Drug-resistant Staphylococcus Aureus Bacteria Detection with the Combination of Surface-enhanced Raman Spectroscopy and Deep Learning Techniques

Fatma Uysal Ciloglu
Erciyes University
Abdullah Caliskan
Munster Technological University
Ayse Mine Saridag
Gaziantep University
Ibrahim Halil Kilic
Gaziantep University
Mahmut Tokmakci
Erciyes University
Mehmet Kahraman
Gaziantep University
Omer Aydin (biomer@umich.edu)
Erciyes University

Research Article

Keywords: Drug-resistant, Staphylococcus Aureus , Bacteria , Raman Spectroscopy, Deep Learning Techniques

Posted Date: May 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-536102/v1

Drug-resistant *Staphylococcus aureus* bacteria detection with the combination of surface-enhanced Raman spectroscopy and deep learning techniques

Fatma Uysal Ciloglu[1], Abdullah Caliskan[2,3], Ayse Mine Saridag[4], Ibrahim Halil Kilic[5], Mahmut Tokmakci[1], Mehmet Kahraman[4*], Omer Aydin[1,6,7*]

[1] Department of Biomedical Engineering, Erciyes University, Kayseri, 38039, Turkey

email: biomer@umich.edu ; Tel: +90-352-207-6666/32984

[2] Department of Biomedical Engineering, Iskenderun Technical University, Hatay, 31200, Turkey

[3] IMaR Technology Gateway, Munster Technological University, Kerry, Ireland

[4] Department of Chemistry, Gaziantep University, Gaziantep, 27310, Turkey

email: mskahraman46@gmail.com ; Tel: +90-342-317-1940

[5] Department of Biology, Gaziantep University, Gaziantep, 27310, Turkey

[6] ERNAM-Nanotechnology Research and Application Center, Erciyes University, Kayseri, 38039, Turkey

[7] ERKAM-Clinical Engineering Research and Application Center, Erciyes University, Kayseri, 38040, Turkey

Abstract

Over the past year, the world's attention has focused on combating COVID-19 disease, but the other threat waiting at the door - antimicrobial resistance should not be forgotten. Although making the diagnosis rapidly and accurately is crucial in preventing antibiotic resistance development, bacterial identification techniques include some challenging processes. To address this challenge, we proposed a deep neural network (DNN) that can discriminate antibiotic-resistant bacteria using surface-enhanced Raman spectroscopy (SERS). Stacked autoencoder (SAE)-based DNN was used for the rapid identification of methicillin-resistant Staphylococcus aureus (MRSA) and methicillin-sensitive S. aureus (MSSA) bacteria using a label-free SERS technique. The performance of the DNN was compared with other traditional classifiers. Since the SERS technique provides high signal-to-noise ratio (SNR) data, some subtle differences were found between MRSA and MSSA in relative band intensities. SAE-based DNN can learn features from raw data and classify them with an accuracy of 97.66%. Moreover, the model discriminates bacteria with an area under curve (AUC) of 0.99. Compared with other traditional classifiers, SAE-based DNN was found superior in accuracy and AUC values. The obtained results are also supported by statistical analysis. These results demonstrate that deep learning has great potential to characterize and detect antibiotic-resistant bacteria by using SERS spectral data. The proposed method is a label-free, easy implemented, and reliable technique with high sensitivity for clinical use.

Introduction

Antimicrobial resistance is a growing problem globally, and 700,000 people die because of resistant infections annually. By 2050, it will threaten 10 million lives a year[1]. Inappropriate prescribing increases unnecessary antibiotic consumption, which triggers antimicrobial resistance with a short period[2]. Antibiotic resistance can be prevented by prescribing the proper antibiotics and raising public awareness. As another solution, new antibiotics can be discovered to compensate for antibiotic resistance. However, the number of discovered and approved antibiotics has declined between 1980 and 2014[3]. Hence, rapid, and correct diagnosis of bacterial infections is required to prescribe the right antibiotic, and this is so crucial to curb antibiotic resistance.

Antimicrobial susceptibility test (AST), categorized as phenotypic and genotypic, is utilized to determine bacteria's antibiotic resistance. Phenotypic AST is reliable; however, it contains a time-consuming culturing step. On the other hand, genotypic AST provides fast results since it eliminates the need for culturing. Although it is highly sensitive, the existence of resistance genes does not mean expressed resistance. Further, genotypic AST requires trained personnel with advance knowledge[4]. Therefore, alternative diagnostic tools are needed for fast and reliable detection of antibiotic resistance.

Surface-enhanced Raman spectroscopy (SERS) is a promising biomedical diagnostic tool and span broad applications in the biomedical field[5-9]. Within the last two decades, it has been successfully applied to discriminate bacteria as well[10-12]. Therefore, the SERS technique also has a significant potential to detect bacteria's antibiotic resistance[13,14]. Although SERS provides unique molecular information, SERS spectra of antibiotic-resistant and susceptible bacteria show subtle spectral differences. Therefore, the SERS technique requires advanced data processing algorithms to capture these minor differences. A vast majority of publications

have reported that machine learning techniques can be employed to discriminate antibiotic-resistant and susceptible bacteria by using data obtained from SERS[15-18].

There are three main steps, including preprocessing, feature extraction, and classification, to determine bacteria from the SERS data by using machine learning techniques. Therefore, obtaining a classification model is very tedious and time-consuming due to the rigid interdependency of the steps. Although some traditional machine learning techniques give reasonable accuracy results to detect the type of bacteria, they have several disadvantages, including overfitting, underfitting, requiring many user-supplied parameters, needing advanced nonlinear optimization techniques, etc. Fortunately, these challenges can be overcome using deep learning models whose achievement originates from large data volumes and sophisticated computational abilities. Deep learning models can learn significant raw data patterns without using advanced preprocessing and feature extraction techniques[19]. Thus, these algorithms seriously reduce the need for feature engineering[20].

In recent years, deep learning algorithms have been applied to analyze spectroscopic signals[19]. However, the number of studies in spectral analysis with deep learning is limited[21-23]. A few studies have been reported to discriminate antibiotic-resistant and susceptible bacteria with deep learning algorithms using Raman spectroscopy and SERS[24,25]. Ho *et al.* have utilized a convolutional neural network (CNN) to classify 30 common bacterial pathogens data obtained from Raman spectroscopy[24]. They have also shown that CNN distinguished Methicillin-resistant *Staphylococcus aureus* (MRSA) and methicillin-susceptible *S. aureus* (MSSA) bacteria with 89±0.1% accuracy using Raman spectral data. Raman spectroscopy has low signal-to-noise ratio (SNR)f due to the low scattering efficiency. This low SNR may be masked easily by background noise. Since noisy Raman spectra make it difficult to detect subtle differences between spectra, the performance of the classifier may be decreased. Thrift *et al.*

reported that variational autoencoders discriminate *Escherichia coli* and *Pseudomonas aeruginosa* bacteria's metabolite profiles based on their SERS data[25].

Among deep learning algorithms, autoencoders have been increasingly used in medical applications in recent years[26,27]. An autoencoder trained in an unsupervised manner is a three-layer feedforward neural network. It consists of an encoder and a decoder[19]. The autoencoder attempts to learn its own input. Thus, it can automatically learn new features from the unlabeled data. The learned features obtained from the autoencoder's hidden layer output are sent to another autoencoder input. This process can be repeated as many times as desired to construct a stacked autoencoder (SAE). A deep neural network (DNN) is formed by combining obtained SAE and a softmax classifier[28].

Herein, we use an SAE-based DNN to classify MRSA and MSSA bacteria's SERS spectra. MRSA has been shown as serious threat according to the 2019 report on antibiotic resistance threats in the United States[29]. These bacteria are resistant to β-lactam antibiotics including penicillin, cephalosporin, and carbapenem[30]. Undoubtedly, rapid and accurate detection of antibiotic resistance profiles of *S. aureus* bacteria will both reduce morbidity and mortality and slow down the development of antibiotic resistance.

We hypothesized that the cell wall structure of MRSA and MSSA might show some differences due to the resistance mechanism. SERS able to reflect these differences at the collected spectra. The discrimination of subtle spectral differences originates from the cell wall structure of MRSA and MSSA is a challenging problem. This study addresses this challenge by using a SAE-based DNN to discriminate between antibiotic-resistant and susceptible bacteria.

