CBIS'24
XX Congresso Brasileiro de Informática em Saúde
08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil

Aprendizado de máquinas para predição de resistência microbiana

Machine learning algorithms for the prediction of bacterial resistance

Aprendizaje automático para la predicción de la resistencia microbiana

Dra. Sofia Clara Lage Rosa [1], Dra. Nathália Irffi Carvalho[1], Dra. Helena Duani[2]

1 Hospital Felício Rocho, Departamento de Infectologia, Belo Horizonte (MG), Brasil.
2 Universidade Federal de Minas Gerais, Departamento de Clínica Médica, Belo Horizonte (MG), Brasil.

Autor correspondente: Dra. Sofia Clara Lage Rosa
E-mail: sofia.lage@hotmail.com

Resumo

A resistência a antibióticos representa uma preocupação significativa para a saúde global, particularmente em unidades de terapia intensiva (UTIs), onde o diagnóstico rápido é essencial. Objetivo do estudo: Testar algoritmos de aprendizado de máquina para prever a resistência bacteriana em UTIs; Métodos: Os fatores idade, gênero, tipo de amostra, antibiótico testado e coloração de Gram das bactérias foram retirados do banco de dados MIMIC-III e usados para treinamento de seis modelos de aprendizado de máquinas. Resultados: O Extreme Gradient Boosting demonstrou a maior precisão de previsão, com 84,53%. Conclusão: o aprendizado de máquina poderia oferecer uma solução para a detecção precoce da resistência a antibióticos, melhorando assim o cuidado do paciente e o manejo dos antibióticos.

Descritores: Resistência Microbiana; Aprendizado de Máquina; Unidade de Terapia Intensiva

Abstract

Antibiotic resistance represents a significant concern for global health, particularly in intensive care units (ICUs), where rapid diagnosis is essential. Study objective: To

test Machine Learning algorithms for predicting bacterial resistance in ICUs; Methods: Factors such as age, gender, sample type, tested antibiotic, and Gram staining of bacteria were extracted from the MIMIC-III database and used for training six machine learning models. Results: The Extreme Gradient Boosting showed the highest prediction accuracy, at 84.53%. Conclusion: Machine Learning could offer a solution for the early detection of antibiotic resistance, thereby improving patient care and antibiotic management.

Keywords: Microbial drug resistance; Machine Learning; Intensive Care Units

Resumen

La resistencia a los antibióticos representa una preocupación significativa para la salud global, especialmente en unidades de cuidados intensivos (UCI), donde el diagnóstico rápido es esencial. Objetivo del estudio: Probar algoritmos de aprendizaje automático para predecir la resistencia bacteriana en UCI; Métodos: Se extrajeron factores como la edad, el género, el tipo de muestra, el antibiótico probado y la tinción de Gram de las bacterias de la base de datos MIMIC-III y se utilizaron para entrenar seis modelos de aprendizaje automático. Resultados: El Extreme Gradient Boosting mostró la mayor precisión en la predicción, con un 84,53%. Conclusión: el aprendizaje automático podría ofrecer una solución para la detección temprana de la resistencia a los antibióticos, mejorando así el cuidado del paciente y el manejo de los antibióticos.

Descriptores: Farmacorresistencia Microbiana; Aprendizaje Automático; Unidades de Cuidados Intensivos

Introduction

Antimicrobial resistance is escalating as a critical concern in healthcare, predicted to cause up to ten million annual deaths by 2050 according to the World Health Organization. [1] Estimates for the year 2019 suggest up to 1.27 million deaths could directly be attributed to drug-resistant bacterial infections. [2] Traditional detection methods like antibiograms are time-intensive, taking at least 24 hours for

results. [3] While faster molecular and genetic assays exist, their cost often renders them impractical for widespread use.

Early detection and treatment are vital; especially in sepsis cases, each hour's delay in initiating adequate therapy significantly elevates mortality risk. [4] Moreover, the antibiotics used to treat multidrug-resistant (MDR) bacteria usually come with a range of unfavourable side effects, adding complexity to already critical patient conditions.

