1) and developed an SVM classifier (Figure S5). The quadratic SVM classifier successfully discriminated against *int*I1 and its mismatches including the single-base mismatch with an accuracy of 84.3% with 2023 SERS spectra correctly classified from 2400 spectra. This result illustrates the discriminatory capacity of the SVM classifier and its base sensitivity. Further, the effect of the base sequence was tested by comparing *int*I1 with *int*I1 with two base position alternations (Figure S6). The two SERS spectra were readily discriminated by an SVM classifier with an accuracy of 100%. This result indicates that there was a detectable spectral difference despite the sequences having the same base composition and may reflect sequence-driven differences in ssDNA secondary structure that alter which ssDNA signals are enhanced by the SERS hot-spots. Accordingly, the measured SERS signals reflect the combined effects of sequence and sequence-dictated 3D conformation.

**Comparison of SVM and Discriminant Analysis.** We compared models to find the most powerful algorithm for discrimination of SERS spectra of different gene sequences: Tr-SVM classifiers with different kernel functions and two discriminant analyses. The set of boxplots in Figure 5 shows the accuracy of different kernels for the identification of gene sequences. Each boxplot consisted of 10 prediction results from the 10-fold cross-validation. At the first tree-decision level, the quadratic, fine, and medium Gaussian SVM classifiers showed comparable accuracy of ~85% based on the post-hoc Duncan's method ($\alpha$ = 0.05). A Gaussian kernel function is a popular function to provide a nonlinear SVM classifier and is an attractive tool for multivariate analysis owing to its computational simplicity.[67] The fine and medium Gaussian kernel functions are versatile means to solve nonlinear optimization problems. This result implied that the dataset at the first tree-decision level showed nonlinearity and the quadratic SVM classifier can perform comparably to the fine and medium Gaussian SVM classifiers. The accuracies for the models followed the order linear SVM > QDA > coarse Gaussian SVM ≈ LDA > cubic SVM. The linear SVM classifier had higher accuracy than LDA and QDA, thus demonstrating the superior power of SVM for discrimination of gene sequences. In theory, the flexibilities of kernel functions increase in the order of linear < quadratic < cubic (polynomial) and fine < medium < coarse (Gaussian). Given the results at the first tree-decision level, the high degree of flexibility of kernel functions such as cubic and coarse Gaussian may increase the chance of overfitting the model and decreased accuracy.

The SVM classifiers for discrimination of *mec*A and *int*I1 species at the second decision levels showed statistically similar accuracy with all kernel functions (post-hoc Duncan's method, $\alpha$ = 0.05). Even discriminant analysis techniques showed accuracy as high as the SVM classifiers. This result demonstrates that such a dataset was relatively linear separable compared to the one at the first tree-decision level such that the linear classifiers had comparably high predictive accuracy.

The stability of the predictive model was evaluated based on the variability in accuracy. The CV of accuracy for the tree-based cubic SVM was 35.4%, while the CVs for the other models ranged from 3.0 to 6.7% at the first tree-decision level. This result implies that the high degree polynomial factor could overfit the hyperplane and result in poor model stability. The CVs at the second tree-decision levels ranged from 7.5 to 17.9% for *mec*A species and 3.8 to 9.5% for *int*I1 species, indicating the relatively high stability of the models.

To recapitulate, the classifiers with linearity (e.g., linear Gaussian SVM and LDA) or flexibility (e.g., quadratic and fine/medium Gaussian SVM) had a significant influence on the accuracy at the first tree-decision level, while all classifiers showed comparable predictive performance at the second tree-decision level. It is expected that a high extent of nonlinearity within a dataset would require a flexible classifier for precise identification. The optimal classifier is strongly dependent on the dataset characteristics. It needs to be determined if a more flexible and robust classifier (e.g., cubic or coarse Gaussian) is required in circumstances with more highly complicated data structures.

