Rapidly predicting vancomycin resistance of *Enterococcus faecium* through MALDI-TOF MS spectrum obtained in real-world clinical microbiology laboratory

Hsin-Yao Wang, MD[1,2], Ko-Pei Lu, MS[3], Chia-Ru Chung, MS[4], Yi-Ju Tseng, PhD[1,5], Tzong-Yi Lee, PhD[1,6,7], Jorng-Tzong Horng, PhD[1,4,8], Tzu-Hao Chang, PhD[9,10], Min-Hsien Wu, PhD[11,12,13], Ting-Wei Lin, MD[1], Tsui-Ping Liu, MS[1], Jang-Jih Lu, MD, PhD[1,14,15]*

1. Department of Laboratory Medicine, Chang Gung Memorial Hospital at Linkou, Taoyuan City, Taiwan
2. Ph.D. Program in Biomedical Engineering, Chang Gung University, Taoyuan City, Taiwan
3. Graduate Program in Biomedical Information, Yuan-Ze University, Taoyuan City, Taiwan
4. Department of Computer Science and Information Engineering, National Central University, Taoyuan City, Taiwan
5. Department of Information Management, Chang Gung University, Taoyuan City, Taiwan
6. School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China
7. Warshel Institute for Computational Biology, The Chinese University of Hong Kong, Shenzhen, China
8. Department of Bioinformatics and Medical Engineering, Asia University, Taichung City, Taiwan
9. Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei City, Taiwan
10. Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei City, Taiwan.
11. Graduate Institute of Biomedical Engineering, Chang Gung University, Taoyuan City, Taiwan
12. Division of Haematology/Oncology, Department of Internal Medicine, Chang Gung Memorial Hospital at Linkou, Taoyuan City, Taiwan
13. Biosensor Group, Biomedical Engineering Research Center, Chang Gung University, Taoyuan City, Taiwan
14. School of Medicine, Chang Gung University, Taoyuan City, Taiwan
15. Department of Medical Biotechnology and Laboratory Science, Chang Gung University, Taoyuan City, Taiwan

*To whom correspondence should be addressed: JJ Lu

Keywords: Vancomycin-resistant *Enterococcus faecium* (VRE*fm*); Antibacterial drug resistance; Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry; Machine learning; Rapid detection

CORRESPONDENCE: Jang-Jih Lu, M.D., Ph.D.
Department of Laboratory Medicine, Chang Gung Memorial Hospital at Linkou
Department of Medical Biotechnology and Laboratory Science, Chang Gung University
5 Fu-Shing St. Kweishan
Taoyuan 333, Taiwan

Abstract

Enterococcus faecium is one of the leading pathogens in the world. In this study, we proposed a strategy to rapidly and accurately distinguish vancomycin-resistant *Enterococcus faecium* (VRE*fm*) and vancomycin-susceptible *E. faecium* (VSE*fm*) to help doctors correctly determine the use of vancomycin by a machine learning (ML)-based algorithm. A predictive model was developed and validated to distinguish VRE*fm* and VSE*fm* by analyzing MALDI-TOF MS spectra of unique *E. faecium* isolates from different specimen types. Firstly, 5717 mass spectra, including 2795 VRE*fm* and 2922 VSE*fm*, were used to develop the algorithm. And 2280 mass spectra of isolates, namely 1222 VRE*fm* and 1058 VSE*fm*, were used to externally validate the algorithm. The random forest-based algorithm demonstrated good classification performances for overall specimens, whose mean AUROC in 5-fold cross validation, time-wise validation, and external validation was all greater than 0.84. For the detection of VRE*fm* in blood, sterile body fluid, urinary tract, and wound, the AUROC in external validation was also greater than 0.84. The predictions with algorithms were significantly more accurate than empirical antibiotic use. The accuracy of antibiotics administration could be improved by 30%. And the algorithm could provide rapid antibiotic susceptibility results at least 24 hours ahead of routine laboratory tests. The turn-around-time of antibiotic susceptibility could be reduced by 50%. In conclusion, a ML algorithm using MALDI-TOF MS spectra obtained in routine workflow accurately differentiated VRE*fm* from VSE*fm*, especially in blood and sterile body fluid, which can be applied to facilitate the clinical testing process due to its accuracy, generalizability, and rapidness.

Introduction

Enterococcus spp. is one of the leading pathogens in healthcare-associated infection.[1] Enterococcal infection could cause urinary tract infection, blood stream infection, and even mortality.[2] Until recently, vancomycin was virtually the only drug that could be consistently relied on for treating multidrug-resistant enterococcal infections[3,4]. Vancomycin-resistant *Enterococcus* (VRE) has led to heavy burden on healthcare worldwide since its first-time isolation.[5,6] *Enterococcus faecalis* and *E. faecium* are the 2 most commonly isolated *Enterococcus* spp. in clinical practice.[1] VRE *faecium* (VRE*fm*) has received considerably more attention than VRE *faecalis* (VRE*fs*) because most of the clinically isolated VRE is *E. faecium* in the recent decades[4,7] and VRE*fm* causes more severe infection than VRE*fs*[8,9]. Early detection of vancomycin resistance is essential for successfully treating VRE*fm* infection.[10] Vancomycin could be discontinued, and antimicrobial agents could be replaced with other antibiotics (eg, linezolid and daptomycin) based on the laboratory results of vancomycin resistance[11,12]. Patients' prognosis could be improved and further drug resistance development could be avoided by using susceptible antibiotics.[11] However, typical tests in clinical microbiology laboratories, such as the minimal inhibitory concentration test or agar-diffusion test, fail to provide results for antibiotic susceptibility rapidly. The antibiotic susceptibility test (AST) of vancomycin is time-consuming, and the Clinical and Laboratory Standards Institute recommended a full 24 hours should be held for accurate detection of vancomycin resistance in enterococci.[13] This would considerably delay accurate prescription of antibiotics against *E. faecium*. Furthermore, prescribing antibiotics based on empirical prescription, without determining AST, would result in low effectiveness (approximately 50%), depending on the local epidemiology of VRE*fm*.[12] Thus, a new tool is needed to provide AST for VRE*fm* rapidly and accurately.

Recently, matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) has become popular among clinical microbiology laboratories worldwide because of its reliability, rapidity, and cost-effectiveness in identifying bacterial species.[14-16] In addition to species identification, MALDI-TOF MS has been promising in other applications, such as strain typing or AST.[17-19] MALDI-TOF MS can generate massive data comprising hundreds of peak signals on the spectra.[17,20] The complex data of MALDI-TOF spectra are overwhelming to even an experienced medical staff.[19] Studies have attempted to identify the characteristic peak through visual inspection.[21,22] The results of the studies have been discordant, which has limited the clinical utility.[23-25]

Machine learning (ML) is a good analytical method for solving classification problems through identification of implicit data patterns from complex data.[26] The ML method outperforms traditional statistical methods because of its excellent ability to handle complex interactions between large amount of predictors and good performance in non-linear classification problems[27] ML has been successfully applied in several clinical fields.[27-36] Thus, the ML algorithm is especially appropriate for analyzing complex data such as MALDI-TOF spectra. However, to our knowledge, few studies have used ML in the analysis of MALDI-TOF spectra for rapidly reporting VRE*fm*, and the case numbers in these studies were insufficient, and so, ML algorithm generalization has been limited.[37-39] Moreover, to date, no study has validated AST prediction ML algorithms by using large real-world data.