Fig. 1 illustrates the general workflow of this study. We first collected SERS spectra of MRSA and MSSA using silver nanoparticles (AgNPs) as SERS substrate. The raw spectral dataset of antibiotic-resistant and susceptible *S. aureus* bacteria was classified using SAE-based

DNN. Further, some traditional machine learning algorithms such as support vector machine (SVM), linear discriminant analysis (LDA), k-nearest neighbors (KNN), decision tree (DT), and a neural network (NN) were used to compare the performances of the DNN and traditional classifiers. To the best of our knowledge, this is the first report for discrimination of MRSA and MSSA SERS data by using SAE-based DNN. This work shows that SERS together SAE-based DNN can successfully discriminate MRSA and MSSA bacteria.

Materials and methods

AgNPs synthesis. AgNPs were synthesized according to the method reported by Lee and Meisel[31]. Briefly, 90 mg silver nitrate ($AgNO_3$, Merck, Darmstadt, Germany) was dissolved in 500 mL distilled water. This solution was heated until boiling. Then, 10 mL aliquot of 1% sodium citrate (Merck, Darmstadt, Germany) was added drop by drop into the solution. The solution was kept boiling for about 1 hour. Synthesized AgNPs were centrifuged at 5500 rpm for 1 hour and discard a portion of the supernatant to form 4x concentrated AgNPs. To characterize AgNPs, the absorption spectrum was measured in the range of 300-700 nm using Thermo Scientific Evolution 201 UV-Vis spectrophotometer (Waltham, USA). The morphology of the nanoparticles was determined using Scanning Transmission Electron Microscope (STEM) imaging (Zeiss GeminiSEM) that is performed with an acceleration voltage of 25 kV.

Bacterial sample preparation. *Staphylococcus aureus* strains (MRSA and MSSA) were used in this study. 19 MRSA and 1 MSSA bacteria were obtained from our microorganism collection (Gaziantep University, Biology Department) with the ethical permission of Gaziantep University clinical research ethics committee (09.06.2014/195). In addition, *S. aureus* type strains ATCC 6538 and ATCC 25923 were purchased from The American Type Culture Collection. The antibiotic resistance of MRSA and MSSA was confirmed the presence or absence of *mec A* gene using the Polymerase Chain Reaction (PCR) technique. Furthermore,

the disc diffusion method was also used for the antibiotic resistance confirmation and details can be found in our previous study[15]. According to PCR and disc diffusion method results methicillin resistance was found in all MRSA bacteria.

The bacteria were grown at 37 ºC on Mueller Hinton agar (Merck, Darmstadt, Germany). The samples were gathered with sterile inoculating loops after the 24 hours cultivation. The collected bacteria were added into 1mL ultrapure water, quickly vortexed, and centrifuged for 5 min at 7500 rpm. The supernatant was discarded, and this procedure was repeated 3 times. A 5 µL aliquot of each washed bacterial sample was added into 100 µL 4x concentrated AgNPs colloidal suspension. Then, samples were shortly vortexed to form homogenous mixtures. A 5 µL of each mixture was immediately dropped on CaF_2 slide and dried at room temperature for SERS measurements.

SERS measurements. SERS measurements were performed using Renishaw inVia Reflex Raman Microscopy System (Renishaw plc., Wotton -under-Edge, UK) using 50x (0.75 NA) microscope objective with 1s exposure from 785 nm excitation under the ~3 mW laser power. The laser spot size was calculated as 1.3 µm (1.22 x λ / NA). The spectra were collected with a 5 µm step size to prevent overlapping. Two datasets were acquired on different days and a total of 1500-1550 spectra were collected from each isolate. Hence, the total dataset consisted of 33,975 spectra. 1200 lines/mm^{-1} grating was used providing a spectral range from 550 to 1700 cm^{-1}.

Outlier detection. To detect outliers, the isolation forest algorithm proposed by Liu et al. was utilized[32]. This algorithm was applied in R programming language v.3.6.343[33] using isolationForest() function in the solitude package[34]. In this technique, a score value was found for every spectrum. The score values close to 1 indicate that those points are definitely anomalies, while less than 0.5 indicates that they are not anomalies. Further, if all the score values are around 0.5 indicates that the whole data do not have any anomaly[32]. In this study,

spectra with score values greater than 0.7 were accepted anomaly and discarded from the dataset.

SAE-based DNN. SAE-based DNN presented here consists of the encoder layers of autoencoders and a softmax classifier.

A single autoencoder consists of encoder and decoder parts. The encoder part takes the input vector x ($x \in R^{Mx1}$) and maps this vector into hidden representation c known as code[35]. This process is as follows:

$$c_i = f(b_{i1} + W_{i1}x) \quad (1)$$

where $c_i \in R^{Mx1}$ is the code, f is the encoding function, $b_{i1} \in R^{Mx1}$ is the bias vector, and $W_{i1} \in R^{MxN}$ is the weight matrix of the encoder. Encoder part of an autoencoder is trained using unsupervised fashion to dig significant feature information.

The decoder part reconstructs the input vector as $\hat{x}$. Thus, an autoencoder tries to generate its input at the output layer by minimizing the error as much as possible between input x and output $\hat{x}$. Decoding of c_i is expressed as follows:

$$\hat{x} = g(b_{i2} + W_{i2}c_i) \quad (2)$$

where g is the encoding function, $b_{i2} \in R^{Nx1}$ is the bias vector, and $W_{i2} \in R^{NxM}$ is the weight matrix of the decoder.

The objective function minimizing the error between the input and output is expressed:

$$J(W_i, b_i, x_i) = \frac{1}{2}\|h_{Wi,bi}(x_i) - x_i\|^2 \quad (3)$$

Two regularization term are added to equation (3) as seen in equation (4). λ is a regularization term and used to prevent overfitting. β is the weight of the sparsity penalty term

and used to allow the autoencoder discovering hidden features related to raw data. The term ρ is the constant sparsity parameter and $\hat{\rho_j}$ is the mean activation value of the j^{th} neuron in the hidden layer.

$$J = \frac{1}{2}\|h_{Wi,bi}(x_i) - x_i\|^2 + \lambda(\|W_i\|_2^2) + \beta\sum_{j=1}^{M} KL(\rho\|\hat{\rho}_j) \quad (4)$$

$KL(\rho\|\hat{\rho_j})$ term expresses the Kullback-Leibler divergence:

$$KL(\rho\|\hat{\rho}_j) = \rho\log\frac{\rho}{\hat{\rho}} + (1-\hat{\rho}_j)\log\frac{1-\rho}{1-\hat{\rho}_j} \quad (5)$$

The number of neurons at the hidden layer is generally chosen lower than the input layer size. Therefore, the autoencoder is forced to extract new features with an unsupervised approach.

Softmax classifier, which generalizes logistic regression, is the supervised layer of the deep learning model[35]. It is based on the softmax function and used to classify the learned features by the autoencoder. The softmax classifier's cost function attempts to decrease the difference between the actual label value and model output by tuning the model parameters.

To form a SAE-based DNN, desired number of the encoder part of the trained autoencoders and softmax classifier are joined together. The performance of the SAE-based DNN can be increased by performing backpropagation on the model. This procedure is known as fine tuning and significantly improved the results of the SAE. Fig. 2 shows frameworks of the SAE-based DNN used in this study. The training procedure of the DNN is as follows:

i. The first autoencoder is trained by feeding a 1024-dimensional input vector to a hidden layer with 30 neurons as illustrated in Fig. 2a.

ii. The second autoencoder is trained to feed the first autoencoder's hidden layer to the second autoencoder's hidden layer with 15 neurons as depicted in Fig. 2b.

iii. The encoder parts of the trained autoencoders and softmax layer are consecutively connected to construct a SAE based deep learning model as seen in Fig. 2c.

iv. Finally, to achieve fine-tuning the backpropagation algorithm is performed and weights are updated with labeled training data.

The hyperparameter selection was done by random search. The specific parameters used to construct SAE-based DNN are given in Supplemental Table S1. 10-fold cross validation was performed to prevent overfitting. SAE based deep learning model was constructed with MATLAB software (The MathWorks, Natick, USA).

To compare the performance of SAE-based DNN with the state-of-the-art classifiers SVM, LDA, KNN, DT, and NN were used. These classifiers were performed using MATLAB software. The whole data were standardized before applying to SAE-based DNN and traditional classifiers.

Statistical analysis. The Mann Whitney U test was utilized for statistical analysis. This analysis was performed using MATLAB software and >95% confidence level was selected ($P < 0.05$ means there is a significant difference between the groups).

Results and discussion

SERS spectra of MRSA and MSSA. The citrate reduced AgNPs were used due to their high SERS activity and providing reproducible spectra. The UV-Vis spectrum and STEM image of the synthesized AgNPs are showed in Supplemental Fig. S1. The maximum absorption of AgNPs was found at 420 nm and they were mostly spherical in the range of 50–60 nm as seen in the inserted image in Supplemental Fig. S1.

