Machine Learning applied to MALDI-TOF data in a clinical setting: a systematic review

Lucía Schmidt Santiago, Alejandro Guerrero-López, Carlos Sevilla-Salcedo, David Rodríguez-Temporal, Belén Rodríguez-Sánchez, and Vanessa Gómez-Verdejo

***Abstract*— Bacterial identification, antimicrobial resistance prediction, and strain typification are critical tasks in clinical microbiology, essential for guiding patient treatment and controlling the spread of infectious diseases. While Machine Learning (ML) has shown immense promise in enhancing MALDI-TOF mass spectrometry applications for these tasks, an up to date comprehensive review from a ML perspective is currently lacking. To address this gap, we systematically reviewed 93 studies published between 2004 and 2024, focusing on key ML aspects such as data size and balance, pre-processing pipelines, model selection and evaluation, open-source data and code availability. Our analysis highlights the predominant use of classical ML models like Random Forest and Support Vector Machines, alongside emerging interest in Deep Learning approaches for handling complex, high-dimensional data. Despite significant progress, challenges such as inconsistent preprocessing workflows, reliance on black-box models, limited external validation, and insufficient open-source resources persist, hindering transparency, reproducibility, and broader adoption. This review offers actionable insights to enhance ML-driven bacterial diagnostics, advocating for standardized methodologies, greater transparency, and improved data accessibility. In addition, we provide guidelines on how to approach ML for MALDI-TOF analysis, helping researchers navigate key decisions in model development and evaluation.**

***Index Terms*— MALDI-TOF MS, Machine Learning, Deep Learning, Antimicrobial resistance, Antimicrobial susceptibility testing, Bacterial identification, Bacterial typing**

I. Introduction

Identifying and characterizing microorganisms are vital steps in clinical microbiology to diagnose infections and determine suitable treatments [1]. Accurate species identification and determination of their antibiotic susceptibility patterns enable clinicians to select the most effective antibiotic therapy, minimizing the risk of treatment failure. Rapid identification is also key for infection control, as studies show that early detection and intervention can reduce hospital-acquired infections by up to 30%, preventing complications and lowering healthcare costs [2].

Effective antibiotic treatment relies on identifying antimicrobial resistance (AMR) profiles, as resistant strains can render standard therapies ineffective, posing a global health threat [3]. Clinicians use Antibiotic Susceptibility Testing (AST), including disk diffusion, broth microdilution, and automated systems, to assess bacterial resistance [4]. However, these methods are resource-intensive and require 24 to 72 hours for results [5], which can impede timely treatment, exacerbating the burden of nosocomial outbreaks and complicating infection control efforts.

Beyond AMR, bacterial strain typification (TYP) is essential for disease management and epidemiological surveillance. TYP is based on molecular techniques such as Polymerase Chain Reaction (PCR), Pulsed-Field Gel Electrophoresis (PFGE), and MultiLocus Sequence Typing (MLST) although the field is increasingly shifting towards Whole-Genome Sequencing (WGS) for its higher resolution. However, WGS remains costly (100€ per sample), technically demanding, and often challenging to interpret, limiting its routine use in diagnostic laboratories.

The advent of Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS) in the late 1980s [6] has revolutionized microbiology, becoming a cornerstone technology for microorganism identification due to its speed, cost-effectiveness, and reliability [7], [8]. Previously, biochemical tests based on bacterial metabolic reactions—such as carbohydrate fermentation and the utilization of various compounds as energy sources—were used for identification. However, these methods often lacked the resolution needed to differentiate between closely related species [9]. By analyzing protein mass spectra, MALDI-TOF MS provides species-level identification within 5 minutes, significantly reducing diagnostic turnaround times. Its ease of use, low cost per sample (0.5€/sample) [10], and its high-throughput capabilities have made it an essential tool in clinical laboratories.

The MALDI-TOF MS procedure begins by applying an organic matrix to the bacterial sample previously deposited on the target plate, facilitating laser-induced ionization and allowing the generated ions to accelerate through a Time-of-Flight (TOF) mass analyzer. This way, ions are separated by their mass-to-charge (m/z) ratio, with lighter ions reaching the detector faster, generating a time-resolved signal and a high-dimensional mass spectrum [11], [12]. These spectra, composed of peaks at specific m/z values, act as molecular fingerprints for microbial identification, often exceeding 20,000 data points [11], [13].