In this study, we aimed to develop and validate a VRE*fm* prediction ML model by using consecutively collected real-world data from 2 tertiary medical centers (Chang Gung Memorial Hospital [CGMH], Linkou branch and CGMH, Kaohsiung branch). Using the largest MALDI-TOF spectrum clinical data to date, the ML algorithm could predict VRE*fm* accurately, rapidly, and in a ready-to-use manner based on the real-world evidence, which is more representative for clinical practice.[40] Moreover, we confirmed the robustness and generalization of the ML

121 algorithm through several validation methods, namely cross-validation, time-wise internal
122 validations (unseen independent testing dataset classified according to time), and external
123 validation (unseen independent testing dataset from another medical center). According to the
124 real-world evidence-based validation, our VRE*fm* prediction ML models are ready to be
125 incorporated into routine workflow.

126 ## Materials and methods

127 ### Data source

128 We designed a novel machine learning approach which can improve accuracy of
129 antibiotics administration and reduce the turn-around-time of antibiotics susceptibility test. We
130 summarized the comparison between the machine learning approach and the traditional
131 approach used in current clinical microbiology laboratory. We schematically illustrated the
132 study design in Figure 1(b). The data used in this retrospective study was consecutively
133 collected from the clinical microbiology laboratories of 2 tertiary medical centers in Taiwan,
134 namely CGMH Linkou branch and CGMH Kaohsiung branch between January 1, 2013 and
135 December 31, 2017. The clinical microbiology laboratories collected and processed all the
136 routine specimens obtained from the hospitals. In total, 7997 *E. faecium* cases were identified
137 and included in this study, whereas 5717 (VRE*fm*: 48.89%) and 2280 (VRE*fm*: 53.60%) cases,
138 respectively, were obtained from Linkou and Kaohsiung branches of CGMH. The *E. faecium*
139 strains were isolated from blood, urinary tract, sterile body fluids, and wound. The detailed
140 description of specimen types is provided in eTable 1 in the Supplement. The study was
141 approved by the Institutional Review Board of Chang Gung Medical Foundation (No.
142 201900767B0). We followed the Standards for Reporting of Diagnostic Accuracy 2015[41] and
143 the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or
144 Diagnosis reporting guidelines.[42]
145

146 **Definition of *E. faecium* and vancomycin susceptibility**

147 *E. faecium* was identified using MALDI-TOF spectra measured using a Microflex LT
148 mass spectrometer and analyzed using Biotyper 3.1 (Bruker Daltonik GmbH, Bremen,
149 Germany). A log score (generated through Biotyper 3.1) larger than 2 was used for confirming
150 *E. faecium*.[17-19] We tested vancomycin susceptibility of *E. faecium* by using the paper disc
151 method. The details of *E. faecium* identification and AST are given in the eMethods in the
152 Supplement.

153

154 **MALDI-TOF mass spectrum data collection and preprocessing**

155 The details were described in the Supplements.

156

157 **Peak selection from MALDI-TOF mass spectra for model development**

158 We applied the embedded feature-selection method to select the most important peaks
159 from MALDI-TOF mass spectra.[43] The peaks were ranked using the p-values of the chi-square
160 test of homogeneity, which was employed to determine whether frequency counts were
161 distributed identically across VRE*fm* and vancomycin-susceptible *E. faecium* (VSE*fm*).
162 Preliminarily, we selected top 10 important peaks to plot a heat map based on the hierarchical
163 clustering (eMethods in the Supplement). All the ranked peaks were incorporated in the model
164 accordingly until the performance did not increase. Consequently, we could obtain the
165 important peaks that were highly related to differentiation of VRE*fm* and VSE*fm* isolates.

166 For determining the number of peaks included in the ML models, we forwardly added
167 them into the ML models and calculated the performance using accuracy as the metric. First,
168 the predictor candidates were sorted in a descending order according to the importance score,
169 and one predictive peak was added at a time into the ML models. On the basis of predictive
170 peak composition, we used different algorithms, namely random forest (RF), support vector

machine (SVM) with a radial basis function kernel, and k-nearest neighbor (KNN) and applied 5-fold cross validation (CV) in the data from the CGMH Linkou branch. The accuracies of the ML models were calculated to determine the adequate number of predictive peaks included in the models.

Development and validation of VRE*fm* prediction models

We aimed to develop and validate a robust VRE*fm* prediction model capable of detecting VRE*fm* earlier than the AST report. Three commonly used ML algorithms, namely RF, SVM with a radial basis function kernel, and KNN, were used for developing the VRE*fm* prediction model. These ML algorithms have demonstrated their successful applications (either classification or prediction) in clinical practice.[17-19,27,28,35,36] The details of these ML algorithms and model training processes are attached in the eMethods in the Supplement.

We thoroughly evaluated the performance and robustness of the VRE*fm* prediction models using 5-fold CV, time-wise internal validation, and external validation. Data from the CGMH Linkou branch were used for 5-fold CV and time-wise internal validation; by contrast, data from the CGMH Kaohsiung branch served as the unseen independent testing data for external validation. For 5-fold CV, data were randomly divided into 5 datasets. Each one of the 5 datasets served as the testing dataset to evaluate the performance of the model developed by the other 4 datasets. In 5-fold CV, we obtained 5 measurements of metrics for evaluating the robustness of VRE*fm* prediction models. Moreover, to evaluate performance using prospectively collected data, we conducted time-wise internal validation: we used data collected between January 1, 2013 and December 31, 2016 as the training dataset for developing VRE*fm* prediction models, while data from January 1, 2017 to December 31, 2017 served as the testing dataset. To test the generalizability of the models, we used data from the CGMH Linkou branch to develop VRE*fm* prediction models and used data from the CGMH