To collect a large dataset 33,975 spectra were acquired from 19 MRSA and 3 MSSA bacterial isolates for 2 measurement times. The isolation forest algorithm was used to determine whether there is an outlier in the data. The results of this algorithm are shown in Fig. 3a. The score values that show whether the spectrum is outlier or not are in the range of [0.572 0.875]. The mean ± standard deviation of them was found to be 0.574 ± 0.009. Further, the 95th percentile of the score values was found as 0.577 in other words 95% of the score values are lower than 0.577. A vast amount of score values are distributed around 0.5 as seen in Fig. 3a indicates that these spectra do not really anomaly. Only 24 score values were determined greater than 0.7 and the spectra to which these values belong were accepted as anomalies and removed from the dataset.

SERS spectra are acquired by illuminating the whole bacterium which interacts with the colloidal AgNPs. Thus, the collected spectra are generated by the outermost of the bacterial cell wall because of the distance dependence of SERS enhancement[36,37]. Since the SERS spectra collected from bacteria reflect composition of the cell wall in close proximity with the SERS substrate. Peptidoglycan layer, teichoic acids, surface proteins, capsular polysaccharides, and phospholipids are the primary components of the bacterial cell wall[38]. The peptidoglycan layer in the bacterial cell wall is a protective envelope found on the outside of the cytoplasmic membrane where composes of glycan strands crosslinked with short peptides[39]. *Staphylococcus aureus* which is a gram-positive bacterium has a thick peptidoglycan layer at the outermost of the cell wall. Peptidoglycan biosynthesis is an excellent target for most of the antibiotics including β-lactams[38]. Correspondingly, some structural differences are anticipated between MRSA which is resistant to β-lactams and MSSA cell wall that SERS could reveal these differences.

The normalized mean SERS spectra ± standard deviation of MRSA and MSSA are depicted in Fig. 3b in the range of 550 - 1700 cm^{-1}. The shade area shows the variations of

measured spectral intensities. SERS spectra of MRSA and MSSA bacteria depict a lot of similar peak positions except for some differences in relative band intensities. The primary SERS spectra of MRSA and MSSA are characterized by strong bands at 658 cm^{-1} (COO^- deformation of guanine)[40], 732 cm^{-1} (flavin adenine dinucleotide derivatives, glycosidic ring mode of the N-acetyl D-glucosamine and N-acetylmuramic)[41-43], 958 cm^{-1} (CN deformation of saturated lipids)[44], 1333 cm^{-1} (C-N stretching mode of Adenine)[45], 1450 cm^{-1} (CH_2 deformation of saturated lipids)[46], and 1576 cm^{-1} (CN stretching of amide II)[47]. It is clearly illustrated in Fig. 3b that the spectral profile of MRSA and MSSA bacteria is quite similar. However, notable difference between MRSA and MSSA is the intensity of 732 cm^{-1} peak position. Interestingly, this sharp peak is drastically increased in MRSA. The source of this band is explicitly assigned by some groups to flavin adenine dinucleotide (FAD) derivatives and glycosidic ring mode of the N-acetyl D-glucosamine (NAG) and N-acetylmuramic (NAM) which are building blocks of the peptidoglycan layer [41-43]. Kahraman et al. reported that both band assignments are correct, and it is possible for bands originating from both NAG and FAD to overlap[48].

The prominent increase in the 732 cm^{-1} band in MRSA may indicate differentiation in the peptidoglycan layer of MRSA. Since β-lactam antibiotics work by targeting Penicillin Binding Proteins (PBPs) in the peptidoglycan layer. It is possible to observe some differences for peak intensities or positions originating peptidoglycan layer. Genotypic changes that cause antibiotic resistance are usually represented in the induced phenotype that inhibits the action of an antibiotic. Garcia et al. measured the cell wall and septum thickness of MRSA and MSSA[49]. They reported that the cell wall and septum thickness of MRSA and MSSA have found statistically different. Besides, they correlated the cell wall thickness of MRSA with the resistance mechanism. There are also some minor differences the intensity of 658, 958, and 1333 cm^{-1} peaks. These peaks are more intense for MRSA than MSSA. The changes between MRSA and MSSA SERS spectra indicate that there is a variation in the amount of some

biomolecules in the cell wall. Thus, SERS has the potential to reveal the variations between MRSA and MSSA.

SAE-based DNN and traditional classifiers for the classification of MRSA and MSSA SERS spectra. The spectral features of MRSA and MSSA are highly similar as clearly seen in Figure 4b. There are only some subtle differences in the relative intensities of the SERS peaks of MRSA and MSSA. This situation necessitates using of a robust algorithm for data analysis. To correctly classify SERS spectral data of MRSA and MSSA bacteria, an SAE-based DNN was utilized.

The total dataset consists of 33,951 SERS spectra of MRSA and MSSA. The 29,452 of them belong to 19 MRSA isolates and the remaining belongs to 3 MSSA isolates. The entire data were used without preprocessing and feature extraction steps. In spectral data analysis, preprocessing and feature extraction are two important steps that show a major impact on the classifier performance. However, misuse of these methods can seriously distort the original data and adversely affect classifier performance[50].

In this study, preprocessing steps such as noise elimination were not required since SERS can provide high signal-to-noise ratio (SNR) data. Moreover, feature extraction which is a challenging process was not used due to the ability of SAE-based DNN about revealing critical features from the raw data. This deep learning model can extract relevant features thanks to the multiple autoencoders. Thus, the dimension of the input data passing through the hidden layers of each autoencoder is significantly reduced. The raw input data were just standardized before applying SAE-based DNN and traditional classifiers. The whole raw data were shuffled randomly before implementing into classifiers. 10-fold cross validation technique was used to measure the performance of the model and this procedure was repeated for 30 runs for each classifier. The mean accuracies of SAE-based DNN and traditional classifiers for 30 runs are depicted in Fig. 4a. It is clearly seen that SAE based deep learning model shows better

classification performance than traditional classifiers. This model provides the best mean accuracy with 97.66 ± 0.26%, among others. Traditional classifiers have close classification performance and SVM gives slightly better result with 95.87 ± 0.01%, among them. The mean, maximum, minimum, and standard deviation of each classifier accuracies acquired from 30 runs are given in Table 1.

A classifier performance can be measured different metrics and accuracy is one of them. However, accuracy is not enough to measure the performance of a classifier. Especially in data sets where the amount of data in classes is unbalanced, measuring the classifier performance with the only accuracy parameter does not give reliable results. The receiver operating characteristic (ROC) curve shows a classifier performance for all classification threshold values. This curve is plotted with the true positive rate (y axis) against the false positive rate (x axis). The area under a ROC curve abbreviated as AUC is frequently used to measure classification performance and one of the most important classifier performance evaluation techniques. The value of AUC is in the range of [0 -1] and when the AUC value is getting closer to 1 classification error decreases. Fig. 4b illustrates the AUC values of SAE-based DNN and traditional classifiers for 30 runs. As seen in Fig. 4b, SAE based deep neural network has the best AUC values through 30 runs. The mean AUC value of it was found 0.993 ± 0.002 which means the deep learning model can distinguish MRSA and MSSA with a high performance as depicted in Fig. 5a. KNN gives the worst AUC values for each run while LDA, SVM, NN, and DT have the better results between traditional classifiers. In addition, the mean, maximum, minimum, and standard deviation of each classifier AUC values obtained from 30 runs are provided in Table 2. The confusion matrix of SAE-based DNN is demonstrated in Fig. 5b. The accuracy, sensitivity, specificity, and precision of the deep learning model were calculated as 97.7%, 99.2%, 87.6%, and 98.2%, respectively. Misdiagnosing MRSA as MSSA causes more

serious results than the reverse situation. Only 236 SERS spectra of MRSA were misdiagnosed as MSSA as seen in Fig. 5b.

The above results show that SAE-based DNN has better classification performance than traditional classifiers. However, these findings should be supported with statistical analysis. Statistical analysis was used to compare the AUC values obtained from SAE-based DNN against state-of-the-art classification techniques such as SVM, NN, DT, LDA, and KNN for 30 runs. The Mann Whitney U test with significance level of 0.05 was used for this purpose and results are presented in Table 3. The statistical results were interpreted according to the *p-values*. The lower *p-values* mean that there is a large difference between the classifiers. It is clearly seen that SAE-based DNN is found better than traditional classifiers in terms of the statistical analysis results for the discrimination of MRSA and MSSA spectral data.