## ■ CONCLUSIONS AND FUTURE OUTLOOK

The SLIPSERS substrate presented herein successfully enabled label-free detection of 100 nM ssDNA of 86 base length by the generation of dense SERS hot-spots. The low spatial variability of SERS hot-spots in the SLIPSERS substrate compared to other commercial SERS substrates illustrates the reproducibility of the approach. Tr-SVM classifiers were successfully built to discriminate 12 different gene sequences including


https://doi.org/10.1021/acs.analchem.0c04576
Anal. Chem. 2021, 93, 9319−9328

*mec*A and *int*I1 gene segments with an accuracy of ~90%. Tr-SVM classifiers discriminated two base mismatches out of 86 bases (2.3% difference). It was also found that one SVM classifier had single-base discriminatory ability. To the best of our knowledge, all prior studies using SERS to interrogate ssDNA have used shorter gene lengths than our 86 genes.[24,35,36] For example, Xu et al. reported SERS detection of 12 nts oligonucleotides with single-base sensitivity. In this case, the sensitivity is 1/12 nts (8.3% difference). Dick and Bell also detected the spectral difference in additional nucleobases from 10 to 24 nts (4.2−9.1% difference) with scaled spectral subtraction. Based on these comparisons, we believe that the base discrimination limit of our approach is low enough to provide single-base sensitivity.

Some challenges remain when discriminating ssDNA using SLIPSERS and the Tr-SVM classifiers. To our knowledge, there is no solid rule for feature selection from SERS spectra to develop discriminatory models. The computer-aided feature selection method used in our study should be collectively compared with other methods. Also, our results reflect that the accuracy of the Tr-SVM classifiers is dependent on the kernel functions and the characteristics of the dataset. A deeper understanding of these spectral datasets will be required for the better development of the models. To expand the applicability of Tr-SVM classifiers for diverse types of ssDNA with respect to gene length and sequence, higher dimensional datasets of ssDNA must be collected and subjected to training. However, recent advances of SERS DNA detection with an aid of computational analysis illustrate the ability to detect such mutations.[68−70] Overall, despite these challenges, this study illustrates the great promise of discriminatory detection of ssDNA using SERS and the functional prediction model (i.e., the Tr-SVM classifiers).

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.0c04576.

SERS spectra of alternative sample production procedures, replicate samples, examination of the effect of oligonucleotide length; principal component analysis; confusion matrix for SVM without decision levels; PCA-SVM confusion matrix; effect of base position alteration; and SVM model accuracy (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

**Peter J. Vikesland** − *Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States; Virginia Tech Institute of Critical Technology and Applied Science (ICTAS) Sustainable Nanotechnology Center (VTSuN), Blacksburg, Virginia 24061, United States;* orcid.org/0000-0003-2654-5132; Email: pvikes@vt.edu

### Authors

**Seju Kang** − *Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States; Virginia Tech Institute of Critical Technology and Applied Science (ICTAS) Sustainable Nanotechnology Center (VTSuN), Blacksburg, Virginia 24061, United States;* orcid.org/0000-0003-3281-4829

**Inyoung Kim** − *Department of Statistics, Virginia Tech, Blacksburg, Virginia 24061, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c04576


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Funding for this work was provided through US NSF grants CBET-1705653 and CBET-2029911 to P.V. Laboratory and instrumentation support was provided by NanoEarth—a node of the NSF-supported NNCI (NSF award number #1542100). Additional support was provided by the Sustainable Nanotechnology Interdisciplinary Graduate Program (VTSuN IGEP) funded by the Virginia Tech Graduate School. Some icons in the graphical abstract were created with BioRender.com.