196 Kaohsiung branch to test the models' performance in a different institute. Additionally, we
197 evaluated the performance of the VRE*fm* prediction model using different types of specimens,
198 namely blood, urinary tract, sterile body fluid, and wound, by using data from the CGMH
199 Kaohsiung branch. We adopted metrics including sensitivity, specificity, accuracy, positive
200 predictive value (PPV), negative predictive value (NPV), receiver operating characteristic
201 (ROC) curve, and area under the receiver operating characteristic curve (AUROC) to access
202 and compare the performance of the VRE*fm* prediction model.
203
204 **Statistical analysis**
205 The confidence intervals for sensitivity, specificity, and accuracy were estimated
206 using the calculation of the confidence interval for a proportion in one sample situation.
207 Specifically, the critical values followed the Z-score table. To compare the percentages in
208 matched samples, Cochran's Q test, a nonparametric approach, was implemented in this
209 study.[44] Then, we employed pairwise McNemar's tests[45] for post hoc analysis and adopted
210 the false discovery rate proposed by Benjamini and Hochberg (1995) to adjust the *P* value.[46]
211 Furthermore, the confidence intervals of AUROCs were determined using the nonparametric
212 approach, and the AUROC comparisons mainly adopted the nonparametric approach
213 proposed by Delong et al.[47]
214
215 # Results
216 **Predictive peaks for detecting VRE*fm***
217 We defined crucial predictive peaks when the occurrence frequency of a peak was
218 significantly different (defined by the chi-square test) in VRE*fm* and VSE*fm*. In the step of
219 extracting predictor candidates, 876 predictor candidates were extracted. From the predictor
220 candidates, we used the chi-square method to select important predictive peaks.

We selected 10 most critical predictive peaks and plotted a heat map to preliminarily visualize the difference between VRE*fm* and VSE*fm* (Figure 2). Peaks of *m/z* 3172, 3302, 3645, 6342, 6356, 6603, and 6690 were found more frequently in VRE*fm*; by contrast, *m/z* 3165, 3681, and 7360 occurred more frequently in VSE*fm*. Although these important predictive peaks were statistically significant, we found them in both VRE*fm* and VSE*fm*. The full list of crucial predictive peaks is provided in eTable 2 in the Supplement.

We selected several important predictive peaks from the predictor candidate list, which was ordered according to the chi-square score. eFigure 4 in the Supplement shows the change in ML models performance when the number of critical predictive peaks increased. For all the ML algorithms used in the study, a similar trend of performance was observed: the accuracies of the ML models reached a steady plateau when the included number of important predictive peaks was larger than 100 (eFigure 4 in the Supplement). Thus, the top 100 crucial predictive peaks were selected as the peak composition for the following experiments.

Performance of VRE*fm* prediction models

We summarized the ML models' performance in Table 1, Table 2, and Figure 3. The details of comparison between different algorithms are described in the Supplement. The RF model outperformed SVM and KNN in 5-fold CV, time-wise internal validation, and external validation (eTable 3 in the Supplement), where the AUROC ranged from 0.8463 to 0.8553 and accuracy ranged from 0.7769 to 0.7855. Moreover, performance robustness was also observed in SVM and KNN. Figure 3 shows typical ROC curves for the 3 algorithms in all the 3 validations. We used Youden's index to select the threshold from the ROC curves in search of balanced sensitivity and specificity. In external validation, the sensitivity and specificity of RF were 0.7791 (95% confidence interval: 0.7620-0.7961) and 0.7930 (95% confidence interval:

0.7764-0.8096). On the basis of the resistance rate (VRE*fm*: 53.60%) in the external validation dataset, the PPV was 0.8130 and the NPV was 0.7565.

Given that the RF algorithm attained the highest performance, additionally, we tested the performance of the RF-based VRE*fm* prediction model using different types of specimens in the independent testing dataset (ie, external validation by using data of the CGMH Kaohsiung branch) (Table 2). The RF-based VRE*fm* prediction model attained higher performance in predicting VRE*fm* in blood and sterile body fluid specimens than the other specimen types. The AUROC of blood specimens reached 0.9103 (95% confidence interval: 0.8727-0.9480), whereas that of sterile body fluid specimens reached 0.8714 (95% confidence interval: 0.8321-0.9106). Moreover, the sensitivity (0.8870, 95% confidence interval: 0.8436-0.9303) and specificity (0.8000, 95% confidence interval: 0.7452-0.8548) of the RF-based VRE*fm* prediction model for the blood specimen were also balanced and significantly higher than those for other specimens. By contrast, the performance of the RF-based VRE*fm* prediction model for urinary tract specimens (0.8494, 95% confidence interval: 0.8258-0.8731) was similar to that for overall specimens (0.8553, 95% confidence interval: 0.8399-0.8706).

Discussion

We developed ML-based models for predicting VRE*fm* rapidly and accurately based on MALDI-TOF MS data. The models were especially effective in predicting VRE*fm* in invasive infections (ie, blood and sterile body fluid). We used the largest up-to-date real-world data to validate the robustness and generalization of the ML-based models by using k-fold CV, time-wise internal validation, and external validation. The rapid and accurate AST of vancomycin is promising for determining antibiotics against VRE*fm* infection.

Our results suggested that AST could be predicted accurately by using ML algorithms to analyze MALDI-TOF MS data. MALDI-TOF MS is a powerful analytical tool in current

clinical microbiology laboratories because of its rapidness and cost-effectiveness in identifying bacterial species.[14-16] On the basis of the massive data produced by MALDI-TOF MS, moreover, some studies have demonstrated that subspecies typing could be predicted from a specific pattern of MS spectra only.[17,19] Furthermore, other studies have shown a good correlation between AST and specific patterns of MS spectra.[18,23-25,48] However, some issues have limited the generalization of these results. First, most of the studies have adopted an additional protein extraction step before analytical measurement of MALDI-TOF MS. The protein extraction step could enhance data quality; however, it is not routinely used in clinical practice because it is labor-intensive, time-consuming, and expensive.[17,18] By contrast, we used the direct deposition method, which is recommended by the manufacturer and is used for everyday works. Thus, our models are more feasible for the existing workflow because they were trained using real-world data. Second, the data sizes in these studies were too small to be representative. We demonstrated that the ML-based models for predicting VRE*fm* can be applied as a clinical decision support tool by using the largest up-to-date datasets collected through the direct deposition method and various validation methods.

Identifying crucial predictive peaks in VRE*fm* classification may not be essential in clinical application; however, the specific combination of crucial predictive peaks would inspire further studies investigating the molecular mechanism of VRE*fm*. Typically, the *vanA* cluster is the most common mediator of vancomycin resistance in enterococci,[49] although many vancomycin resistance genes have been identified.[50] In brief, many factors together attribute to antibiotic resistance. Moreover, the complex mechanisms of antibiotic resistance would evolve in response to the selective pressures of their competitive environment (eg, antibiotic use).[49] Thus, identifying the important predictive peaks for VRE*fm* could help us understand the mechanism behind resistance. In this study, for example, peaks of *m/z* 6603, 6631, and 6635 were found frequently for VRE*fm* (eTable 2 in the Supplement). The finding is consistent with

a previous study where Griffin et al. reported *m/z* 6603 is specific for *vanB*-positive VRE*fm*, while *m/z* 6631 and 6635 are specifically found for *vanA*-positive VRE*fm*.[38] These peaks are worthy of further identification in future investigations. Moreover, new antibiotics against VRE*fm* can be developed based on these predictive peaks for VRE*fm*.