The SAE-based DNN more accurately classified SERS spectral data of MRSA and MSSA bacteria. Our model, applied here for rapid and reliable identification of antibiotic-resistant and susceptible bacteria requires minimum sample preparation procedure, does not require special labels, and eliminates long incubation times required for phenotypic AST. Although raw data were used in our study, high classification accuracy and AUC were found thanks to the SERS technique with high SNR using SAE-based DNN which successfully extracts features from the data. Our group previously has applied traditional classifiers such as KNN, SVM, DT and naïve Bayes (NB) for the discrimination of MRSA, MSSA, and *Legionella pneumophila* bacteria[15]. KNN classifier has provided best accuracy with 97.8% among other techniques. However, as the size of the data set grows, the success of traditional classifiers falls behind the deep learning algorithms. Therefore, SAE-based DNN can provide more successful results for SERS spectral data of antibiotic-resistant bacteria with high accuracy and sensitivity.

In this study, SERS spectral data of MRSA and MSSA bacteria have been successfully characterized and identified by SAE-based DNN. The results show that the proposed technique

has a potential application for the detection of antibiotic-resistant bacteria in clinical utilization. Compared with the phenotypic or genotypic AST techniques frequently used, proposed method has advantageous in terms of easy use and fast detection.

Conclusions

Rapid bacterial diagnosis is essential to combat antibiotic resistance. Label-free SERS provides fingerprint spectrum of the sample with high SNR. Therefore, it is an attractive technique for bacterial identification studies. However, interpreting of SERS spectra is a difficult process due to the high molecular similarities of bacterial species. Detection of spectral differences between antibiotic-resistant and susceptible bacteria becomes even more difficult. Advanced data analysis techniques are indispensable at this step. Deep learning algorithms perform outstanding success by using SERS data for discrimination of antibiotic-resistant bacteria.

Here we illustrate that SAE-based DNN can be used for the SERS-based label-free identification of antibiotic-resistant and susceptible strains of *S. aureus* bacteria. SERS technique providing high SNR reveals the subtle spectral differences between MRSA and MSSA. SAE-based DNN automatically extracts features needed for classification from the raw spectral data. Therefore, complex preprocessing and feature extraction steps are eliminated. Compared with the traditional classifiers, SAE-based DNN shows a more accurate diagnostic model with an accuracy and AUC of 97.66%, 0.99, respectively. The proposed method provides label free, rapid, and reliable technique with a high sensitivity.

In conclusion, the proposed method has a great potential for clinical use considering that rapid diagnostic methods have a great effect on combating antibiotic resistance. Further, this technique has a high application potential not only in the detection of antibiotic-resistant bacteria, but also for a lot of label-free SERS applications in the biomedical field.

References

1 Jim, O. N. *Tackling drug-resistant infections globally: final report and recommendations*, https://www.biomerieuxconnection.com/wp-content/uploads/2018/04/Tackling-Drug-Resistant-Infections-Globally_-Final-Report-and-Recommendations.pdf Accessed 3 Jan 2021

2 Fleming-Dutra, K. E. *et al.* Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010-2011. *JAMA* **315**, 1864-1873 (2016).

3 Ventola, C. L. The antibiotic resistance crisis: part 1: causes and threats. *P T* **40**, 277-283 (2015).

4 Baltekin, Ö., Boucharin, A., Tano, E., Andersson, D. I. & Elf, J. Antibiotic susceptibility testing in less than 30 min using direct single-cell imaging. *Proc Natl Acad Sci USA* **114**, 9170-9175 (2017).

5 Aydin, Ö., Altaş, M., Kahraman, M., Bayrak, Ö. F. & Çulha, M. Differentiation of Healthy Brain Tissue and Tumors Using Surface-Enhanced Raman Scattering. *Appl Spectrosc* **63**, 1095-1100 (2009).

6 Aydin, Ö., Kahraman, M., Kiliç, E. & Çulha, M. Surface-Enhanced Raman Scattering of Rat Tissues. *Appl Spectrosc* **63**, 662-668 (2009).

7 Kahraman, M. & Wachsmann-Hogiu, S. Label-free and direct protein detection on 3D plasmonic nanovoid structures using surface-enhanced Raman scattering. *Analytica Chimica Acta* **856**, 74-81 (2015).

8 Park, J. *et al.* Exosome Classification by Pattern Analysis of Surface-Enhanced Raman Spectroscopy Data for Lung Cancer Diagnosis. *Anal. Chem.* **89**, 6695-6701 (2017).

9 Li, S. *et al.* Characterization and noninvasive diagnosis of bladder cancer with serum surface enhanced Raman spectroscopy and genetic algorithms. *Scientific Reports* **5**, 9582; 10.1038/srep09582 (2015).

10 Kahraman, M., Keseroğlu, K. & Çulha, M. On Sample Preparation for Surface-Enhanced Raman Scattering (SERS) of Bacteria and the Source of Spectral Features of the Spectra. *Appl Spectrosc* **65**, 500-506 (2011).

11 Kahraman, M., Zamaleeva, A. I., Fakhrullin, R. F. & Culha, M. Layer-by-layer coating of bacteria with noble metal nanoparticles for surface-enhanced Raman scattering. *Anal Bioanal Chem* **395**, 2559-2567 (2009).

12 Mosier-Boss, P. A. Review on SERS of Bacteria. *Biosensors* **7**, 51; 10.3390/bios7040051 (2017).

13 Liu, C.-Y. *et al.* Rapid bacterial antibiotic susceptibility test based on simple surface-enhanced Raman spectroscopic biomarkers. *Scientific Reports* **6**, 23375; 10.1038/srep23375 (2016).

14 Wang, K., Li, S., Petersen, M., Wang, S. & Lu, X. Detection and Characterization of Antibiotic-Resistant Bacteria Using Surface-Enhanced Raman Spectroscopy. *Nanomaterials (Basel)* **8**, 762; 10.3390/nano8100762 (2018).

15 Ciloglu, F. U. *et al.* Identification of methicillin-resistant Staphylococcus aureus bacteria using surface-enhanced Raman spectroscopy and machine learning techniques. *Analyst* **145**, 7559-7570 (2020).

16 Chen, X. *et al.* Surface-enhanced Raman scattering method for the identification of methicillin-resistant Staphylococcus aureus using positively charged silver nanoparticles. *Microchim Acta* **186**, 102; 10.1007/s00604-018-3150-6 (2019).

17 Cheong, Y., Kim, Y. J., Kang, H., Choi, S. & Lee, H. J. Rapid label-free identification of Klebsiella pneumoniae antibiotic resistant strains by the drop-coating deposition surface-enhanced Raman scattering method. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **183**, 53-59 (2017).

18 Li, J. *et al.* Rapid identification and antibiotic susceptibility test of pathogens in blood based on magnetic separation and surface-enhanced Raman scattering. *Microchim Acta* **186**, 475; 10.1007/s00604-019-3571-x (2019).

19 Yang, J. *et al.* Deep learning for vibrational spectral analysis: Recent progress and a practical guide. *Analytica Chimica Acta* **1081**, 6-17 (2019).

20 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444; 10.1038/nature14539 (2015).

21 Berisha, S. *et al.* Deep learning for FTIR histology: leveraging spatial and spectral features with convolutional neural networks. *Analyst* **144**, 1642-1653 (2019).

22 Liu, J. *et al.* Deep convolutional neural networks for Raman spectrum recognition: a unified solution. *Analyst* **142**, 4067-4074 (2017).

23 Shin, H. *et al.* Early-Stage Lung Cancer Diagnosis by Deep Learning-Based Spectroscopic Analysis of Circulating Exosomes. *ACS Nano* **14**, 5435-5444 (2020).

24 Ho, C.-S. *et al.* Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nature Communications* **10**, 4927; 10.1038/s41467-019-12898-9 (2019).

25 Thrift, W. J. *et al.* Deep Learning Analysis of Vibrational Spectra of Bacterial Lysate for Rapid Antimicrobial Susceptibility Testing. *ACS Nano* **14**, 15336-15348 (2020).

26 Caliskan, A., Badem, H., Basturk, A. & Yuksel, M. E. Diagnosis Of The Parkinson Disease By Using Deep Neural Network Classifier. *IU-Journal of Electrical & Electronics Engineering* **17**, 3311-3318 (2017).

27 Caliskan, A., Yuksel, M. E., Badem, H. & Basturk, A. A Deep Neural Network Classifier for Decoding Human Brain Activity Based on Magnetoencephalography. *ELEKTRON ELEKTROTECH* **23**, 63-67 (2017).

