Original Article

TITLE: Development and Validation of a MALDI-TOF-Based Model to Predict Extended-Spectrum Beta-Lactamase and/or Carbapenemase-Producing in *Klebsiella pneumoniae* Clinical Isolates

Alejandro Guerrero-López (1)(2) *, Ana Candela (2), Carlos Sevilla-Salcedo (1), Marta Hernández-García (3)(4), Pablo M. Olmos (1)(2), Emilia Cercenado (2)(5)(6)(7), Rafael Cantón (3)(4), Patricia Muñoz (2)(5)(6)(7), Rosa del Campo (3)(4), Vanessa Gómez-Verdejo (1), Belén Rodríguez-Sánchez (2)(5).

(1) Department of Signal Theory and Communication, University Carlos III of Madrid, Madrid, Spain.
(2) Gregorio Marañón Health Research Institute, Hospital Gregorio Marañón, Madrid, Spain.
(3) Department of Microbiology, Hospital Ramón y Cajal and Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS). Madrid, Spain.
(4) CIBER en Enfermedades Infecciosas. Madrid, Spain.
(5) Clinical Microbiology and Infectious Diseases Department, Hospital General Universitario Gregorio Marañón, Madrid, Spain
(6) CIBER de Enfermedades Respiratorias (CIBERES CB06/06/0058), Madrid, Spain
(7) Medicine Department, Faculty of Medicine, Universidad Complutense de Madrid, Madrid, Spain

***Corresponding author**

Alejandro Guerrero-López

Theory Signal Group, Dpt. of Signal Theory and Communications

Universidad Carlos III de Madrid, Office 4.3.A03

Av. de la Universidad, 30, Leganés, 28911, Madrid, Spain,

Email: alexjorguer@tsc.uc3m.es

Phone: +34 916 248 839

FILE STATEMENT:

- **ABSTRACT LENGTH:** 238 (250 max)
- **PAPER LENGTH:** 2468 (2500 max)

ABSTRACT

Objectives: MALDI-TOF Mass Spectrometry (MS) is a reference method for microbial identification at clinical microbiology laboratories. We have designed and validated a new multiview model based on machine learning from MS spectra to predict antibiotic resistance mechanisms 24 h before phenotypic results are available.

Methods: Antibiotic susceptibility of 402 clinical *Klebsiella pneumoniae* isolates was determined in two collections, discriminating among Wild Type (WT), Extended-Spectrum Beta-Lactamases (ESBL) producers, and ESBL and Carbapenemases (ESBL+CP) producers. Each isolate was subcultured 3 consecutive days and 2 independent spectra were acquired in each replica (6 MS spectra/isolate). Spectra were automatically classified by a kernelized Bayesian factor analysis model (KSSHIBA), using two independent strategies: 1) the model was designed with isolates from a single centre and validated with isolates from the other centre; and 2) in a second stage all isolates were used at the same time for design and validation processes.

Results: Higher prediction values were obtained when integrating all isolates with hospital collection of origin information. Our model exhibited higher prediction capability than current state-of-the-art models, particularly in intercollection scenarios because local epidemiology could introduce relevant variables affecting prediction accuracy.

Conclusions: Compared to previously reported studies, our model demonstrated the highest ability to predict ESBL and/or CP production in clinical *K. pneumoniae* isolates and it provided an efficient way to combine information from different centres. Its implementation in microbiological laboratories could improve the detection of multi-drug resistant isolates, optimizing the therapeutic decision.

INTRODUCTION

Multidrug-resistant *Klebsiella pneumoniae* is considered a global public health threat according to the major international health organizations due to its rapid spread, its high morbidity and mortality and the economic burden associated with its treatment and control [1–3]. Resistance to carbapenems is a major challenge since this antibiotic group represents one of the last therapeutic options. In fact, some Carbapenemases (CP) have been shown to hydrolyse almost all beta-lactam antibiotics [4]. Thus, besides the routinely antimicrobial susceptibility testing (AST), rapid diagnostic methods such as MALDI-TOF Mass Spectrometry (MS) should be implemented in clinical microbiology laboratories beyond identification for early detection of multidrug resistant isolates.