Our ML models persistently performed well in 5-fold CV, time-wise internal validation, and external validation. Moreover, all the ML algorithms used in this study exhibited good performance (AUROC > 0.8). It could be explained that discriminating VRE*fm* from VSE*fm* is generally achievable after adequate feature extraction and feature selection processes. In time-wise internal validation, we intended to simulate a prospective study for a model trained by the "past data" to analyze the "future data." Based on the performance of time-wise internal validation, we concluded that the trained ML models could also perform well on the prospectively collected data, which are unseen in the training process. Previous study results differentiating VRE*fm* from VSE*fm* by using MALDI-TOF MS spectra could not be generalized.[23-25,38] The inconsistent results could be because less features (<10) were used. A review article reported that peak-level reproducibility of MALDI-TOF mass is approximately 80%.[51] The classification performance is compromised when essential peaks are few and happen to be absent on the mass spectra. In our study, the ML models performed stably when the included peaks were more than 100 (eFigure 4 in the Supplement). The steady and good performance of our ML models could be explained by the much more included peaks: when some of the essential peaks are not reproduced in the mass spectra, we can still use other alternative essential peaks to conduct an accurate classification. The number of essential peaks somehow compensated the insufficient reproducibility of MALDI-TOF mass. By contrast, regarding predicting VRE*fm* for various specimens, we found that the RF-based model performed especially well in blood and sterile body fluids. The superior prediction performance could be attributed to the relatively fewer number of VRE*fm* strains in blood and sterile body

fluids. Bacterial infection in blood or sterile body fluids is typically regarded as invasive infection.[52] Only a few VRE*fm* strains (sequence type (ST)17, ST18, ST78, and ST203) cause invasive infections in blood or sterile body fluids according to the studies in Taiwan[53] and Ireland.[54] The nature of the classification problem would be more simple when the number of labels is fewer.

Limitations

This study has several limitations. First, although the models were evaluated using unseen external data from different medical centers, all the training data and testing data were collected from only 2 tertiary medical centers in Taiwan. Directly applying the ML models in hospitals of other areas or countries as well as in primary care institutes may not be appropriate. However, we believe that the method, but not the trained model, could be generalized. Although our ML models were validated comprehensively using 3 different approaches and the results show that the difference in MALDI-TOF mass spectra between VRE*fm* and VSE*fm* can be distinguished through all the ML algorithms we used, we suggest others collecting their locally relevant data for training and validating the VRE*fm* predicting model given that the epidemiology of VRE*fm* could be fairly different site by site. Second, our primary goal was to develop and validate a practical and ready-to-use ML model in real-world practice. We found some crucial predictive peaks for VRE*fm*; however, we did not confirm the identities for these peaks. It is worthy of identifying these peaks in further investigations. Third, we did not use the deep learning (DL) algorithm for predicting VRE*fm*, although DL has been successful in the image classification or radiology field.[32,33] In this study, VRE*fm* could be accurately predicted using several classic algorithms (ie, RF, SVM, and KNN) that require less resource and time in training and using models. Moreover, DL usually requires more training samples and is financially and computationally more expensive than classical ML algorithms.[55] DL

utility in analyzing MS data rather than image data could be another promising issue in the bioinformatics field. Fourth, no strain typing data were included. Thus, the molecular epidemiology of VRE*fm* used in this study is unknown.

Conclusions

We developed and validated robust ML models capable of discriminating VRE*fm* from VSE*fm* based on MALDI-TOF MS spectra. These models were especially good at detecting VRE*fm* causing invasive diseases. The accurate and rapid detection of VRE*fm* by using the ML models would facilitate more appropriate antibiotic prescription.

Acknowledgments

This manuscript was edited by Wallace Academic Editing.

Author Contributions

HYW, KPL, and CRC had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of data analysis. HYW, KPL, CRC, and YJT analyzed/interpreted the data, performed experiments, designed the study, and wrote the manuscript. HYW, CRC, YJT, JTH, TYL, THC, MHW, TPL, and JJL reviewed/edited the manuscript for important intellectual content and provided administrative, technical, or material support. JJL obtained funding and supervised the study.

Funding

This work was supported by Chang Gung Memorial Hospital (CMRPG3F1721, CMRPG3F1722, CMRPD3I0011) and the Ministry of Science and Technology, Taiwan

(MOST 107-2320-B-182A-021-MY3, MOST 108-2636-E-182-001, and MOST 107-2636-E-182-001).

Competing interests

The authors have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

1 Arias, C. A. & Murray, B. E. The rise of the Enterococcus: beyond vancomycin resistance. *Nat Rev Microbiol* **10**, 266-278, doi:10.1038/nrmicro2761 (2012).

2 Marra, A. R. *et al.* Nosocomial Bloodstream Infections in Brazilian Hospitals: Analysis of 2,563 Cases from a Prospective Nationwide Surveillance Study. *Journal of Clinical Microbiology* **49**, 1866-1871, doi:10.1128/JCM.00376-11 (2011).

3 Cetinkaya, Y., Falk, P. & Mayhall, C. G. Vancomycin-Resistant Enterococci. *Clinical Microbiology Reviews* **13**, 686-707, doi:10.1128/CMR.13.4.686 (2000).

4 Arias, C. A., Contreras, G. A. & Murray, B. E. Management of multidrug-resistant enterococcal infections. *Clin Microbiol Infect* **16**, 555-562, doi:10.1111/j.1198-743X.2010.03214.x (2010).

5 Leclercq, R., Derlot, E., Duval, J. & Courvalin, P. Plasmid-mediated resistance to vancomycin and teicoplanin in Enterococcus faecium. *N Engl J Med* **319**, 157-161, doi:10.1056/NEJM198807213190307 (1988).

6 Sahm, D. F. *et al.* In vitro susceptibility studies of vancomycin-resistant Enterococcus faecalis. *Antimicrobial Agents and Chemotherapy* **33**, 1588-1591, doi:10.1128/aac.33.9.1588 (1989).

7 Sader, H. S., Moet, G. J., Farrell, D. J. & Jones, R. N. Antimicrobial susceptibility of daptomycin and comparator agents tested against methicillin-resistant Staphylococcus aureus and vancomycin-resistant enterococci: trend analysis of a 6-year period in US medical centers (2005–2010). *Diagnostic Microbiology and Infectious Disease* **70**, 412-416, doi:10.1016/j.diagmicrobio.2011.02.008 (2011).

8 Lodise, T. P., McKinnon, P. S., Tam, V. H. & Rybak, M. J. Clinical outcomes for patients with bacteremia caused by vancomycin-resistant enterococcus in a level 1 trauma center. *Clin Infect Dis* **34**, 922-929, doi:10.1086/339211 (2002).