28 Liu, G., Bao, H. & Han, B. A Stacked Autoencoder-Based Deep Neural Network for Achieving Gearbox Fault Diagnosis. *Mathematical Problems in Engineering* **2018**, 5105709; 10.1155/2018/5105709 (2018).

29 CDC. Antibiotic-resistant Germs: New Threats. *Centers for Disease Control and Prevention* https://www.cdc.gov/drugresistance/biggest-threats.html Accessed 29 Dec 2020.

30 Fishovitz, J., Hermoso, J. A., Chang, M. & Mobashery, S. Penicillin-Binding Protein 2a of Methicillin-Resistant Staphylococcus aureus. *IUBMB Life* **66**, 572-577 (2014).

31 Lee, P. C. & Meisel, D. Adsorption and surface-enhanced Raman of dyes on silver and gold sols. *The Journal of Physical Chemistry* **86**, 3391-3395 (1982).

32 Liu, F. T., Ting, K. M. & Zhou, Z.-H. Isolation-Based Anomaly Detection. *ACM Trans. Knowl. Discov. Data* **6**, 3; 10.1145/2133360.2133363 (2012).

33 R Core Team. *R: a language and environment for statistical computing*. https://www.r-project.org/ Accessed 26 Dec 2020.

34 Srikanth, K. S. *Solitude: An Implementation of Isolation Forest*, https://cran.r-project.org/web/packages/solitude/index.html Accessed 28 Dec 2020.

35 Zhang, Y., Zhang, E. & Chen, W. Deep neural network for halftone image classification based on sparse auto-encoder. *Engineering Applications of Artificial Intelligence* **50**, 245-255 (2016).

36 Kovacs, G. J., Loutfy, R. O., Vincett, P. S., Jennings, C. & Aroca, R. Distance dependence of SERS enhancement factor from Langmuir-Blodgett monolayers on metal island films: evidence for the electromagnetic mechanism. *Langmuir* **2**, 689-694 (1986).

37 Premasiri, W. R. *et al.* The biochemical origins of the surface-enhanced Raman spectra of bacteria: a metabolomics profiling by SERS. *Anal Bioanal Chem* **408**, 4631-4647 (2016).

38 Rajagopal, M. & Walker, S. Envelope Structures of Gram-Positive Bacteria. *Curr Top Microbiol Immunol* **404**, 1-44 (2017).

39 Vollmer, W. in *Molecular Medical Microbiology (Second Edition)* (ed Yi-Wei Tang *et al.*) 105-124 (Academic Press, 2015).

40 Schuster, K. C., Reese, I., Urlaub, E., Gapes, J. R. & Lendl, B. Multidimensional Information on the Chemical Composition of Single Bacterial Cells by Confocal Raman Microspectroscopy. *Anal. Chem.* **72**, 5529-5534 (2000).

41 Jarvis, R. M. & Goodacre, R. Discrimination of Bacteria Using Surface-Enhanced Raman Spectroscopy. *Anal. Chem.* **76**, 40-47 (2004).

42 Walter, A., März, A., Schumacher, W., Rösch, P. & Popp, J. Towards a fast, high specific and reliable discrimination of bacteria on strain level by means of SERS in a microfluidic device. *Lab Chip* **11**, 1013-1021 (2011).

43 Zeiri, L., Bronk, B. V., Shabtai, Y., Eichler, J. & Efrima, S. Surface-Enhanced Raman Spectroscopy as a Tool for Probing Specific Biochemical Components in Bacteria. *Appl Spectrosc* **58**, 33-40 (2004).
44 Zhou, H. *et al.* SERS Detection of Bacteria in Water by in Situ Coating with Ag Nanoparticles. *Anal. Chem.* **86**, 1525-1533 (2014).
45 Kneipp, K. *et al.* Detection and identification of a single DNA base molecule using surface-enhanced Raman scattering (SERS). *Phys. Rev. E* **57**, R6281-R6284 (1998).
46 Maquelin, K. *et al.* Identification of medically relevant microorganisms by vibrational spectroscopy. *Journal of Microbiological Methods* **51**, 255-271 (2002).
47 Jarvis, R. M., Brooker, A. & Goodacre, R. Surface-Enhanced Raman Spectroscopy for Bacterial Discrimination Utilizing a Scanning Electron Microscope with a Raman Spectroscopy Interface. *Anal. Chem.* **76**, 5198-5202 (2004).
48 Kahraman, M., Yazıcı, M. M., Şahin, F. & Çulha, M. Convective Assembly of Bacteria for Surface-Enhanced Raman Scattering. *Langmuir* **24**, 894-901 (2008).
49 García, A. B., Viñuela-Prieto, J. M., López-González, L. & Candel, F. J. Correlation between resistance mechanisms in Staphylococcus aureus and cell wall and septum thickening. *Infect Drug Resist* **10**, 353-356 (2017).
50 Xu, L. *et al.* Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration. *Analytica Chimica Acta* **616**, 138-143 (2008).

Acknowledgement

This work was financially supported by the Scientific and Technological Research Council of Turkey (Project Number: 120F097).

Author Contributions

F.U.C. and A.C. conceptualized and trained the algorithms and analyzed classification results. F.U.C. wrote the main manuscript text and prepared all figures. F.U.C., A.M.S., and I.H.K. prepared sample cultures. F.U.C. and A.M.S. collected the dataset. M.T., M.K., and O.A. designed the study protocol and supervised the project. All authors contributed to the editing of the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Additional information

Supplementary information is available for this paper.

List of figures

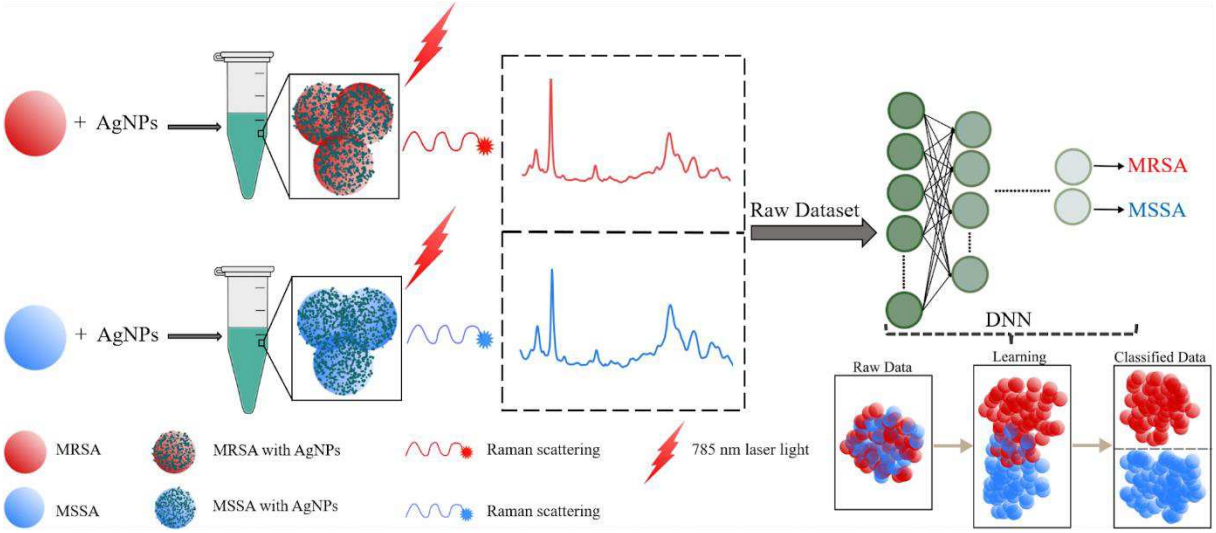


Figure 1. General workflow of deep learning-based spectral data analysis for the discrimination of antibiotic-resistant bacteria.