## ■ REFERENCES


(1) Opalinska, J. B.; Gewirtz, A. M. *Nat. Rev. Drug Discovery* **2002**, *1*, 503−514.
(2) Sidransky, D. *Science* **1997**, *278*, 1054−1058.
(3) Dahm, R. *Hum. Genet.* **2008**, *122*, 565−581.
(4) Kim, T. H.; Park, J.; Kim, C. J.; Cho, Y. K. *Anal. Chem.* **2014**, *86*, 3841−3848.
(5) Kuchta, T.; Knutsson, R.; Fiore, A.; Kudirkiene, E.; Höhl, A.; Horvatek Tomic, D.; Gotcheva, V.; Pöpping, B.; Scaramagli, S.; To Kim, A.; Wagner, M.; De Medici, D. *Lett. Appl. Microbiol.* **2014**, *59*, 263−271.
(6) Griffiths, R. I.; Whiteley, A. S.; O'Donnell, A. G.; Bailey, M. J. *Appl. Environ. Microbiol.* **2000**, *66*, 5488−5491.
(7) Wintzingerode, F. V.; Göbel, U. B.; Stackebrandt, E. *FEMS Microbiol. Rev.* **1997**, *21*, 213−229.
(8) Vogelstein, B.; Kinzler, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9236−9241.
(9) Korbie, D. J.; Mattick, J. S. *Nat. Protoc.* **2008**, *3*, 1452−1456.
(10) Siebert, P. D.; Larrick, J. W. *Nature* **1992**, *359*, 557−558.
(11) Metzker, M. L. *Nat. Rev. Genet.* **2010**, *11*, 31−46.
(12) Buermans, H. P. J.; den Dunnen, J. T. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2014**, *1842*, 1932−1941.
(13) Venkatesan, B. M.; Bashir, R. *Nat. Nanotechnol.* **2011**, *6*, 615−624.
(14) Raman, C. V.; Krishnan, K. S. *Nature* **1928**, *122*, 169.
(15) Jeanmaire, D. L.; Van Duyne, R. P. *J. Electroanal. Chem. Interfacial Electrochem.* **1977**, *84*, 1−20.
(16) Wang, H. H.; Liu, C. Y.; Wu, S. B.; Liu, N. W.; Peng, C. Y.; Chan, T. H.; Hsu, C. F.; Wang, J. K.; Wang, Y. L. *Adv. Mater.* **2006**, *18*, 491−495.
(17) Li, Y.; Gao, T.; Xu, G.; Xiang, X.; Zhao, B.; Han, X. X.; Guo, X. *Anal. Chem.* **2019**, *91*, 7980−7984.
(18) Graham, D.; Thompson, D. G.; Smith, W. E.; Faulds, K. *Nat. Nanotechnol.* **2008**, *3*, 548−551.
(19) Bailey, M. R.; Pentecost, A. M.; Selimovic, A.; Martin, R. S.; Schultz, Z. D. *Anal. Chem.* **2015**, *87*, 4347−4355.
(20) Liang, Z.; Zhou, J.; Petti, L.; Shao, L.; Jiang, T.; Qing, Y.; Xie, S.; Wu, G.; Mormile, P. *Analyst* **2019**, *144*, 1741−1750.
(21) Sun, Y.; Li, T. *Anal. Chem.* **2018**, *90*, 11614−11621.
(22) Braun, G.; Seung, J. L.; Dante, M.; Nguyen, T. Q.; Moskovits, M.; Reich, N. *J. Am. Chem. Soc.* **2007**, *129*, 6378−6379.
(23) Bell, S. E. J.; Sirimuthu, N. M. S. *J. Am. Chem. Soc.* **2006**, *128*, 15580−15581.
(24) Dick, S.; Bell, S. E. J. *Faraday Discuss.* **2017**, *205*, 517−536.
(25) Ngo, H. T.; Wang, H. N.; Fales, A. M.; Vo-Dinh, T. *Anal. Bioanal. Chem.* **2016**, *408*, 1773−1781.