MALDI-TOF MS is designed for microbial identification, but also allows the detection of extended-spectrum beta-lactamases (ESBL) and CP by the different molecular weight of the antibiotic after its hydrolysis by resistant bacteria [5]. This approach is faster than conventional AST (30-60 min vs. 18-24 h) but requires highly trained personnel and it is of limited use in clinical laboratories.

More recently, machine-learning methods such as Support Vector Machines (SVM), Random Forest (RF), K-Nearest Neighbours (KNN), naïve Bayes and Logistic Regression have been successfully applied to predict CP-producing isolates from MS spectra [6]; as well as other approaches based on deep learning methods [7]. Supervised learning is a powerful classification tool but is not yet optimized for the usual high-dimensional MS data. Consequently, pre-processing is needed to reduce dimensions; in this sense, some authors have proposed the use of a genetic algorithm in combination with a SVM using ClinProTools [8]. Bayesian models are starting to be implemented, as they get rid of cross-validation issues while can provide a probability prediction with a confidence measurement. A recent study has proposed the use of an ad-*hoc* non-linear kernel followed by a Gaussian Process [9].

To address all these aspects, we developed and validated a novel Bayesian model called Kernelized Sparse Semi-Supervised Interbattery Bayesian Analysis (KSSHIBA) [10,11] that using the MS spectra and the hospital collection of origin predicts the phenotypic/genotypic AST. As phenotypic AST data reproducibility between laboratories is also an unresolved issue, we have included clinical collections characterized at separate centres, representing wide lineages variability from distant epidemiological environments.

MATERIAL AND METHODS

Isolates selection and processing. We included 282 consecutive clinical *K. pneumoniae* isolates (2014-2019) isolated at Hospital Gregorio Marañón (GM), and 120 isolates previously characterized in surveillance programs (STEP and SUPERIOR) [12,13] and sourcing from 8 Spanish and 11 Portuguese hospitals. This collection of isolates was merged into the Hospital Ramón y Cajal collection (RyC). AST determination was performed for each collection in their origin centre by the automated broth microdilution method Microscan® System (Beckman-Coulter, CA, USA), using EUCAST criteria (2021). Presence of ESBL/CP genetic resistant mechanisms was corroborated by molecular testing. Isolates were categorized as Wild Type (WT) -n=94-, ESBL-producers (n=67) or ESBL+CP-producers (n=241).

Isolates were kept frozen at -80ºC in skimmed milk and, after thawing, cultured overnight at 37ºC in Columbia Blood agar (bioMérieux, Lyon, France) during 3 subcultures for metabolic activation. MS analysis was centralized and performed by the same operator using an MBT Smart MALDI Biotyper mass spectrometer (Bruker Daltonics, Bremen), in 6 separated replicas (2 positions in 3 consecutive days). Protein extraction was performed adding 1µl 100% formic acid further dry at room temperature. Then, 1µl of HCCA matrix solution (Bruker Daltonics) was added to each spot. MS spectra were acquired in the positive linear mode in the range of 2,000 to 20,000 Da, using default settings [14], although only data between 2,000-12,000 *m/z* was further analysed applying Total Ion Current (TIC) normalization [15,16].

Model development. The proposed model, SSHIBA [10], considers a common low dimensional latent variable, Z, responsible for generating heterogeneous observations of each view (i.e., it can model either continuous, categorical or multilabel observations) and, besides, it automatically adjusts the dimension of this latent space, finding the relationships between the different views and making an interpretability analysis easier. To deal with the high dimensionality of the MS spectra data, we applied here the KSSHIBA [11] extension since it is able to model kernelized version data to work in the dual space by means of the kernel trick and avoiding the high dimensionality problem of the MS spectra. In our setting, KSSHIBA model used three complementary views: the Kernelized MS spectra, the hospital collection origin (GM or RyC), and the prediction of Antibiotic Resistance (AR) category (WT, ESBL and ESBL+CP) (Figure 1).