Machine Learning (ML), a growing subfield of artificial intelligence, offers potential solutions. Previous research in healthcare has explored ML for genomic data processing to identify MDR genes, evaluating metabolism and cellular functions for potential antibiotic targets, and employing natural language processing (NLP) for antimicrobial stewardship. [5]

Intriguingly, recent research has indicated the feasibility of predicting MDR organisms using only demographic and lab data. [6,7] Such an approach is particularly advantageous in resource-limited settings, offering a fast and cost-effective clinical decision-making tool. This study aims to extend this line of research by testing similar predictive models on a larger dataset, seeking to replicate or exceed the accuracy reported in earlier studies.

Materials and methods

The dataset

The study employed the MIMIC-III database, a comprehensive and publicly accessible repository of ICU data collected from 2008 to 2014. Sourced from Philips CareVue Clinical Information System and iMDsoft MetaVision ICU, the database comprises a variety of data types including billing information, demographics, medication records, and lab results. [8,9,10] It is maintained by the Massachusetts Institute of Technology (MIT).

Data Extraction

Relevant data was extracted from the 'microbiology events' table, which contains information on bacterial samples, susceptibility profiles, and patient

identifiers. These identifiers link to the 'admissions' and 'patients' tables, facilitating the extraction of demographic data. [11]

Antibiotic Susceptibility and Minimal Inhibitory Concentration (MIC) Interpretation

Antibiotic susceptibility was evaluated based on MIC values, representing the minimal concentration of an antibiotic required to inhibit bacterial growth. Breakpoints distinguishing sensitive, intermediate, and resistant samples were derived, although the database documentation did not specify whether these were based on CLSI or EUCAST standards. [12] Both CLSI and EUCAST define 'sensitive' strains as those likely to be eradicated by the therapeutic concentration of the antibiotic in question, while 'resistant' strains are unlikely to be affected. The 'intermediate' category is less straightforward; its interpretation has evolved and varies between standards. [13,14] For the purposes of this study, strains classified as 'intermediate' were considered 'sensitive', considering that therapeutic success could still be achieved through dose modification or naturally higher concentrations of a given drug on the intended site of action. [15]

Feature Selection and Data Merging

Key features were extracted from selected tables in the MIMIC-III database: patient identifiers, gender, and date of birth were taken from the 'patients' table; sample type, bacterial species, tested antibiotic, and susceptibility profiles were extracted from the 'microbiology events' table; and admission date and ID were selected from the 'admissions' table. Data from these tables were then combined using inner joins, first between the 'admissions' and 'patients' tables, and subsequently with the 'microbiology events' table. The merges were performed based on subject ID and admission ID.

Data Cleaning

Entries lacking either bacterial species or tested antibiotics were dropped. Observations with pending antibiogram results were also excluded. Age was computed by subtracting the date of birth from the admission date. Patients listed with an age over 90—a protected identifier shifted to 300 in the MIMIC-III

database—were excluded to mitigate bias. The dataset was further refined to include only adult patients (18 years and older).

Gram Stain and Sample Type

Although performing a gram stain of samples immediately after collection is common, such results are not usually recorded. That is the case for the MIMIC-III database, where only the final culture results are available. For this study, we instead use the gram stain of the germ identified on the final culture results. Samples were categorized by type: urine, sputum, blood culture, etc. Samples that could represent colonization rather than infection, such as screening and swab samples, were excluded.

Antibiotic Categories

The following classes and antibiotics were included in this study: aminoglycosides (Gentamicin, Amikacin, Tobramycin), quinolones (Levofloxacin, Ciprofloxacin), Cephalosporins (Cefazolin, Cefuroxime, Ceftriaxone, Ceftazidime, Cefepime), Sulfamethoxazole/Trimethoprim, Carbapenems (Meropenem, Imipenem), Glycopeptides (Vancomycin), Penicillins (Oxacillin, Penicillin, Ampicillin, Ampicillin/Sulbactam, Piperacillin, Piperacillin/Tazobactam), Macrolides (Erythromycin), Tetracycline, Clindamycin, Nitrofurantoin, Rifampin, Linezolid, Chloramphenicol, and Daptomycin.

Observation Definition

An observation was defined as a unique combination of demographic data, sample type, bacterial gram stain, and tested antibiotic. The outcome label for each observation was the antibiotic susceptibility profile. Due to the multiple-drug testing nature of antibiograms, a single antibiogram was the source of multiple observations.