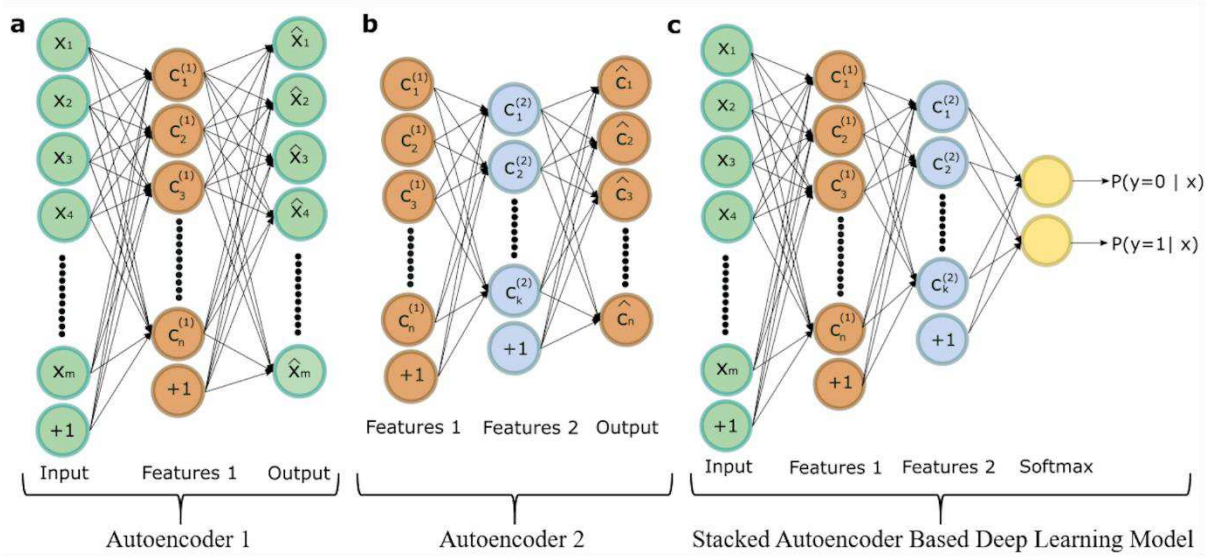
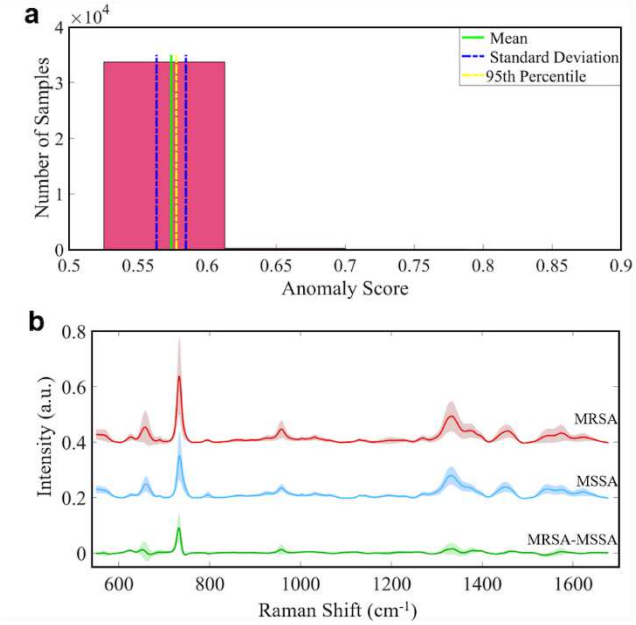


Figure 2. The structure of the proposed deep learning model. (**a**) First autoencoder that reveals the features from the input data. (**b**) Second autoencoder that reveal the new features from the first autoencoder's hidden layer. (**c**) The deep learning model that is formed stacking the two autoencoders and softmax classifier.



561

562 **Figure 3.** (**a**) Histogram plot of the anomaly score values obtained by the isolation forest
563 algorithm. (**b**) Normalized average SERS spectra ± standard deviation of MRSA, MSSA and
564 the difference of MRSA-MSSA.

565

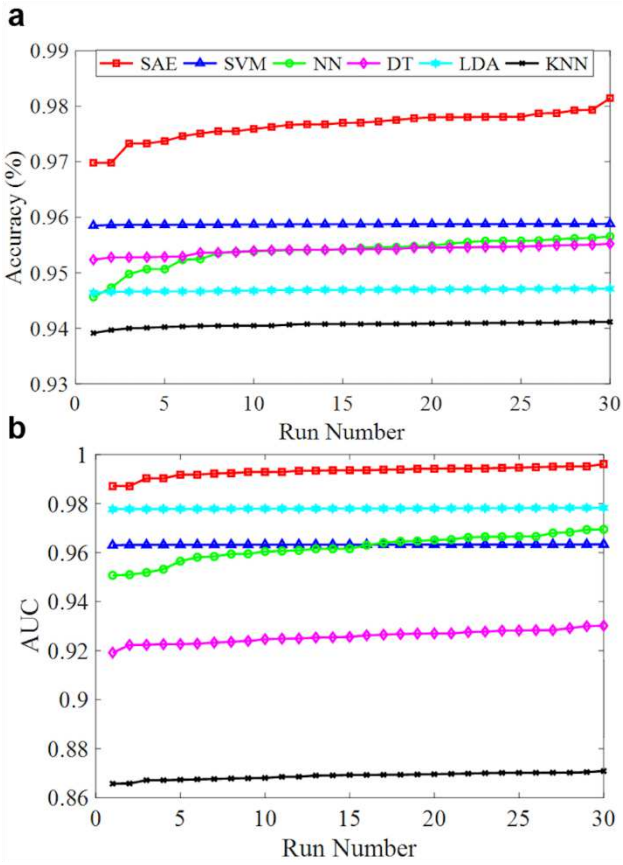


Figure 4. Performance comparisons of SAE-based DNN and traditional classifiers. (**a**) Accuracies of classifiers for 30 runs. (**b**) AUC values obtained from ROC curve of classifiers for 30 runs.

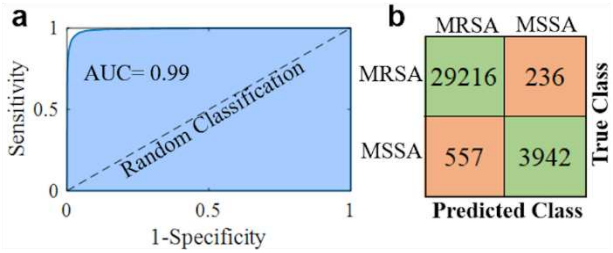


Figure 5. Binary classification results of MRSA and MSSA by the SAE based deep learning model. (**a**) The ROC curve with an AUC of 0.99. (**b**) Confusion matrix showing the results of 10-fold cross validated bacterial identification.

576 **List of tables**

577 **Table 1.** Mean accuracies of the classifiers with maximum, minimum and standard deviation

578 for 30 runs

Method	Mean (%)	Maximum (%)	Minimum (%)	Standard Deviation
SAE	**97.66**	**98.14**	**96.98**	0.26
SVM	95.87	95.88	95.85	0.01
NN	95.38	95.66	94.56	0.26
DT	95.41	95.52	95.23	0.08
LDA	94.69	94.71	94.64	0.02
KNN	94.06	94.12	93.91	0.05

579

580

581 **Table 2.** Mean AUC values of the classifiers with maximum, minimum and standard deviation

582 for 30 runs

Method	Mean	Maximum	Minimum	Standard Deviation
SAE	**0.9931**	**0.9961**	**0.9872**	0.0020
SVM	0.9632	0.9634	0.9630	0.0001
NN	0.962	0.9695	0.9507	0.0054
DT	0.9257	0.9302	0.9192	0.0026
LDA	0.9780	0.9783	0.9777	0.0002
KNN	0.8688	0.8709	0.8657	0.0014

583

584

585

586

587

588 **Table 3.** Statistical analysis results for comparing the SAE-based deep learning model against
589 state-of-the-art machine learning techniques.