9 Ghanem, G., Hachem, R., Jiang, Y., Chemaly, R. F. & Raad, I. Outcomes for and Risk Factors Associated With Vancomycin-Resistant Enterococcus faecalis and Vancomycin-Resistant Enterococcus faecium Bacteremia in Cancer Patients. *Infection Control & Hospital Epidemiology* **28**, 1054-1059, doi:10.1086/519932 (2015).

10 Ozsoy, S. & Ilki, A. Detection of vancomycin-resistant enterococci (VRE) in stool specimens submitted for Clostridium difficile toxin testing. *Braz J Microbiol* **48**, 489-492, doi:10.1016/j.bjm.2016.12.012 (2017).

11 Balli, E. P., Venetis, C. A. & Miyakis, S. Systematic Review and Meta-Analysis of Linezolid versus Daptomycin for Treatment of Vancomycin-Resistant Enterococcal Bacteremia. *Antimicrobial Agents and Chemotherapy* **58**, 734-739, doi:10.1128/AAC.01289-13 (2014).

12 Crank, C. & O'Driscoll, T. Vancomycin-resistant enterococcal infections: epidemiology, clinical manifestations, and optimal management. *Infection and Drug Resistance*, doi:http://dx.doi.org/10.2147/IDR.S54125 (2015).

13 CLSI. Performance Standards for Antimicrobial Susceptibility Testing. 27th ed. CLSI supplement M100. *Wayne, PA: Clinical and Laboratory Standards Institute* (2017).

14 Hrabak, J., Chudackova, E. & Walkova, R. Matrix-assisted laser desorption ionization-time of flight (maldi-tof) mass spectrometry for detection of antibiotic resistance mechanisms: from research to routine diagnosis. *Clin Microbiol Rev* **26**, 103-114, doi:10.1128/CMR.00058-12 (2013).

15 Idelevich, E. A., Sparbier, K., Kostrzewa, M. & Becker, K. Rapid detection of antibiotic resistance by MALDI-TOF mass spectrometry using a novel direct-on-target microdroplet growth assay. *Clin Microbiol Infect*, doi:10.1016/j.cmi.2017.10.016 (2017).

16 Suarez, S. *et al.* Ribosomal proteins as biomarkers for bacterial identification by mass spectrometry in the clinical microbiology laboratory. *J Microbiol Methods* **94**, 390-396, doi:10.1016/j.mimet.2013.07.021 (2013).

17 Wang, H.-Y. *et al.* Application of a MALDI-TOF analysis platform (ClinProTools) for rapid and preliminary report of MRSA sequence types in Taiwan. *PeerJ* **6**, doi:10.7717/peerj.5784 (2018).

18 Wang, H. Y. *et al.* Rapid Detection of Heterogeneous Vancomycin-Intermediate Staphylococcus aureus Based on Matrix-Assisted Laser Desorption Ionization Time-of-Flight: Using a Machine Learning Approach and Unbiased Validation. *Front Microbiol* **9**, 2393, doi:10.3389/fmicb.2018.02393 (2018).

19 Wang, H. Y. *et al.* A new scheme for strain typing of methicillin-resistant Staphylococcus aureus on the basis of matrix-assisted laser desorption ionization time-of-flight mass spectrometry by using machine learning approach. *PLoS One* **13**, e0194289, doi:10.1371/journal.pone.0194289 (2018).

20 Lopez-Fernandez, H. *et al.* Mass-Up: an all-in-one open software application for MALDI-TOF mass spectrometry knowledge discovery. *BMC Bioinformatics* **16**, 318, doi:10.1186/s12859-015-0752-4 (2015).

21 Lasch, P. *et al.* Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of Enterococcus faecium and Staphylococcus aureus isolates. *Journal of microbiological methods* **100**, 58-69, doi:10.1016/j.mimet.2014.02.015 (2014).

22 Wolters, M. *et al.* MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant Staphylococcus aureus lineages. *International journal of medical microbiology : IJMM* **301**, 64-68, doi:10.1016/j.ijmm.2010.06.002 (2011).

23 Burckhardt, I. & Zimmermann, S. Susceptibility Testing of Bacteria Using Maldi-Tof Mass Spectrometry. *Front Microbiol* **9**, 1744, doi:10.3389/fmicb.2018.01744 (2018).

24 Vrioni, G. *et al.* MALDI-TOF mass spectrometry technology for detecting biomarkers of antimicrobial resistance: current achievements and future perspectives. *Ann Transl Med* **6**, 240, doi:10.21037/atm.2018.06.28 (2018).

25 Kostrzewa, M., Sparbier, K., Maier, T. & Schubert, S. MALDI-TOF MS: an upcoming tool for rapid detection of antibiotic resistance in microorganisms. *Proteomics Clin Appl* **7**, 767-778, doi:https://doi.org/10.1002/prca.201300042 (2013).

26 Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J. *Data Mining: Practical machine learning tools and techniques*. (Morgan Kaufmann, 2016).

27 Lo-Ciganic, W.-H. *et al.* Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions. *JAMA Network Open* **2**, doi:10.1001/jamanetworkopen.2019.0968 (2019).

28 Tseng, Y.-J. *et al.* Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies. *International Journal of Medical Informatics*, doi:10.1016/j.ijmedinf.2019.05.003 (2019).

29 Kuppermann, N. *et al.* A Clinical Prediction Rule to Identify Febrile Infants 60 Days and Younger at Low Risk for Serious Bacterial Infections. *JAMA Pediatr*, doi:10.1001/jamapediatrics.2018.5501 (2019).

30 Norgeot, B. *et al.* Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. *JAMA Network Open* **2**, doi:10.1001/jamanetworkopen.2019.0606 (2019).

31 Karter, A. J. *et al.* Development and Validation of a Tool to Identify Patients With Type 2 Diabetes at High Risk of Hypoglycemia-Related Emergency Department or Hospital Use. *JAMA Internal Medicine* **177**, doi:10.1001/jamainternmed.2017.3844 (2017).

32 Gulshan, V. *et al.* Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *Jama* **316**, doi:10.1001/jama.2016.17216 (2016).

33 Hwang, E. J. *et al.* Development and Validation of a Deep Learning-Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs. *JAMA Netw Open* **2**, e191095, doi:10.1001/jamanetworkopen.2019.1095 (2019).

34 Elfiky, A. A., Pany, M. J., Parikh, R. B. & Obermeyer, Z. Development and Application of a Machine Learning Approach to Assess Short-term Mortality Risk Among Patients With Cancer Starting Chemotherapy. *JAMA Netw Open* **1**, e180926, doi:10.1001/jamanetworkopen.2018.0926 (2018).

35 Lin, W. Y. *et al.* Predicting post-stroke activities of daily living through a machine learning-based approach on initiating rehabilitation. *Int J Med Inform* **111**, 159-164, doi:10.1016/j.ijmedinf.2018.01.002 (2018).