Exploratory Data Analysis (EDA)

After the data preprocessing steps, the resultant dataset included 9,214 individual patients, generating a total of 210,559 observations. The average age in the dataset was 64.34 years with a standard deviation of 15.85. Males constituted

52.81% (4,866) of the dataset. Out of the total observations, 128,846 (61.19%) were identified as gram-negative bacteria, and 152,607 (72.46%) were categorized as sensitive to the tested antibiotics. Summary statistics are provided in Tables 1 and 2 for a comprehensive understanding of the data landscape.

Table 1 – Summary demographic statistics

Age (Years)	Gender
Mean 65,34	Male (52,81%)
St.Dev 15,85	Female (47,18%)
1st quartile 55,03	
2nd quartile 67,91	
3rd quartile 78,24	
Max 89,06	

Table 2 – Summary sample statistics

Gram stain	Category
Positive (37,82%)	Sensitive (72,47%)
Negative (62,18%)	Resistant (27,52%)
Type of sample	
Urine (34,02%)	Sputum (31,86%)
Blood (21,36%)	Catheter tip (5,23%)
Tissue (2,92%)	Peritoneal fluid (1,59%)
Bile (1,29%)	Pleural fluid (0,07%)
Bronchial washings (0,06%)	Joint fluid (0,02%)
Antibiotics (class)	
Aminoglycosides (15,12%)	Quinolones (11,04%)
Cephalosporins (20,25%)	Sulfa/Trimethoprim (5,09%)
Carbapenems (7,13%)	Penicillins (21,20%)
Others (20,59%)	

Model Evaluation Metrics and Hyperparameter Tuning

- **Train-Test Split**

The dataset was randomly shuffled and split into a training set containing 90% of the samples and a test set containing the remaining 10%, stratified by antibiotic

resistance profiles. This shuffle-split process was executed prior to preprocessing to avoid data leakage.

- **Evaluation Metrics**

For performance evaluation, confusion matrices were generated to calculate true positives, true negatives, false positives, and false negatives. Other metrics such as the Receiver Operating Characteristic (ROC) curve, precision, and recall were also utilized. Due to their clinical relevance, sensitivity and specificity were also chosen as evaluation metrics alongside the area under the ROC curve and overall accuracy.

- **Implementation Details**

All models were implemented in Python. The RAPIDS library was chosen for its GPU acceleration capabilities, facilitating quicker computation. Fine-tuning was executed using Python's scikit-learn for grid search, XGBTune, and Keras Tuner for the Deep Learning model built with TensorFlow's Keras package.

- **Addressing Class Imbalance**

The skewness of class distribution in the dataset can heavily influence the model's performance, especially when the minority class (in our case, antibiotic-resistant samples) is of greater clinical importance. Therefore, data balancing techniques were employed to circumvent this issue. For balancing, we utilized the Synthetic Minority Oversampling Technique (SMOTE). SMOTE generates synthetic instances of the minority class based on the distance between a random observation and its neighbors. [18]

Models were trained on two variations of the dataset: the original imbalanced dataset and a balanced version achieved through SMOTE. This setup allowed us to assess the impact of balancing strategies rigorously.

Results

Accuracy and area under the ROC curve for each model have been summarized on table 3 and detailed below.