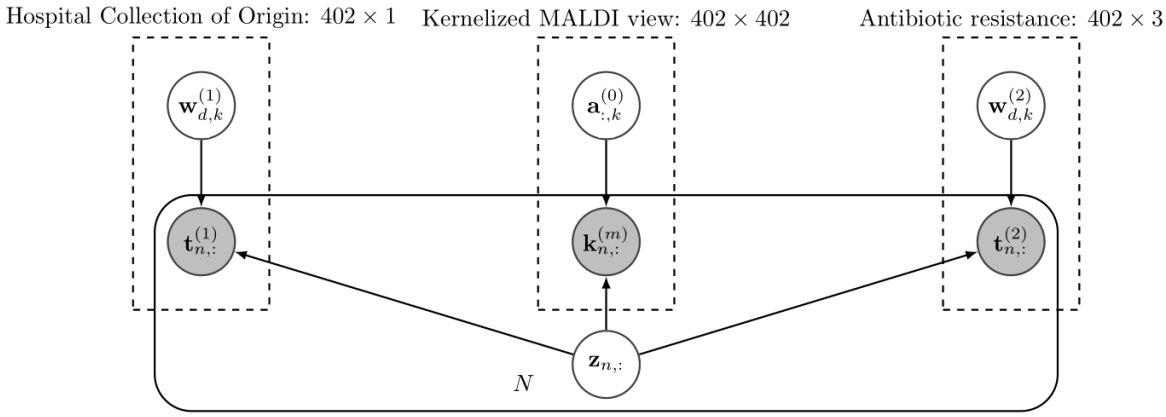

Figure 1. KSSHIBA graphical model proposed integrating 3 views: kernelized MS spectra, Hospital Collection of Origin (GM or RyC), and antibiotic resistance category (WT, ESBL, and ESBL+CP).

Model validation. To test whether intra and intercollection distributions could improve the learning process, two different scenarios were proposed. First, GM and RyC collections were tested separately (intra-collection analysis), and in a second stage all isolates were merged in a single collection (inter-collection analysis).

In the first experiment, we split each dataset into 5 random train-validation folds. Each training fold was processed to correct the unbalance in the class population by oversampling the minority class on each antimicrobial resistance category ultimately resulting in stratified folds with a consistent class ratio. In the second experiment using the global collection, data were again split into 5 random train-test folds maintaining the previous unbalance correction technique. Moreover, we defined two frameworks: (1) directly combining both datasets and (2) merging them with an extra view identifying as “0” isolates from GM and as “1” those from RyC.

Comparison with state-of-the-art methods. Firstly, KSSHIBA was compared with a SVM and a GP since both approaches can also work in a dual space. For these models we tried out a nonlinear kernel called Radial Basis Kernel (RBF) and a linear kernel. As we were solving a multidimensional problem with MS spectra and both models work for single output prediction, we ran independent SVMs and GPs for each prediction task. We also compared KSSHIBA to a RF able to jointly estimate all the prediction tasks. Finally, we explored our model implementing the kernel called Peak Information KErnel (PIKE) [9], which exploits non-linear correlations between MS peaks. In this case, MS spectra were pre-processed by a topological peak selection keeping only 200 peaks per sample, as they explain in their work.

Hyper-parameter cross-validation was done by an inner 5-fold over the training folds. We cross-validated the C value (0.01, 0.1, 1, 10) for the SVM and the number of estimators (50, 100, 150) and the maximum number of features (auto, log2) for the RF. For both KSSHIBA and GP, the hyper-parameters were optimized by maximizing the evidence lower bound and the log marginal likelihood of the data, respectively. Then, we ran both models 5 times for each one to ensure that the learnt parameters do not correspond to a local maximum. When KSSHIBA was combined with the PIKE kernel we fixed the kernel smoothing parameter "t" to 5, based on the influence analysis carried out in the original research.

The extended comparison to other baselines, such as KNN, RF or SVM and GP with other kernels are included in the Table S2.

Performance metric. Prediction of AR category was calculated by the AUC measuring. The Receiver Operator Characteristic curve is an evaluation metric used in binary classification problems that show the True Positive Rate against the False Positive Rate at different thresholds. The AUC measures the ability of the model to distinguish between classes (positive and negative) for different thresholds in the probability prediction and is a summary of the ROC curve. Higher values of AUC means that the models distinguish better between WT and non-WT isolates.

Repository. The model is implemented in Python using Pytorch and Pyro libraries. The code used to obtain the presented results, an explanation on how to use KSSHIBA and both datasets are publicly available in [17].

Ethics Statement. The Ethics Committee from the GM and RyC hospitals (codes MICRO.HGUGM.2020-002, and 087–16, respectively) approved this study. The study was performed from microbiological samples, not human products and informed consent from the patients was not necessary.