36 Wang, H. Y. *et al.* Cancers Screening in an Asymptomatic Population by Using Multiple Tumour Markers. *PLoS One* **11**, e0158285, doi:10.1371/journal.pone.0158285 (2016).

37 Nakano, S. *et al.* Differentiation of vanA-positive Enterococcus faecium from vanA-negative E. faecium by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Int J Antimicrob Agents* **44**, 256-259, doi:10.1016/j.ijantimicag.2014.05.006 (2014).

38 Griffin, P. M. *et al.* Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. *J Clin Microbiol* **50**, 2918-2931, doi:10.1128/JCM.01000-12 (2012).

39 Huang, T. S. *et al.* Evaluation of a matrix-assisted laser desorption ionization-time of flight mass spectrometry assisted, selective broth method to screen for vancomycin-resistant enterococci in patients at high risk. *PLoS One* **12**, e0179455, doi:10.1371/journal.pone.0179455 (2017).

40 Corrigan-Curay, J., Sacks, L. & Woodcock, J. Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. *Jama* **320**, doi:10.1001/jama.2018.10136 (2018).

41 Bossuyt, P. M. *et al.* STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* **351**, h5527, doi:10.1136/bmj.h5527 (2015).

42 Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* **350**, g7594, doi:10.1136/bmj.g7594 (2015).

43 Saeys, Y., Inza, I. & Larranaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **23**, 2507-2517, doi:10.1093/bioinformatics/btm344 (2007).

44 Cochran, W. G. The Comparison of Percentages in Matched Samples. *Biometrika* **37**, doi:10.2307/2332378 (1950).

45 McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **12**, 153-157, doi:10.1007/BF02295996 (1947).

46 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289-300, doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x (1995).

47 DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* **44**, doi:10.2307/2531595 (1988).

48 Mather, C. A., Werth, B. J., Sivagnanam, S., SenGupta, D. J. & Butler-Wu, S. M. Rapid Detection of Vancomycin-Intermediate Staphylococcus aureus by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. *J Clin Microbiol* **54**, 883-890, doi:10.1128/JCM.02428-15 (2016).

49 Miller, W. R., Munita, J. M. & Arias, C. A. Mechanisms of antibiotic resistance in enterococci. *Expert Review of Anti-infective Therapy* **12**, 1221-1236, doi:10.1586/14787210.2014.956092 (2014).

50 Lebreton, F. *et al.* D-Ala-d-Ser VanN-type transferable vancomycin resistance in Enterococcus faecium. *Antimicrob Agents Chemother* **55**, 4606-4612, doi:10.1128/AAC.00714-11 (2011).

51 Croxatto, A., Prod'hom, G. & Greub, G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS microbiology reviews* **36**, 380-407, doi:10.1111/j.1574-6976.2011.00298.x (2012).

52 Lee, J. H. *et al.* Etiology of invasive bacterial infections in immunocompetent children in Korea (1996-2005): a retrospective multicenter study. *J Korean Med Sci* **26**, 174-183, doi:10.3346/jkms.2011.26.2.174 (2011).

53 Kuo, A. J. *et al.* Vancomycin-resistant Enterococcus faecium at a university hospital in Taiwan, 2002-2015: Fluctuation of genetic populations and emergence of a new structure type of the Tn1546-like element. *J Microbiol Immunol Infect* **51**, 821-828, doi:https://doi.org/10.1016/j.jmii.2018.08.008 (2018).

54 Ryan, L. *et al.* Epidemiology and molecular typing of VRE bloodstream isolates in an Irish tertiary care hospital. *J Antimicrob Chemother* **70**, 2718-2724, doi:https://doi.org/10.1093/jac/dkv185 (2015).

55 Liu, P., Choo, K.-K. R., Wang, L. & Huang, F. SVM or deep learning? A comparative study on remote sensing image classification. *Soft Computing* **21**, 7053-7065, doi:https://doi.org/10.1007/s00500-016-2247-2 (2016).

Figure legends

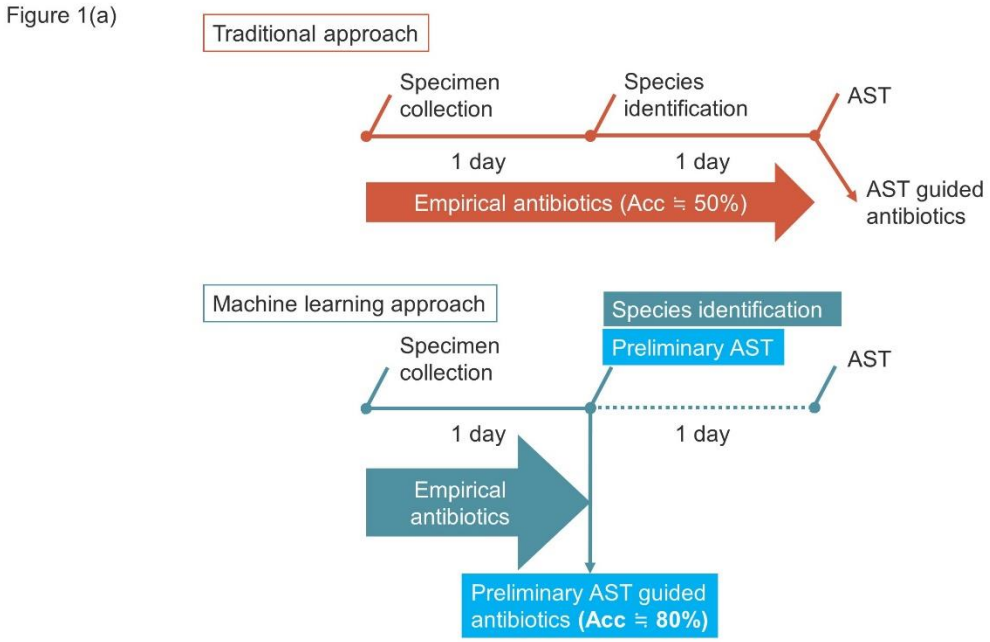
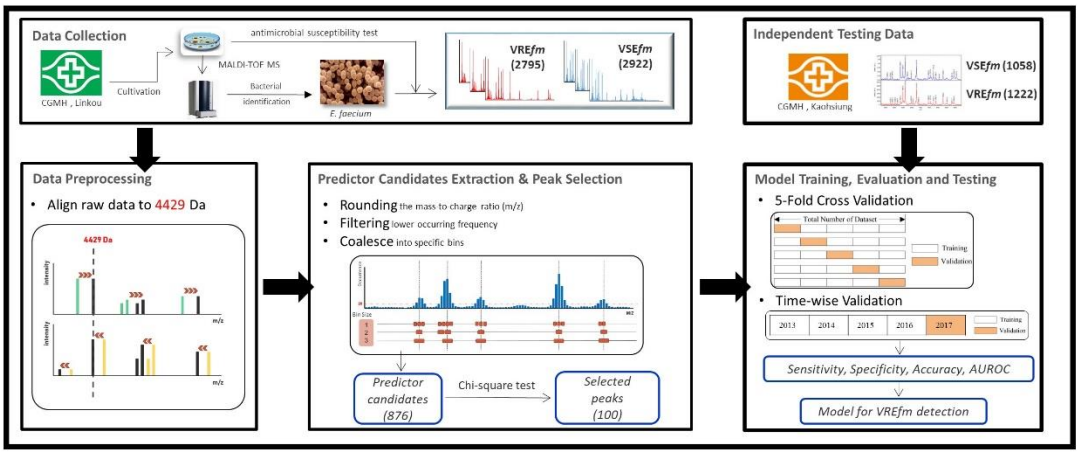