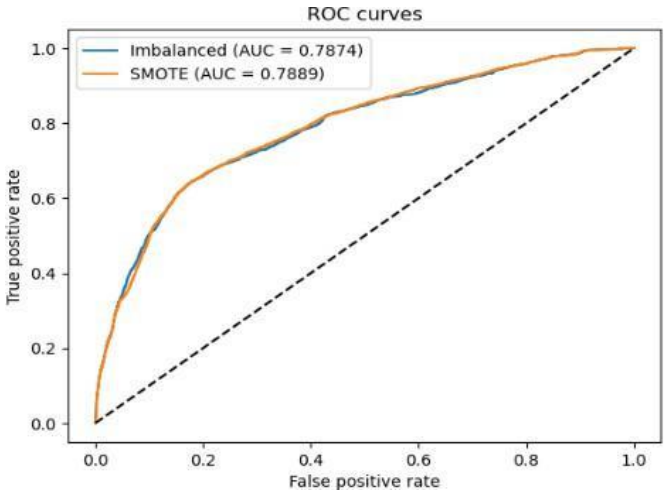
Table 3 – Model metrics summary

Model	Accuracy (imbalanced/oversampled)	AUC
Logistic Regression	0.79 | 0.77	0.7874 | 0.7889
Support Vector Machine	0.79 | 0.78	0.7770 | 0.7878
K-nearest neighbors	0.80 | 0.77	0.8174 | 0.8021
Random Forest	0.79 | 0.79	0.7863 | 0.7948
Extreme Gradient Boosting	0.81 | 0.79	0.8453 | 0.8413
Multilayer Perceptron	0.79 | 0.78	0.7987 | 0.8092

Logistic Regression

For the imbalanced dataset, the best results yielded a sensitivity of 90.64%, specificity of 48.65%, and an overall accuracy of 79.07%. The AUC was 0.7874. For the oversampled dataset, the best results obtained were a sensitivity of 82.20%, specificity of 64.31%, and an overall accuracy of 77.27%. The AUC was slightly higher at 0.7889. ROC curves are displayed in Figure 1 below.

Figure 1 – Logistic Regression ROC curves.



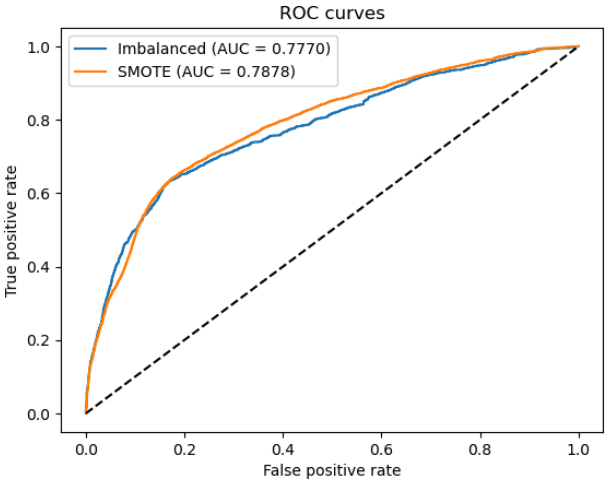
Support Vector Machine

For the imbalanced dataset, the best results led to a sensitivity of 90.67%, specificity of 49.09%, and an overall accuracy of 79.21%. The AUC (Area Under the Receiver Operating Characteristic Curve) was 0.7770.

For the oversampled dataset, the optimal model led to a sensitivity of 83.57%, specificity of 62.66%, and an overall accuracy of 77.81%. The AUC for this model

was 0.7878. Both curves are shown in Figure 2 below.

Figure 2 – Linear Support Vector Machine ROC curves.

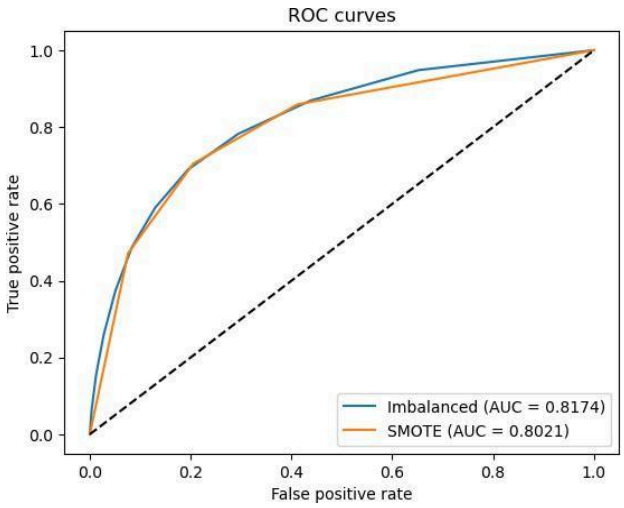


K-nearest Neighbours

For the imbalanced dataset, the best performing model had a sensitivity of 91.67%, a specificity of 48.67%, and an overall accuracy rate of 79.82%. The AUC for this model was 0.8174.

On the other hand, for the oversampled dataset, the optimal setup exhibited a sensitivity of 79.48%, a specificity of 70.47%, and an overall accuracy of 77.00%. The AUC for this model was 0.8021. Both curves are shown in Figure 3 below.