RESULTS

A detailed description of the datasets decomposition in each scenario can be found in Supplementary Material A. The experiment's source code can be found at the GitHub repository [17].

Intra-collection scenario

Results obtained for GM (n=282) and RyC (n=120) separated collections are summarized in Table 1. For GM isolates, KSSHIBA performs better AUC scores than the baselines regardless of the predicted AR category. Specifically, nonlinear kernels provided the best results, where the RBF kernel performed better in the prediction of both ESBL and ESBL+CP and the PIKE kernel worked better for WT prediction. In the RyC collection KSSHIBA was more accurate for WT prediction while performing competitive results in ESBL prediction.

Table 1. AUC mean and standard deviation for GM and RyC intracollection analysis. The kernel type is detailed in brackets, if used, and values in bold correspond to the high prediction value for each antibiotic resistance category.

Dataset	Category	KSSHIBA (RBF)	KSSHIBA (LINEAR)	KSSHIBA (PIKE)	GP (LINEAR)	SVM (RBF)	RF
GM	WT	0.61±0.14	0.70±0.15	**0.71±0.16**	0.70±0.18	0.67±0.12	0.70±0.17
	ESBL	**0.57±0.28**	0.46±0.19	0.56±0.32	0.54±0.18	0.40±0.29	0.39±0.21
	ESBL+CP	**0.85±0.14**	0.77±0.16	0.78±0.09	0.80±0.20	0.82±0.19	0.80±0.19
RyC	WT	0.47±0.35	0.49±0.22	**0.64±0.19**	0.48±0.28	0.45±0.15	0.57±0.26
	ESBL	0.70±0.10	0.59±0.08	0.43±0.09	0.58±0.14	**0.72±0.14**	0.69±0.10
	ESBL+CP	0.67±0.12	0.66±0.05	0.43±0.09	0.62±0.06	0.71±0.17	**0.71±0.07**

205 **Inter-collection scenario**

206 Table 2 shows the results obtained when training simultaneously GM and RyC
207 collections (n=402). Labelled KSSHIBA, which means that each sample is labelled indicating
208 from which collection is coming from, outperforms every baseline for GM isolates while also
209 performs better for the prediction of WT and ESBL+CP isolates in RyC. The lower
210 performance of the baselines without the source label indicates that using the data from both
211 datasets without identifying their origin produced biased results. However, merging both
212 collections, by adding an extra label with the collection origin, clearly contributed to improve
213 the results in terms of AUC in all AR categories apart from ESBL, where there is a large
214 imbalance (see Table S1).

215

216 **Table 2**. AUC mean and standard deviation when all isolates were merged in a single
217 collection. Each model is defined with its name and the type of kernel in brackets, if used.
218 Values in bold correspond to the high prediction for each antibiotic resistance category.
219 Labelled means that the samples have the hospital collection origin view, whereas unlabelled
220 means that this information is not considered by model.

Dataset	Category	KSSHIBA (LINEAR) LABELED	KSSHIBA (LINEAR) UNLABELED	GP (LINEAR) UNLABELED	SVM (RBF) UNLABELED
GM	WT	**0.77±0.11**	0.72±0.14	0.76±0.10	0.62±0.13
	ESBL	**0.46±0.19**	0.39±0.21	0.43±0.20	0.39±0.21
	ESBL+CP	**0.88±0.08**	0.86±0.10	0.86±0.08	0.85±0.08
RyC	WT	**0.70±0.16**	0.66±0.16	0.68±0.17	0.59±0.20
	ESBL	0.55±0.09	0.49±0.09	0.60±0.10	**0.69±0.12**
	ESBL+CP	**0.68±0.10**	0.64±0.06	0.64±0.04	0.66±0.14

221

222

223

224

225

Latent space analysis of our model

The learnt weight matrix W of each view to evaluate the importance of each latent variable and analyse how they relate to each other is shown in Figure 2; in MS view these weights are averaged over the input MS dimensions to obtain an intensity value per latent factor. Here, due to the sparsity imposed, the model automatically learned which latent features were relevant for each view and, in turn, some of the views only used selected features.

KSSHIBA projected the 3 input views into 76 latent features (Figure 2), which were ordered by importance in the prediction task. Noticeably, only 13 latent factors were used to predict the AR category, all of them shared information with MS while only 3 of them correlated simultaneously all the information available. Finally, note 51 latent features private to MS spectra view, which corresponds to an unsupervised projection of the data that can be understood as a principal component analysis.