Over the past few years, progress in Machine Learning (ML) has shown significant promise in automating these tasks. These technologies facilitate the accurate automatic identification (ID) of bacterial species [14]–[18] and AMR profile prediction [19]–[23], reducing resource burden. Algorithms such as Support Vector Machines (SVM), Random Forests (RF), and Neural Networks (NN) uncover complex patterns in spectral data that are inaccessible by conventional methods [24]. Such advances address the pressing need for precise and rapid diagnostic tools in clinical settings [25], [26] and present a promising alternative for the effective characterization of strains and the management of outbreaks. [27]–[31].

This work is partially supported by grant PID2023-146684NB-I00 funded by MCIN/AEI/10.13039/501100011033 and ERDF/EU and by the project TEC-2024/COM-89 (*Corresponding author*: L. Schmidt Santiago).

L. Schmidt Santiago, A. Guerrero-López, C. Sevilla-Salcedo, and V. Gómez Verdejo are with the Dept. of Signal Processing and Communications, Universidad Carlos III de Madrid, Leganés, 28911 Spain (e-mail: lschmidt@pa.uc3m.es; alguerre@pa.uc3m.es; casevill@ing.uc3m.es; vanessag@ing.uc3m.es).

A. Guerrero-López and V. Gómez Verdejo are also with the Signal Processing Group, Gregorio Marañón Health Research Institute, Madrid, 28040 Spain.

D. Rodríguez Temporal and B. Rodríguez-Sánchez are with the Dept. of Clinical Microbiology and Infectious Diseases, Hospital General Universitario Gregorio Marañón. Gregorio Marañón Health Research Institute, Madrid, 28040 Spain (e-mail: david.rodriguez@iisgm.com; mbelen.rodriguez@iisgm.com).

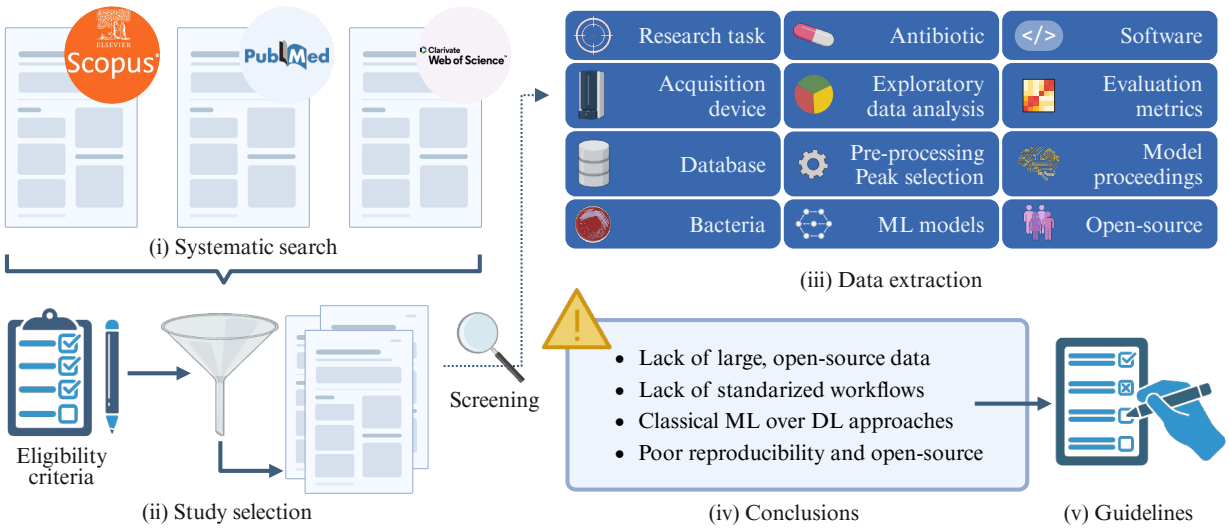

Fig. 1: **Graphical overview of the systematic review** on Machine Learning (ML) and Deep Learning (DL) applications in MALDI-TOF.