Comparison	**Z**	***p-value***	**Sig. (p<0.05)**
SAE vs. SVM	6.6463	0.000	SAE
SAE vs. NN	6.6463	0.000	SAE
SAE vs. DT	6.6463	0.000	SAE
SAE vs. LDA	6.6463	0.000	SAE
SAE vs. KNN	6.6464	0.000	SAE

590

Figures

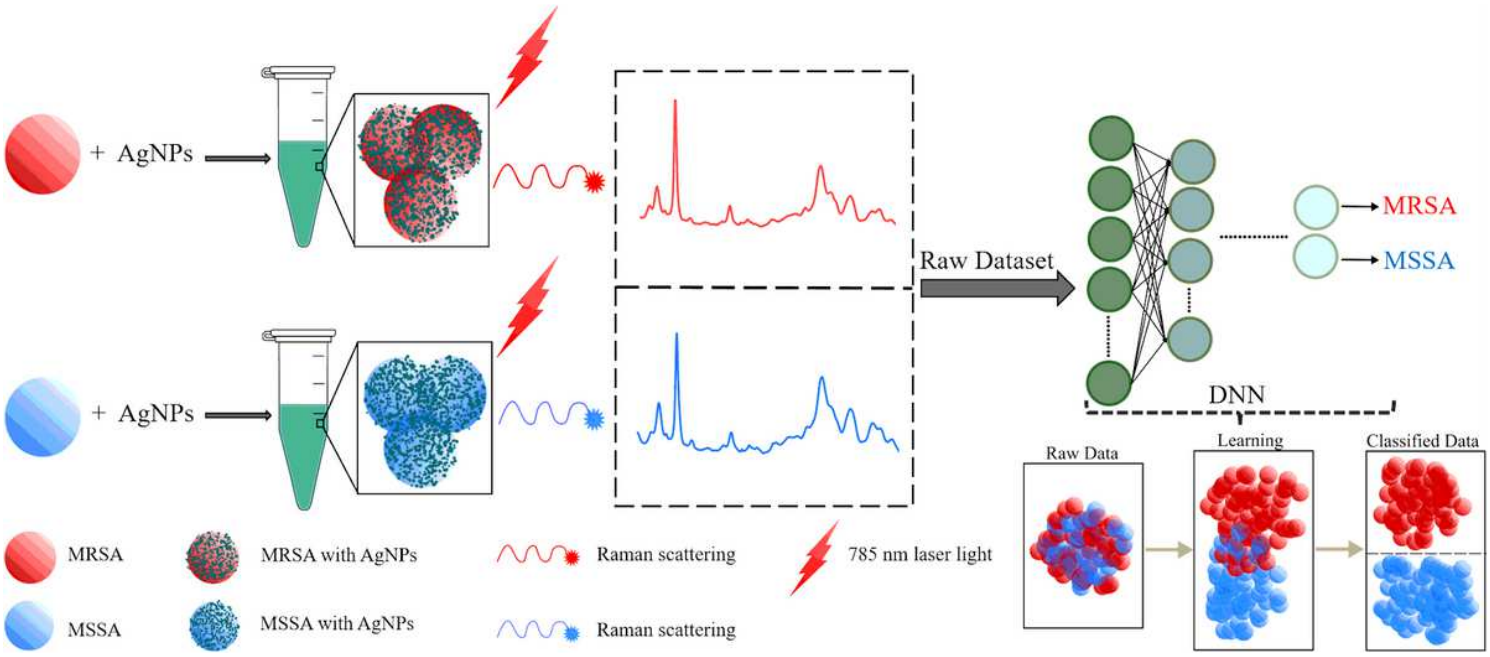


Figure 1

General workflow of deep learning-based spectral data analysis for the discrimination of antibiotic-resistant bacteria.

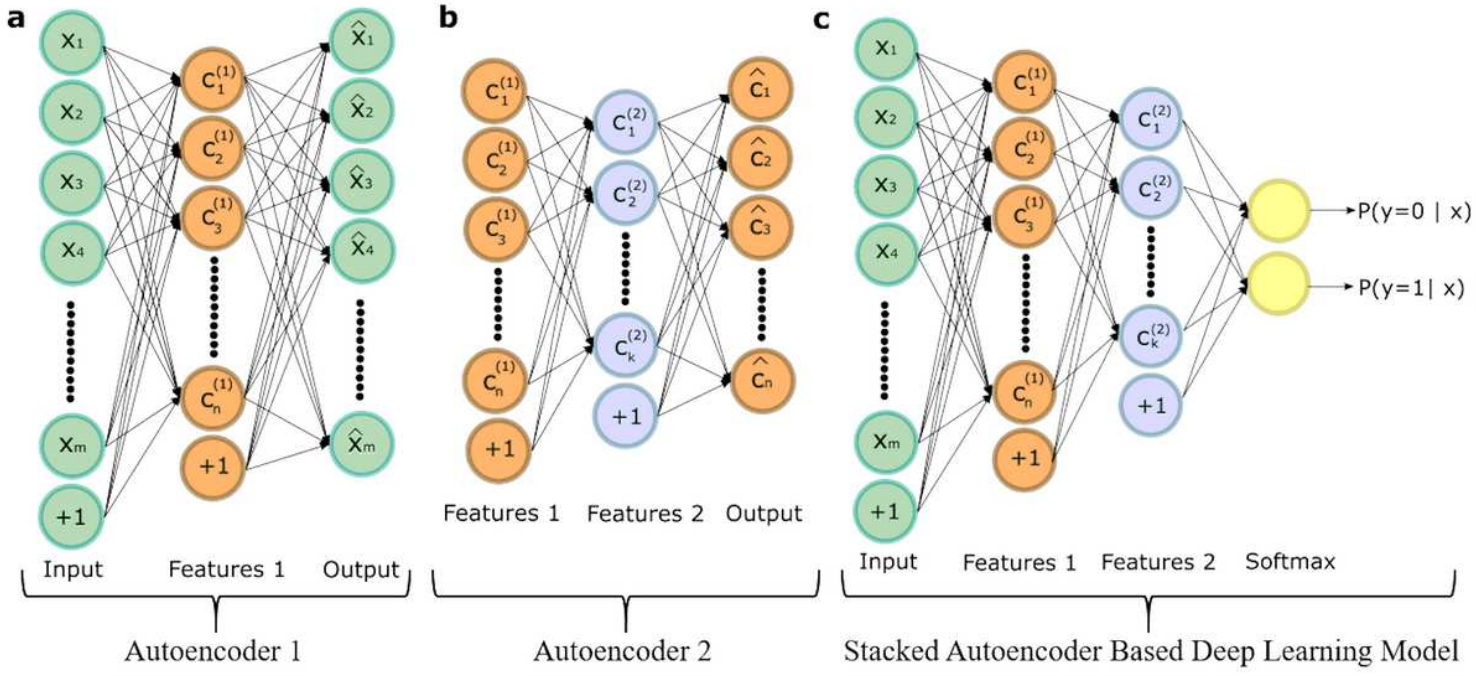


Figure 2

The structure of the proposed deep learning model. (a) First autoencoder that reveals the features from the input data. (b) Second autoencoder that reveal the new features from the first autoencoder's hidden

layer. (c) The deep learning model that is formed stacking the two autoencoders and softmax classifier.