Figure 1(a). Scheme of using the VRE*fm* Model. We plotted a timeline of bacterial culture test in current clinical microbiology laboratory (i.e., traditional approach) and a modified timeline when the VRE*fm* model is incorporated (i.e., machine learning approach). In the traditional approach, specimens are collected for bacterial culture test. One day is usually needed for growth of a single colony for species identification (by MALDI-TOF MS). Antibiotics susceptibility test (AST) of vancomycin for VRE*fm* will cost another day to report. By contrast, in the machine learning approach, the VRE*fm* model can provide preliminary AST at the time when bacterial species is identified by MALDI-TOF MS. For treating VRE*fm*, the machine learning approach can improve accuracy of antibiotics use by around 30% (from 50% accuracy of empirical antibiotics use in the traditional approach to 80% accuracy of preliminary AST provided by the machine learning approach). Meanwhile, the turn-around-time of bacterial culture test can be reduced to one day, which is 50% reduction.

Figure 1(b)

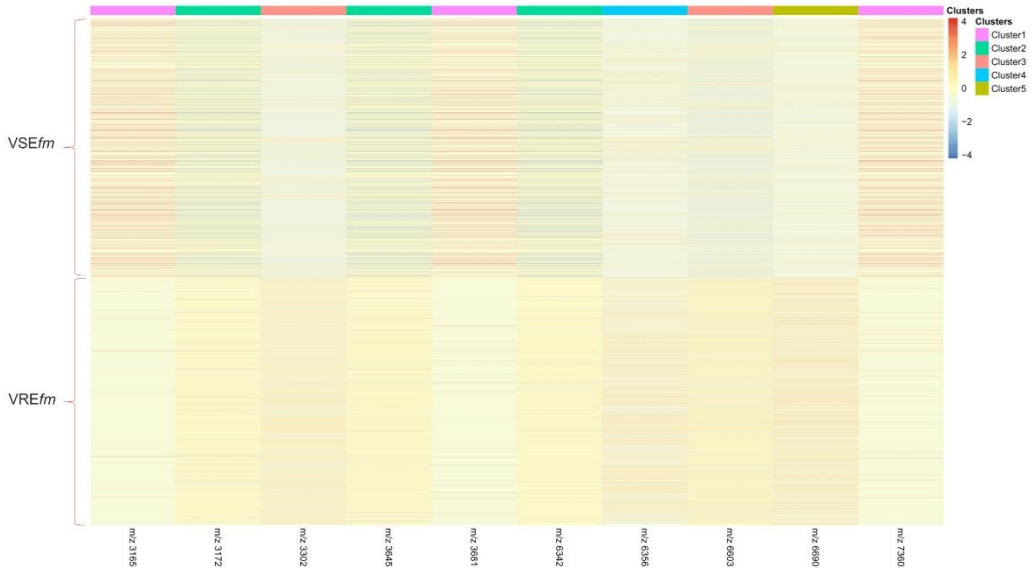


570

571 **Figure 1(b). Schematic Illustration of the Study Design.** We developed and validated a VRE*fm*
572 prediction model. The study included several steps, namely data collection, data
573 preprocessing, predictor candidate extraction and important predictor selection, model
574 training, evaluation, and testing. In data collection, data were obtained from 2 tertiary
575 medical centers (Linkou and Kaohsiung branches of CGMH). The data included mass
576 spectra and results of the vancomycin susceptibility test of *E. faecium*. Data from the CGMH
577 Linkou branch were used for model training and validation, while data from the CGMH
578 Kaohsiung branch served as an independent testing data. In the steps of data preprocessing
579 and predictor candidate extraction and important predictor selection, a specific set of crucial
580 predictors would be used for model training. K-fold, time-wise CV, and external validation
581 were used to confirm the models' robustness. The VRE*fm* prediction model can detect
582 VRE*fm* accurately at least 1 day earlier than the current method.

583

Figure 2



584

585 **Figure 2. Heat map.** We selected top 10 discriminative peaks by chi-square testing the occurrence
586 frequency of peaks in VRE*fm* and VSE*fm*. The heat map was plotted based on the hierarchical
587 clustering of all the VRE*fm* and VSE*fm* isolates from the CGMH Linkou branch. Rows
588 represent the isolates, and columns represent the top 10 discriminative peaks. The values in
589 the heat map represent the MS spectral intensity which was $\log_{10}$-normalized and z-score
590 standardized. Red color indicates relatively higher peak intensity while blue color indicates
591 lower peak intensity. The isolates are grouped into 5 clusters. VRE*fm* and VSE*fm* isolates can
592 be visually differentiated by using the top 10 discriminative peaks.

593

Figure 3(a)

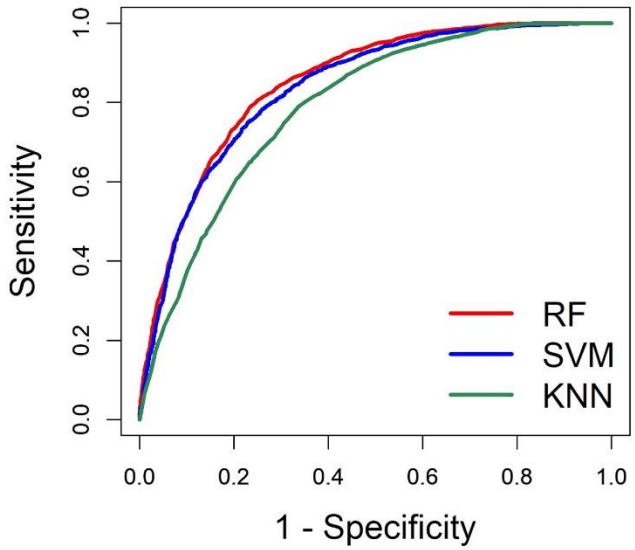


Figure 3(a). ROC Curves for Different Algorithms in Terms of Linkou 5-Fold CV

Figure 3(b)