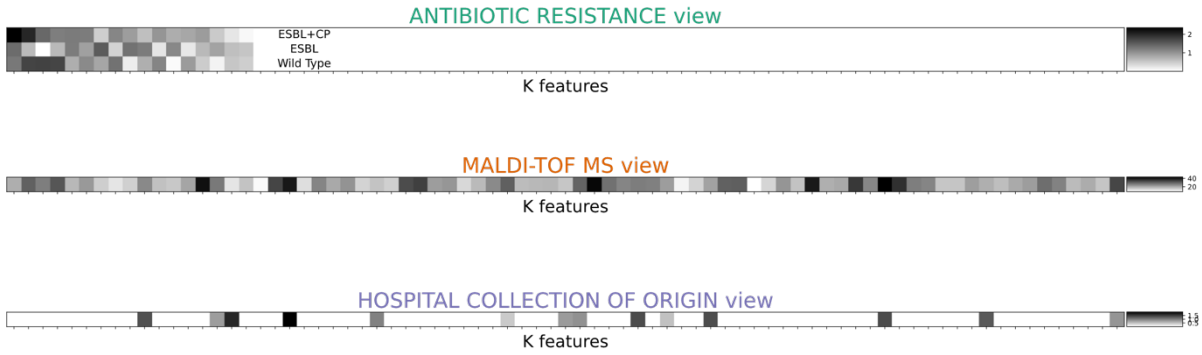

Figure 2. KSSHIBA latent space projection (d=76) for hospital collection origin, MS spectra, 244 and antibiotic resistance category, including data with the extra hospital collection of origin 245 view. Each subfigure refers to a W matrix associated with each view.

KSSHIBA showed that there existed a correlation between the hospital collection of origin of each strain and their AST profile shown in the latent space that these views share. This latent space representation allowed KSSHIBA to outperform all baselines in the intercollection scenario. As seen in Table 2, KSSHIBA performed 0.88+-0.08, 0.46+-0.19 and 0.77+-0.11 in GM for ESBL+CP, ESBL and WT, respectively, outperforming the state-of-the-art models. Likewise, in RyC data our proposal performed better in both ESBL+CP and WT prediction with AUC values of 0.68+-0.10 and 0.70+-0.16, respectively.

251 **DISCUSSION**

252 MS allows rapid and accurate bacterial identification, and the resulting spectra can be
253 analysed by machine learning approaches to predict AR, as it has been previously suggested
254 [18]. The most relevant limitation for this methodology is the great complexity of MS spectra,
255 which is also influenced by the particularities of each lineage and its accessory genome. In this
256 sense, two isolates carrying the same CP gene could differ in the MS spectra due to their
257 particular genetic background. For the present work, two different bacterial sets were included,
258 one of them sourcing from the same hospital collection without any particular criteria of
259 inclusion and the other one grouped strain from 18 geographical disperse hospitals selected by
260 their phenotypic and genotypic beta-lactams resistance. The latter collection was characterized
261 by whole-genome sequencing and includes both frequent and rare clonal lineages.

262 ESBL and CP categories groups a highly variable set of different proteins but with a
263 common phenotypic pattern of antibiotic susceptibility. Here, we developed and validated a
264 novel model in *K. pneumoniae* clinical isolates for ESBL and CP-producing prediction based
265 on MS spectra. Our major contribution is provided by the multiview nature of KSSHIBA since
266 it was able to learn from intercollection distribution without getting biased by intracollection
267 distributions using the hospital collection of origin as an extra view. Moreover, our model also
268 reduced the training complexity by kernel application and getting rid of cross-validation issues
269 by the optimization of the evidence lower bound exploiting the Bayesian framework. As a
270 direct consequence of these actions, the training period was considerably reduced.