(26) Wang, H. N.; Vo-Dinh, T. *Nanotechnology* **2009**, *20*, No. 065101.
(27) Ngo, H. T.; Gandra, N.; Fales, A. M.; Taylor, S. M.; Vo-Dinh, T. *Biosens. Bioelectron.* **2016**, *81*, 8−14.
(28) Yang, S.; Dai, X.; Stogin, B. B.; Wong, T. S. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 268−273.
(29) Zhang, D.; You, H.; Yuan, L.; Hao, R.; Li, T.; Fang, J. *Anal. Chem.* **2019**, *91*, 4687−4695.
(30) Wong, T.-S.; Kang, S. H.; Tang, S. K. Y.; Smythe, E. J.; Hatton, B. D.; Grinthal, A.; Aizenberg, J. *Nature* **2011**, *477*, 443−447.
(31) Shore, A. C.; Coleman, D. C. *Int. J. Med. Microbiol.* **2013**, *303*, 350−359.
(32) Vandenbroucke-Grauls, C. M. J. E.; Kusters, J. G.; Gala, J. L. *J. Clin. Microbiol.* **1996**, *34*, No. 1599.
(33) Gillings, M. R.; Gaze, W. H.; Pruden, A.; Smalla, K.; Tiedje, J. M.; Zhu, Y. G. *ISME J.* **2015**, *9*, 1269−1279.
(34) Stedtfeld, R. D.; Stedtfeld, T. M.; Waseem, H.; Fitschen-Brown, M.; Guo, X.; Chai, B.; Williams, M. R.; Shook, T.; Logan, A.; Graham, A.; Chae, J. C.; Sul, W. J.; VanHouten, J.; Cole, J. R.; Zylstra, G. J.; Tiedje, J. M.; Upham, B. L.; Hashsham, S. A. *J. Environ. Manage.* **2017**, *198*, 213−220.
(35) Xu, L. J.; Lei, Z. C.; Li, J.; Zong, C.; Yang, C. J.; Ren, B. *J. Am. Chem. Soc.* **2015**, *137*, 5149−5154.
(36) Papadopoulou, E.; Bell, S. E. J. *Angew. Chem., Int. Ed.* **2011**, *50*, 9058−9061.
(37) Ladwani, V. M. Support Vector Machines and Applications. In *Computer Vision: Concepts, Methodologies, Tools, and Applications*; IGI Global, 2018; pp 1381−1390.
(38) Noble, W. S. *Nat. Biotechnol.* **2006**, *24*, 1565−1567.
(39) Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press, 2000.
(40) Kusić, D.; Kampe, B.; Rösch, P.; Popp, J. *Water Res.* **2014**, *48*, 179−189.
(41) Volkmann, H.; Schwartz, T.; Bischoff, P.; Kirchen, S.; Obst, U. *J. Microbiol. Methods* **2004**, *56*, 277−286.
(42) Hardwick, S. A.; Stokes, H. W.; Findlay, S.; Taylor, M.; Gillings, M. R. *FEMS Microbiol. Lett.* **2008**, *278*, 207−212.
(43) Brown, K. R.; Walter, D. G.; Natan, M. J. *Chem. Mater.* **2000**, *12*, 306−313.
(44) Wang, Y.; Wei, Z.; Zhang, Y.; Chen, Y. *Electrophoresis* **2019**, *40*, 2104−2111.
(45) Liu, H.; Yang, Z.; Meng, L.; Sun, Y.; Wang, J.; Yang, L.; Liu, J.; Tian, Z. *J. Am. Chem. Soc.* **2014**, *136*, 5332−5341.
(46) Su, J.; Wang, G. J.; Lin, X. G.; Liu, D. Z. *Acta Autom. Sin.* **2008**, *34*, 1040−1046.
(47) Leng, W.; Pati, P.; Vikesland, P. J. *Environ. Sci. Nano* **2015**, *2*, 440−453.
(48) Piella, J.; Bastús, N. G.; Puntes, V. *Bioconjugate Chem.* **2017**, *28*, 88−97.