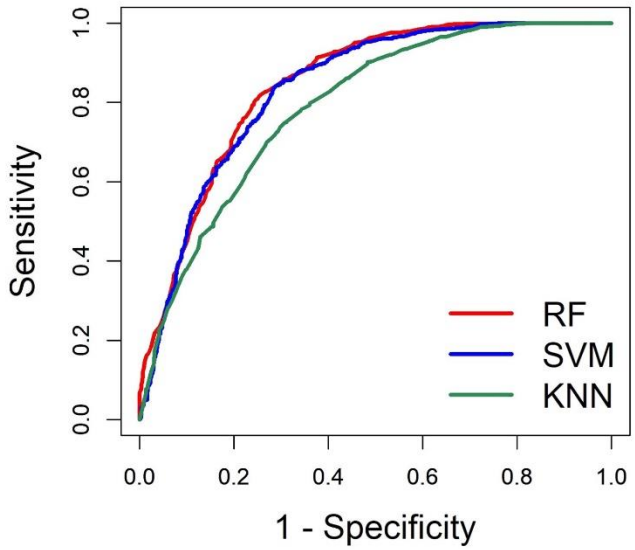


Figure 3(b). ROC Curves for Different Algorithms in Terms of Time-Wise Validation

Figure 3(c)

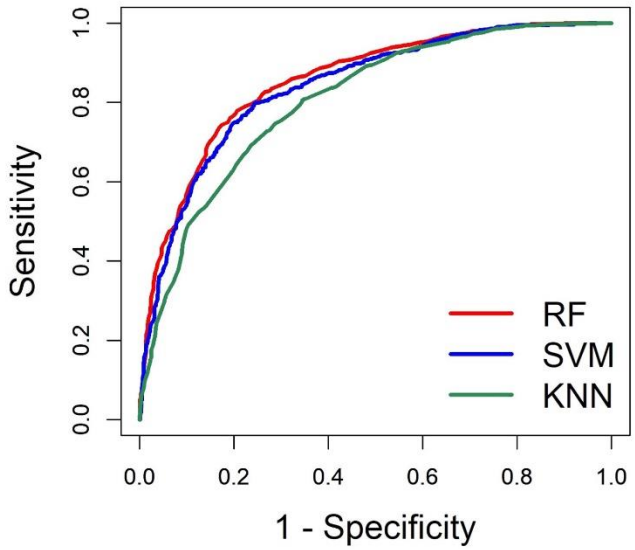
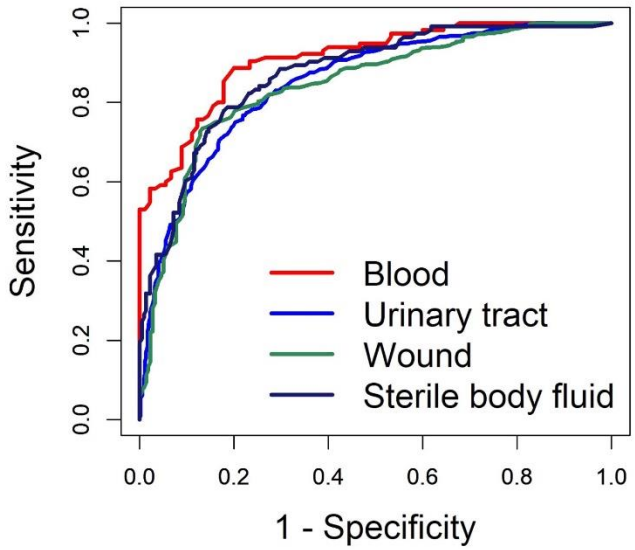


Figure 3(c). ROC Curves for Different Algorithms in Terms of External Validation

Figure 3(d)



603

604 **Figure 3(d). ROC Curves for the RF-Based VRE*fm* Model With Different Types of**

605 **Specimens**

606

607 **Table 1. Performance of VRE*fm* Prediction Models in Terms of k-Fold CV, Time-Wise**

608 **Validation, and External Validation**

AUROC	RF	SVM	KNN
5-fold CV	0.8495 (0.8397, 0.8594)	0.8367 (0.8264, 0.8471)	0.7908 (0.7792, 0.8024)
Time-wise validation	0.8463 (0.8273, 0.8654)	0.8368 (0.8169, 0.8566)	0.7908 (0.7690, 0.8127)
External validation	0.8553 (0.8399, 0.8706)	0.8407 (0.8246, 0.8569)	0.8050 (0.7872, 0.8227)
Accuracy			
5-fold CV	0.7769 (0.7660, 0.7878)	0.7610 (0.7499, 0.7721)	0.7248 (0.7131, 0.7364)
Time-wise validation	0.7840 (0.7640, 0.8039)	0.7815 (0.7615, 0.8016)	0.7228 (0.7011, 0.7445)
External validation	0.7855 (0.7687, 0.8024)	0.7781 (0.7610, 0.7951)	0.7355 (0.7174, 0.7536)
Sensitivity			
5-fold CV	0.8054 (0.7951, 0.8517)	0.7826 (0.7719, 0.7934)	0.7873 (0.7767, 0.7980)
Time-wise validation	0.8153 (0.7965, 0.8341)	0.8415 (0.8238, 0.8592)	0.7491 (0.7281, 0.7702)
External validation	0.7791 (0.7620, 0.7961)	0.7954 (0.7789, 0.8120)	0.8044 (0.7881, 0.8207)
Specificity			
5-fold CV	0.7497 (0.7384, 0.7609)	0.7403 (0.7289, 0.7517)	0.6649 (0.6526, 0.6772)
Time-wise validation	0.7477 (0.7266, 0.7688)	0.7120 (0.6900, 0.7340)	0.6922 (0.6698, 0.7146)
External validation	0.7930 (0.7764, 0.8096)	0.7580 (0.7405, 0.7756)	0.6560 (0.6365, 0.6755)

609

610 AUROC, area under the receiver operating characteristic curve.

611

612 **Table 2. Performance of the RF-Based VRE*fm* Detection Model With Different Types of**

613 **Specimens in Terms of External Validation**

	Type			
Metrics	Blood	Urinary tract	Sterile body fluid	Wound
AUROC	0.9103 (0.8727, 0.9480)	0.8494 (0.8258, 0.8731)	0.8714 (0.8321, 0.9106)	0.8432 (0.8121, 0.8743)
Accuracy	0.8488 (0.7997, 0.8978)	0.7743 (0.7482, 0.8004)	0.8077 (0.7657, 0.8497)	0.7740 (0.7436, 0.8043)
Sensitivity	0.8870 (0.8436, 0.9303)	0.7672 (0.7409, 0.7936)	0.7788 (0.7345, 0.8230)	0.7339 (0.7018, 0.7659)
Specificity	0.8000 (0.7452, 0.8548)	0.7805 (0.7547, 0.8063)	0.8222 (0.7815, 0.8630)	0.8676 (0.8430, 0.8922)

614

615