271 When we used a complex (non-linear) kernel, our model got better results in
272 intracollection data: 0.85+-0.14 in ESBL+CP, 0.57+-0.28 in ESBL and 0.71+-0.16 in WT in
273 GM and 0.64+-0.19 in WT in RyC (Table 1). Although the results seemed to indicate that the
274 solution was to use complex kernels, when we combined both collections (intercollection
275 scenario, Table 2), complex kernels obtained poorer results, pointing to a possible over-fit to
276 the intracollection distributions in the first scenario. For the intracollection scenario, KSSHIBA
277 (UNLABELLED) failed to achieve better results than the baselines, but this result could be a
278 consequence of the overrepresentation of ESBL/CP in the RyC collection with respect to the
279 GM collection without inclusion criteria. This imbalance was particularly visible in the AUC
280 performance as all unlabelled models predicted ESBL significantly worse in the GM dataset
281 (unbalanced) than in the RyC dataset (balanced). Likewise, WT isolates were significantly
282 more balanced only in the GM dataset. Therefore, the unlabelled models predicted worse
283 unbalanced scenarios, getting biased by the data distribution. However, KSSHIBA

(LABELLED) proved that exploiting the multi-view heterogeneous features allowed to add additional information to the learning process, such as the dataset source, being able to properly model different data distributions getting rid of the introduced bias by the data itself. Therefore, our model represents a step forward in the prediction of AR, particularly to beta-lactam antibiotics, as it obtains better performance in terms of AUC than previous models, while providing new features: adding more data to the learning base, reducing dimensionality and providing interpretability of how the data sets interact with each other to predict.

Previous reports suggested some protein peaks are associated with specific mechanisms of AR [19]. These observations were manually performed by direct visualization of the protein spectra, but obviously automation avoids operator-related bias and provides more information about the optimal areas of the spectrum for discrimination. Likewise, some person-related discrepancies may occur in AST for the WT/resistant categories. Although the same AST methodology was used in both collections, we cannot rule-out possible discrepancies linked to each centre/person. On the contrary, a single worker in the same instrument performed all MS spectra. A limitation of our work was a reduced number of isolates for machine-learning methodology. Also, more geographically unrelated isolates should be included, always combining the phenotypic and the genotypic previous characterization of the antimicrobial susceptibility profile. Although machine-learning applications on MS spectra to predict resistance to antibiotics are still in an initial stage, their great potential should encourage us to continue work in this direction.

TRANSPARENCY DECLARATION

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The work of Pablo M. Olmos is supported by Spanish MINECO (Agencia Estatal de Investigación) grants TEC2017-92552-EXP and RTI2018-099655-B-100; the Comunidad de Madrid under grants IND2017/TIC-7618, IND2018/TIC-9649, IND2020/TIC-17372, and Y2018/TCS-4705; the BBVA Foundation under the Domain Alignment and Data Wrangling with Deep Generative Models (Deep-DARWiN) project; and the European Union (European Regional Development Fund and the European Research Council) through the European Union's Horizon 2020 Research and Innovation Program under grant 714161. C. Sevilla-Salcedo, A. Guerrero-López and V. Gómez-Verdejo's are partly funded by the Spanish MINECO (Agencia Estatal de Investigación) grants TEC2017-83838-R and PID2020-115363RB-I00. A. Guerrero-López was also funded by the Intramural Program of the Gregorio Marañón Health Research Institute. This work was supported by the projects PI15/01073 and PI18/00997 from the Health Research Fund (Instituto de Salud Carlos III. Plan Nacional de I+D+I 2013-2016) of the Carlos III Health Institute (ISCIII, Madrid, Spain) partially financed by the European Regional Development Fund (FEDER) 'A way of making Europe'. Belén Rodríguez-Sánchez is a recipient of a Miguel Servet contract (CPII19/00002) supported by the Health Research Fund.

Acknowledgments

We acknowledge the SUPERIOR Study Group which provided the RyC collection data and includes the following members: Antonio Oliver and Xavier Mulet (Hospital Universitario Son Espases, Palma de Mallorca, Spain); Emilia Cercenado (Hospital General Universitario Gregorio Marañón, Madrid, Spain); Germán Bou and M. Carmen Fernández (Hospital Universitario A Coruña, A Coruña, Spain); Álvaro Pascual and Mercedes Delgado-Valverde (Hospital Universitario Virgen Macarena, Sevilla, Spain); Concepción Gimeno and suppNuria Tormo (Consorcio Hospital General Universitario de Valencia, Valencia, Spain); Jorge Calvo, Jesús Rodríguez-Lozano and Ana Ávila Alonso (Hospital Universitario Marqués de Valdecilla, Santander, Spain); Jordi Vila, Francesc Marco and Cristina Pitart (Hospital Clínic, Barcelona, Spain); and María García del Castillo, Sergio García-Fernández, Marta Hernández-García, Marta Tato and Rafael Cantón (Hospital Universitario Ramón y Cajal, Madrid, Spain). This study was sponsored by MSD Spain. The STEP Study Group includes the following members: José Melo-Cristino (Serviço de Microbiologia Centro Hospitalar Lisboa Norte, Lisboa, Portugal); Margarida F. Pinto, Cristina Marcelo, Helena Peres, Isabel Lourenço, Isabel Peres, João Marques, Odete Chantre and Teresa Pina (Laboratório de Microbiologia, Serviço de Patologia Clínica, Centro Hospitalar Universitário Lisboa Central, Lisboa, Portugal); Elsa Gonçalves and Cristina Toscano (Laboratório de Microbiologia Clínica Centro Hospitalar de Lisboa Ocidental, Lisboa, Portugal); Valquíria Alves (Serviço de Microbiologia, Unidade