Figure 3 – K-nearest neighbors ROC curves.

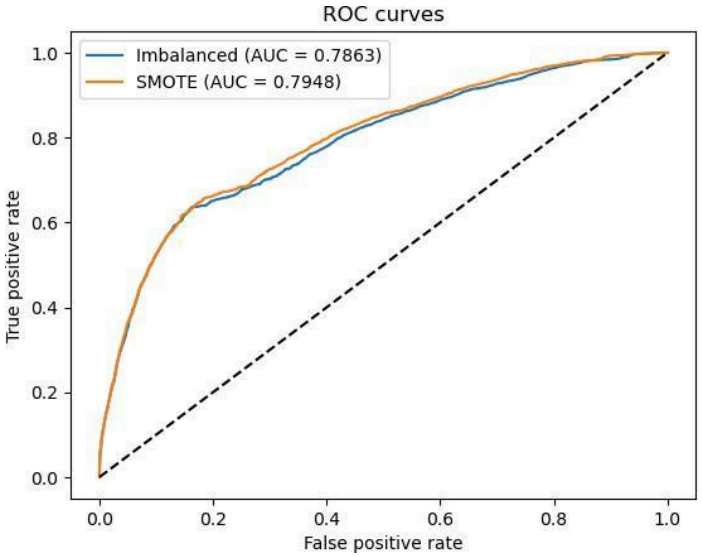


Random Forest

For the imbalanced dataset, the optimal settings yielded a sensitivity of 93.87%, specificity of 40.18%, and overall accuracy of 79.08%. The AUC was 0.7863.

For the oversampled dataset, the best settings achieved a sensitivity of 85.41%, a specificity of 61.70%, and an overall accuracy of 78.88%. The AUC was slightly higher at 0.7948. Both curves are shown in Figure 4 below.

Figure 4 – Random Forest ROC curves.

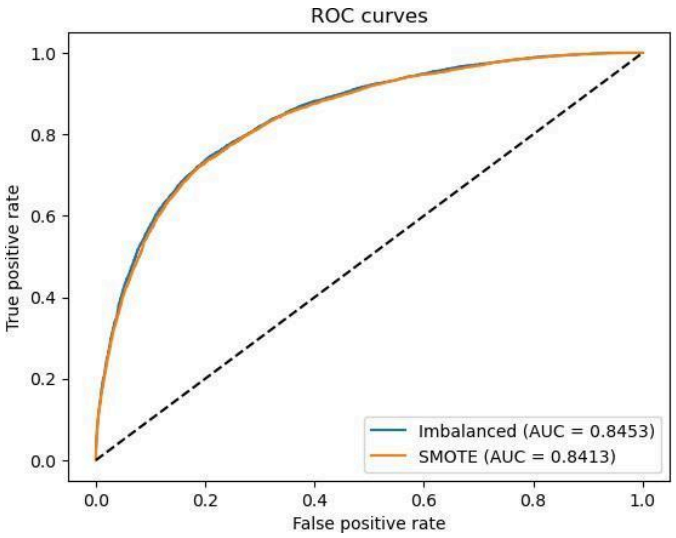


Extreme Gradient Boosting

For the imbalanced dataset, the best model had an accuracy of 81.11%, sensitivity of 91.7%, specificity of 53%, and an AUC of 0.8453.

For the oversampled dataset, the optimized parameters resulted in an accuracy of 78.78%, sensitivity of 81.61%, specificity of 71.36%, and an AUC of 0.8413. Both curves are shown in Figure 5 below.

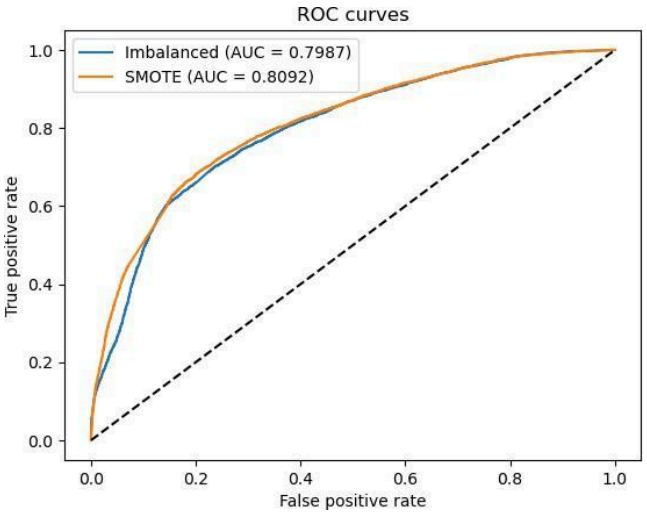
Figure 5 – Extreme Gradient Boosting ROC curves.