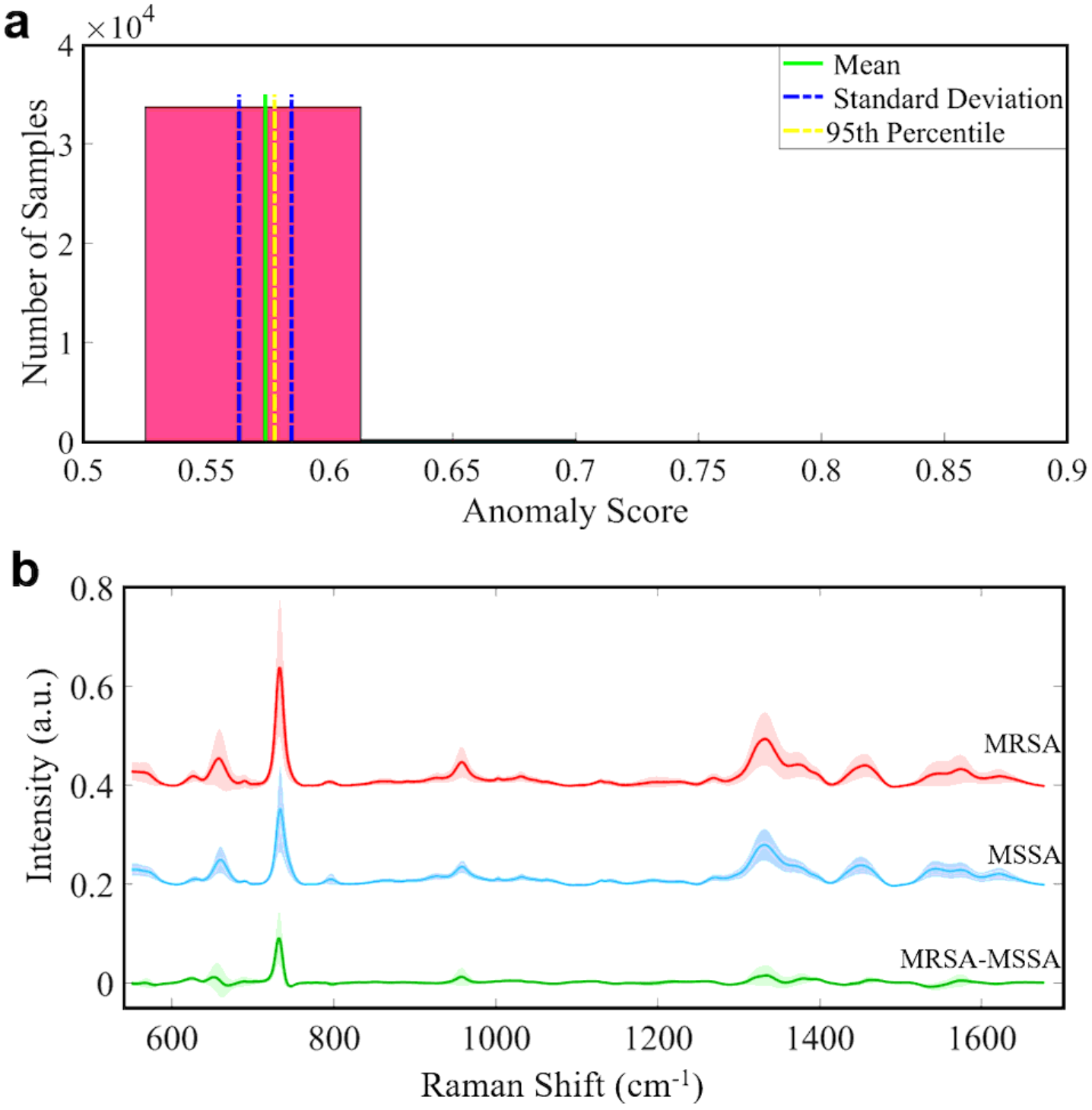


Figure 3

(a) Histogram plot of the anomaly score values obtained by the isolation forest algorithm. (b) Normalized average SERS spectra ± standard deviation of MRSA, MSSA and the difference of MRSA-MSSA.

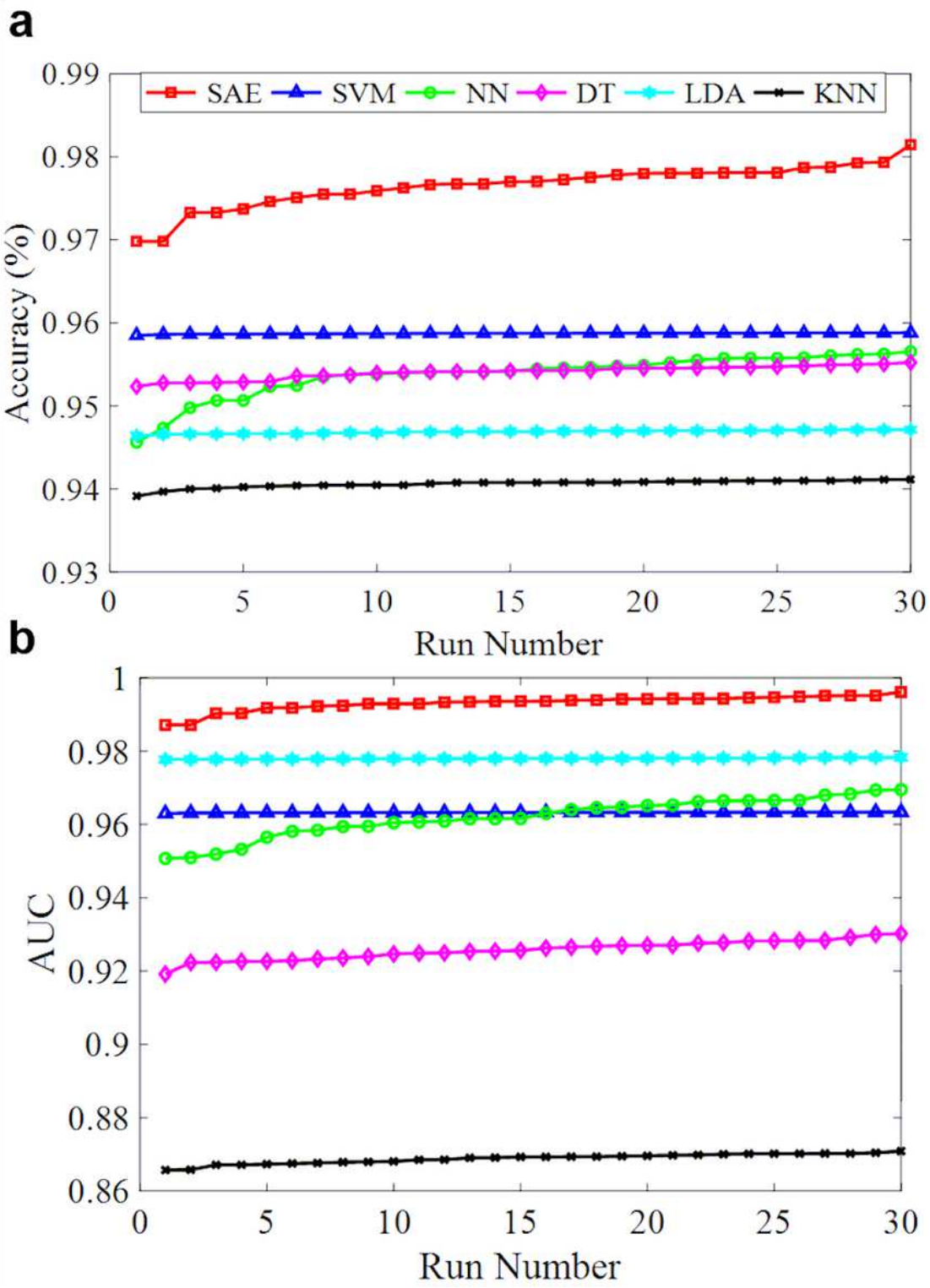


Figure 4

Performance comparisons of SAE-based DNN and traditional classifiers. (a) Accuracies of classifiers for 30 runs. (b) AUC values obtained from ROC curve of classifiers for 30 runs.

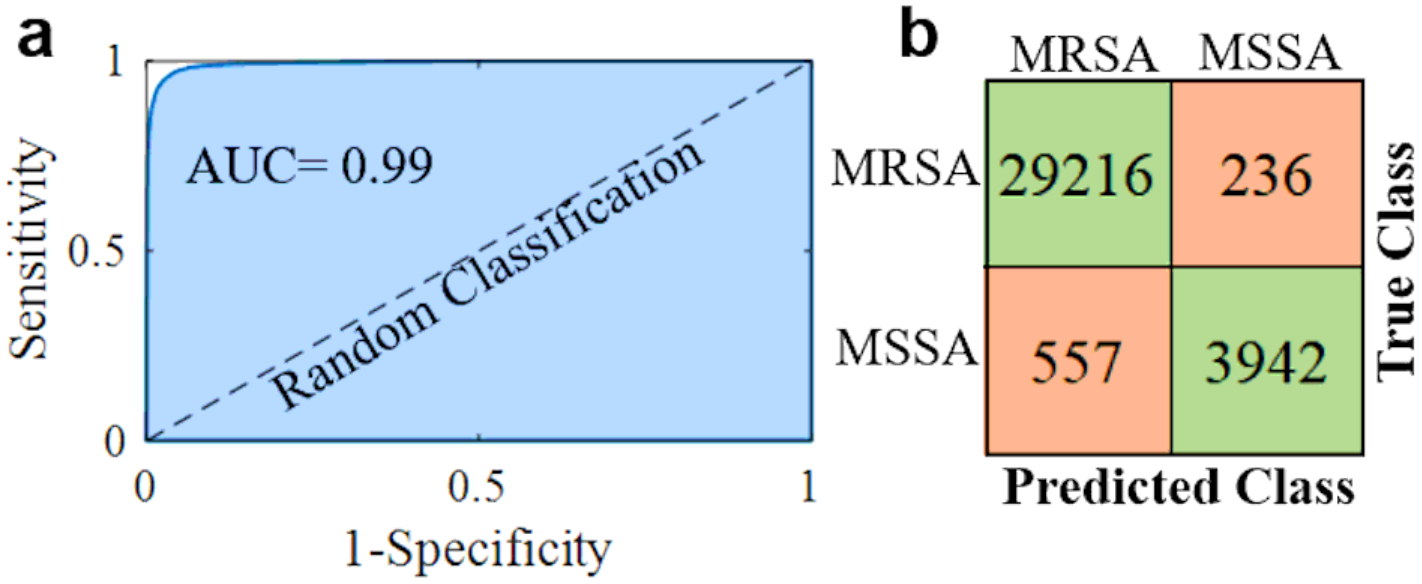


Figure 5

Binary classification results of MRSA and MSSA by the SAE based deep learning model. (a) The ROC curve with an AUC of 0.99. (b) Confusion matrix showing the results of 10-fold cross validated bacterial identification.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryinformationa.docx
- OnlineGraphicalAbstract.png