Local de Saúde de Matosinhos, Matosinhos, Portugal); Manuela Ribeiro, Eliana Costa and Ana Raquel Vieira (Serviço Patologia Clínica, Centro Hospitalar Universitário São João, Porto, Portugal); Sónia Ferreira, Raquel Diaz and Elmano Ramalheira (Serviço Patologia Clínica, Hospital Infante Dom Pedro, Aveiro, Portugal); Sandra Schäfer, Luísa Tancredo and Luísa Sancho (Serviço de Patologia Clínica, Fernando Fonseca, Amadora, Portugal); Ana Rodrigues and José Diogo (Serviço de Microbiologia, Hospital Garcia de Orta, Almada, Portugal); Rui Ferreira (Serviço de Patologia Clínica—Microbiologia—CHUA—Unidade de Portimão, Portugal); Helena Ramos, Tânia Silva and Daniela Silva (Serviço de Microbiologia, Centro Hospitalar Universitário do Porto, Porto, Portugal); Catarina Chaves, Carolina Queiroz and Altair Nabiev (Serviço de Microbiologia, Centro Hospitalar Universitário de Coimbra, Coimbra, Portugal); Leonor Pássaro, Laura Paixao, João Romano and Carolina Moura (MSD Portugal, Paço de Arcos, Portugal).

Access to the data

The data and the code used in this work are publicly available at the GitHub repository: https://github.com/alexjorguer/RMPrediction

Authors contributions

A.G.L. developed the model, validated the model, and wrote the original draft. A.C. acquired and analysed the data. C.S.S., P.M.O. and V.G.V. supervised model development. M.H.G. and E.C. supervised AST data. M.H.G, R.D.C and B.R.S. supervised model validation. P.M.O., P.M., V.G.V. and B.R.S. acquired funding for the development of this study. P.M.O., R.C., V.G.V. and B.R.S. conceived the study design, did the formal analysis, and supervised. A.G.L., A.C., C.S.S, M.H.G., P.M.O., R.D.C., P.M., R.C., V.G.V. and B.R.S. reviewed and approved the final manuscript.

REFERENCES

[1] Edward R. Carbapenem-resistant Enterobacteriaceae - Second update 2019:17.

[2] Carbapenem-resistant Enterobacteriaceae (CRE) n.d.:2.

[3] Tacconelli E. GLOBAL PRIORITY LIST OF ANTIBIOTIC-RESISTANT BACTERIA TO GUIDE RESEARCH, DISCOVERY, AND DEVELOPMENT OF NEW ANTIBIOTICS n.d.:7.

[4] Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect Dis 2018;18:318–27. https://doi.org/10.1016/S1473-3099(17)30753-3.

[5] Oviaño M, Bou G. Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry for the Rapid Detection of Antimicrobial Resistance Mechanisms and Beyond. Clin Microbiol Rev 2018;32:e00037-18. https://doi.org/10.1128/CMR.00037-18.

[6] Huang T-S, Lee SS-J, Lee C-C, Chang F-C. Detection of carbapenem-resistant Klebsiella pneumoniae on the basis of matrix-assisted laser desorption ionization time-of-flight mass spectrometry by using supervised machine learning approach. PLoS ONE 2020;15:e0228459. https://doi.org/10.1371/journal.pone.0228459.