Multilayer Perceptron

For the imbalanced dataset, the best sensitivity was 52.92%, and the specificity was 88.79%, resulting in an overall accuracy of 78.91% and an AUC of 0.7987. For the oversampled dataset, the best sensitivity was 59.89%, and the specificity was 85.58%, resulting in an overall accuracy of 78.50% and an AUC of 0.8092. Both curves are shown on Figure 6 below.

Figure 6 – Multilayer Perceptron ROC curves.



Discussion

Our study successfully applied several Machine Learning algorithms to detect antibiotic resistance using a minimal set of laboratory data. Among these, XGBoost emerged as a standout, achieving an accuracy of 81.11%, a sensitivity of 91.7%, and a specificity of 53.28%. While these figures might not meet the rigorous standards expected of diagnostic tools in healthcare, it is crucial to recognize their efficiency and scalability. We carried out all our training and testing on a single mid-range GPU, with individual predictions taking seconds to compute.

The adoption of Machine Learning represents a paradigm shift in healthcare research methodologies. Unlike traditional medical studies, which seek to establish individual predictors' impact through rigorous control of confounding variables, Machine Learning algorithms minimize prediction error by simultaneously evaluating multiple features. This approach allows for a more dynamic and intricate understanding of variable relationships.

However, our models had limitations, most notably their sensitivity to data imbalances and outliers. Additionally, they required larger sample sizes than typically available in single-research-center studies. To address these issues, methods like few- shot learning, data augmentation, and transfer learning are actively being refined. Furthermore, public databases like MIMIC-III could offer an invaluable resource for training robust models capable of functioning in smaller, more specific datasets.

Epidemiological differences in hospital-acquired infections also pose challenges. Our models were trained on data heavily influenced by the prevalence of certain bacteria, which might not be universally applicable. For instance, the MIMIC-III dataset, gathered from a medical center in Massachusetts, closely aligns with the prevalence of gram-positive bacteria in North America, potentially limiting its applicability in environments where gram-negative bacteria are more common.

Yet, this limitation may also be seen as a strength: Machine Learning models could be fine-tuned to the unique epidemiological profiles of different healthcare institutions. Such customized models could not only enhance patient-specific decision- making but also offer hospitals a sophisticated tool for monitoring trends in multidrug- resistant bacteria. Therefore, while there is still work to be done, the initial results are promising and warrant further investigation into the application of Machine Learning algorithms in healthcare.

Conclusion

This study evaluated the efficacy of Machine Learning models in predicting antibiotic resistance using cost-effective, non-invasive techniques. Remarkably, even with a limited number of features, most of the models achieved an accuracy rate close to 80%. Future research may attempt to improve upon these results by adding more variables, and to measure the impact of epidemiological factors on model accuracy. As Machine Learning continues to evolve rapidly, such research is crucial in advancing the development of efficient approaches to the problem of antibiotic resistance.

References

1. IACG. No Time To Wait: Securing the Future From Drug-Resistant Infections Report To the Secretary-General of the United Nations [Internet]. World Health Organization. 2019. Available from: https://www.who.int/antimicrobial-resistance/interagency-coordination-group/IACG_final_report_EN.pdf?ua=1

2. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global

burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet [Internet]. 2022 Feb 12 [cited 2022 Sep 10];399(10325):629–55. Available from: http://www.thelancet.com/article/S0140673621027240/fulltext

3. March-Rosselló GA. Rapid methods for detection of bacterial resistance to antibiotics. Enfermedades Infecc y Microbiol Clínica (English Ed [Internet]. 2017 Mar 1 [cited 2022 Sep 10];35(3):182–8. Available from: https://www.elsevier.es/en-revista-enfermedades-infecciosas-microbiologia-clinica-english-428-articulo-rapid-methods-for-detection-bacterial-S2529993X17300606

4. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. Crit Care Med [Internet]. 2021 Nov 1 [cited 2022 Sep 10];49(11):E1063–143. Available from: https://journals.lww.com/ccmjournal/Fulltext/2021/11000/Surviving_Sepsis_Campaign_International.21.aspx

5. Anahtar MN, Yang JH, Kanjilal S. Applications of Machine Learning to the Problem of Antimicrobial Resistance: an Emerging Model for Translational Research. J Clin Microbiol [Internet]. 2021 Jun 18 [cited 2022 Apr 15];59(7):e0126020. Available from: https://journals.asm.org/journal/jcm

6. Feretzakis G, Sakagianni A, Loupelis E, Kalles D, Skarmoutsou N, Martsoukou M, et al. Machine Learning for Antibiotic Resistance Prediction: A Prototype Using Off-the-Shelf Techniques and Entry-Level Data to Guide Empiric Antimicrobial Therapy. Healthc Inform Res [Internet]. 2021 Jul 1 [cited 2022 Apr 16];27(3):214. Available from: /pmc/articles/PMC8369050/

7. Feretzakis G, Loupelis E, Sakagianni A, Kalles D, Martsoukou M, Lada M, et al. Using Machine Learning Techniques to Aid Empirical Antibiotic Therapy Decisions in the Intensive Care Unit of a General Hospital in Greece. Antibiot 2020, Vol 9, Page 50 [Internet]. 2020 Jan 31 [cited 2022 Apr 16];9(2):50. Available from: https://www.mdpi.com/2079-6382/9/2/50/htm

8. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet. Circulation [Internet]. 2000 Jun 13 [cited 2023 Jan 10];101(23). Available from: https://www.ahajournals.org/doi/abs/10.1161/01.cir.101.23.e215

9. Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data. 2016 May 24;3.

10. Johnson A, Pollard T MR. MIMIC-III Clinical Database (version 1.4). PhysioNet. [Internet]. MIMIC-III Clinical Database (version 1.4). PhysioNet. 2016. Available from: https://doi.org/10.13026/C2XW26

11. The microbiologyevents table | MIMIC [Internet]. [cited 2023 Jan 10]. Available from: https://mimic.mit.edu/docs/iii/tables/microbiologyevents/

12. Kahlmeter G, Brown DFJ, Goldstein FW, Macgowan AP, Mouton JW, Österlund A, et al. European harmonization of MIC breakpoints for antimicrobial susceptibility testing of

bacteria. J Antimicrob Chemother [Internet]. 2003 [cited 2023 Jan 10];52:145–8. Available from: http://www.srga.org

13. Humphries RM. Re-Exploring the Intermediate Interpretive Category. CLSI AST News Updat [Internet]. 2021 [cited 2023 Jan 10];6(1):7–9. Available from: https://clsi.org/media/3mrkwh4j/ast-news-update-volume-6-issue-1-april-2021_5-17-21.pdf

14. Yusuf E, Zeitlinger M, Meylan S. A narrative review of the intermediate category of the antimicrobial susceptibility test : relation with dosing and possible impact on antimicrobial stewardship. 2022;1–8.

15. Kahlmeter G, Steering Committee E. EUCAST proposes to change the definition and usefulness of the susceptibility category `Intermediate'. Clin Microbiol Infect [Internet]. 2017 [cited 2023 Jan 10];23:894–5. Available from: http://dx.doi.org/10.1016/j.cmi.2017.08.015

16. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. https://doi.org/101214/09-SS054 [Internet]. 2010 Jan 1 [cited 2023 Jan 11];4(none):40–79. Available from: https://projecteuclid.org/journals/statistics-surveys/volume-4/issue-none/A-survey-of-cross-validation-procedures-for-model-selection/10.1214/09-SS054.full

17. Burkov A. The Hundred-Page Machine Learning Book-Andriy Burkov. Expert Syst [Internet]. 2019;5(2). Available from: http://doi.wiley.com/10.1111/j.1468-0394.1988.tb00341.x

18. Chawla N V, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res. 2002;16:321–57