(49) Wei, H.; Leng, W.; Song, J.; Liu, C.; Willner, M. R.; Huang, Q.; Zhou, W.; Vikesland, P. J. *Environ. Sci. Technol.* **2019**, *53*, 575−585.
(50) Wei, H.; McCarthy, A.; Song, J.; Zhou, W.; Vikesland, P. J. *Faraday Discuss.* **2017**, *205*, 491−504.
(51) Wei, H.; Leng, W.; Song, J.; Willner, M. R.; Marr, L. C.; Zhou, W.; Vikesland, P. J. *Anal. Chem.* **2018**, *90*, 3227−3237.
(52) Wei, H.; Huang, Q.; Vikesland, P. J. *Environ. Sci. Technol. Lett.* **2019**, *6*, 199−204.
(53) Li, Y.; Gao, T.; Xu, G.; Xiang, X.; Han, X.; Zhao, B.; Guo, X. *J. Phys. Chem. Lett.* **2019**, *10*, 3013−3018.
(54) Notingher, I.; Green, C.; Dyer, C.; Perkins, E.; Hopkins, N.; Lindsay, C.; Hench, L. L. *J. R. Soc., Interface* **2004**, *1*, 79−90.
(55) Ruiz-Chica, A. J.; Medina, M. A.; Sánchez-Jiménez, F.; Ramírez, F. J. *J. Raman Spectrosc.* **2004**, *35*, 93−100.
(56) Chan, J. W.; Taylor, D. S.; Zwerdling, T.; Lane, S. M.; Ihara, K.; Huser, T. *Biophys. J.* **2006**, *90*, 648−656.
(57) Qiu, L.; Liu, P.; Zhao, L.; Wen, M.; Yang, H.; Fan, S.; Zhou, L. *Vib. Spectrosc.* **2014**, *72*, 134−141.
(58) Stone, N.; Kendall, C.; Smith, J.; Crow, P.; Barr, H. *Faraday Discuss.* **2004**, *126*, 141−157.
(59) Cheng, W. T.; Liu, M. T.; Liu, H. N.; Lin, S. Y. *Microsc. Res. Tech.* **2005**, *68*, 75−79.
(60) Krafft, C.; Neudert, L.; Simat, T.; Salzer, R. *Spectrochim. Acta, Part A* **2005**, *61*, 1529−1535.
(61) Rastogi, S. K.; Jabal, J. M. F.; Zhang, H.; Gibson, C. M.; Haler, K. J.; Qiang, Y.; Aston, D. E.; Branen, A. L. *J. Nanomed. Nanotechnol.* **2011**, *2*, 2−9.
(62) Papadopoulou, E.; Bell, S. E. J. *Chem. Commun.* **2011**, *47*, 10966−10968.
(63) Betz, J. F.; Cheng, Y.; Rubloff, G. W. *Analyst* **2012**, *137*, 826−828.
(64) Halvorson, R. A.; Vikesland, P. J. *Environ. Sci. Technol.* **2010**, *44*, 7749−7755.
(65) Hauschild, A. C.; Kopczynski, D.; D'Addario, M.; Baumbach, J. I.; Rahmann, S.; Baumbach, J. *Metabolites* **2013**, *3*, 277−293.
(66) Astuti, W.; Adiwijaya. *J. Phys.: Conf. Ser.* **2018**, *971*, No. 012003.
(67) Seeger, M. *Int. J. Neural Syst.* **2004**, *14*, 69−106.
(68) Wu, L.; Teixeira, A.; Garrido-Maestu, A.; Muinelo-Romay, L.; Lima, L.; Santos, L. L.; Prado, M.; Diéguez, L. *Biosens. Bioelectron.* **2020**, *165*, No. 112392.
(69) Kowalczyk, A.; Krajczewski, J.; Kowalik, A.; Weyher, J. L.; Dzięcielewski, I.; Chłopek, M.; Góźdź, S.; Nowicka, A. M.; Kudelski, A. *Biosens. Bioelectron.* **2019**, *132*, 326−332.
(70) Wu, L.; Garrido-Maestu, A.; Guerreiro, J. R. L.; Carvalho, S.; Abalde-Cela, S.; Prado, M.; Diéguez, L. *Nanoscale* **2019**, *11*, 7781−7789.


https://doi.org/10.1021/acs.analchem.0c04576
*Anal. Chem.* 2021, 93, 9319−9328