[7] Esener N, Green MJ, Emes RD, Jowett B, Davies PL, Bradley AJ, et al. Discrimination

of contagious and environmental strains of Streptococcus uberis in dairy herds by means of mass spectrometry and machine-learning. Sci Rep 2018;8:17517. https://doi.org/10.1038/s41598-018-35867-6.

[8] Chen JHK, She KKK, Wong O-Y, Teng JLL, Yam W-C, Lau SKP, et al. Use of MALDI Biotyper plus ClinProTools mass spectra analysis for correct identification of Streptococcus pneumoniae and Streptococcus mitis/oralis. J Clin Pathol 2015;68:652–6. https://doi.org/10.1136/jclinpath-2014-202818.

[9] Weis C, Horn M, Rieck B, Cuénod A, Egli A, Borgwardt K. Topological and kernel-based microbial phenotype prediction from MALDI-TOF mass spectra. Bioinformatics 2020;36:i30–8. https://doi.org/10.1093/bioinformatics/btaa429.

[10] Sevilla-Salcedo C, G'omez-Verdejo V, Olmos P. Sparse Semi-supervised Heterogeneous Interbattery Bayesian Analysis. Pattern Recognit 2021. https://doi.org/10.1016/j.patcog.2021.108141.

[11] Sevilla-Salcedo C, Guerrero-López A, Olmos PM, Gómez-Verdejo V. Bayesian Sparse Factor Analysis with Kernelized Observations. ArXiv200600968 Cs Stat 2021.

[12] Hernández-García M, García-Fernández S, García-Castillo M, Melo-Cristino J, Pinto MF, Gonçalves E, et al. Confronting Ceftolozane-Tazobactam Susceptibility in Multidrug-Resistant Enterobacterales Isolates and Whole-Genome Sequencing Results (STEP Study). Int J Antimicrob Agents 2021;57:106259. https://doi.org/10.1016/j.ijantimicag.2020.106259.

[13] Hernández-García M, García-Fernández S, García-Castillo M, Bou G, Cercenado E, Delgado-Valverde M, et al. WGS characterization of MDR Enterobacterales with different ceftolozane/tazobactam susceptibility profiles during the SUPERIOR surveillance study in Spain. JAC-Antimicrob Resist 2020;2:dlaa084. https://doi.org/10.1093/jacamr/dlaa084.

[14] Rodríguez-Sánchez B, Marín M, Sánchez-Carrillo C, Cercenado E, Ruiz A, Rodríguez-Créixems M, et al. Improvement of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry identification of difficult-to-identify bacteria and its impact in the workflow of a clinical microbiology laboratory. Diagn Microbiol Infect Dis 2014;79:1–6. https://doi.org/10.1016/j.diagmicrobio.2014.01.021.

[15] Zvezdanova ME, Arroyo MJ, Méndez G, Candela A, Mancera L, Rodríguez JG, et al. Detection of azole resistance in Aspergillus fumigatus complex isolates using MALDI-TOF mass spectrometry. Clin Microbiol Infect 2021. https://doi.org/10.1016/j.cmi.2021.06.005.

[16] Rodrigues C, Passet V, Rakotondrasoa A, Brisse S. Identification of Klebsiella pneumoniae, Klebsiella quasipneumoniae, Klebsiella variicola and Related Phylogroups by MALDI-TOF Mass Spectrometry. Front Microbiol 2018;9:3000. https://doi.org/10.3389/fmicb.2018.03000.

[17] Guerrero-López A. Klebsiellas Pneumoniae Resistant Mechanisms and Antimicrobial Resistance Prediction library. 2021.

[18] Lau AF, Wang H, Weingarten RA, Drake SK, Suffredini AF, Garfield MK, et al. A Rapid Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry-Based Method for Single-Plasmid Tracking in an Outbreak of Carbapenem-Resistant Enterobacteriaceae. J Clin Microbiol 2014;52:2804–12. https://doi.org/10.1128/JCM.00694-14.

[19] Correa-Martínez CL, Idelevich EA, Sparbier K, Kostrzewa M, Becker K. Rapid Detection of Extended-Spectrum β-Lactamases (ESBL) and AmpC β-Lactamases in Enterobacterales: Development of a Screening Panel Using the MALDI-TOF MS-Based Direct-on-Target Microdroplet Growth Assay. Front Microbiol 2019;10:13. https://doi.org/10.3389/fmicb.2019.00013.