Despite advancements in this area, several challenges remain. Large high-quality datasets, standardized pre-processing, and strategies for data imbalance are essential, especially for AMR and TYP tasks. Emerging technologies such as generative Artificial Intelligence (AI) and advanced Deep Learning (DL) models present opportunities to mitigate these issues through improved data augmentation [32], improved model interpretability [33], and expanded applications of MALDI-TOF MS in clinical microbiology. Furthermore, the development of open-access spectral repositories like the Database of Resistance Information on Antimicrobials and MALDI-TOF Mass Spectra (DRIAMS) [34] and the Robert Koch Institute (RKI) [35] databases underscores the importance of collaborative data sharing to accelerate innovation in the field [21].

This review explores the impact of integrating MALDI-TOF MS with AI models on microbial diagnostics, examining its applications in ID, AMR profiling, and strain TYP, while addressing current limitations and future directions. Using the synergy between MALDI-TOF MS and cutting-edge AI technologies, we envision a future where diagnostic microbiology is faster, more precise, and accessible.

A. Motivation

The integration of ML with MALDI-TOF MS has significantly advanced clinical microbiology, enhancing microbial ID, strain TYP, and AMR prediction. However, existing systematic literature reviews, while providing valuable insights, often lack a strong technical ML perspective. *Weis et al.* [5] reviewed bacterial characterization and AMR prediction up to 2020, focusing on bacterial species, antibiotics, and model types while briefly addressing ML aspects. *López-Cortés et al.* [36] analyzed ML-based AMR prediction from 2016 to 2024 but excluded bacterial ID and strain TYP, similarly overlooking key ML workflow components such as pre-processing, data balance, and hyperparameter tuning. *Pennisi et al.* [37] took a broader approach, reviewing AI in antimicrobial stewardship beyond MALDI-TOF MS. While these reviews highlight ML effectiveness for microbiologists, they lack in-depth discussions on model design, validation practices, and open-source data, limiting transparency and reproducibility. Our review systematically evaluates ML workflows in studies from 2004 to 2024, covering every stage from data pre-processing to clinical applicability. By bridging microbiology and ML perspectives, we provide methodological guidelines for researchers with clinical backgrounds to design robust ML workflows. Additionally, while we reference *Weis et al.* for pre-2020 studies, we apply stricter selection criteria, ensuring a rigorous and transparent assessment of ML applications, specifically for human bacteria.

B. Objective

This review examines ML applications for bacteria ID, TYP, and AMR detection using MALDI-TOF MS, expanding on *Weis et al.* [5] and incorporating studies from 2020–2024. Unlike previous reviews, we take a deeper ML-focused approach, critically evaluating:

- Trends in task scope, data size, instrumentation, and bacterial/antibiotic coverage.
- Exploratory data analysis, pre-processing, feature selection, and data imbalance handling, highlighting the lack of homogenous pre-processing pipelines.
- Model selection, highlighting black-box reliance and its impact on transparency, interpretability, and reproducibility.
- Evaluation metrics, Cross-Validation (CV), and external validation, identifying gaps in standardization.
- Open-source challenges in dataset and software accessibility, proposing strategies for greater transparency and reliability.
- Limitations of ML in MALDI-TOF MS and future directions for clinical integration.

Additionally, we propose best practices for future ML-driven clinical microbiology studies. This way, this review serves as a roadmap to improve transparency, standardization, and real-world applicability in this interdisciplinary field. Fig 1 provides a structured overview of this paper.

II. Methods

A. Search strategy

A comprehensive literature search was conducted using *PubMed/Medline*, *Scopus*, and *Web of Science*, covering the period from January 1st, 2020 up to December 31st, 2024. The search aimed to identify studies that applied ML techniques to the analysis of MALDI-TOF MS data. The search included related terms[1] applied to titles, abstracts, and keywords when available.

[1]The search query was: 'machine learning' OR 'classification' OR 'support vector machine' OR 'random forest' OR 'logistic regression' OR 'neural network' OR 'deep learning' OR 'transformer' OR 'multi-layer perceptron' OR 'embedding' OR 'clustering' OR 'unsupervised' OR 'self-supervised' OR 'semi-supervised' OR 'selfsupervised' OR 'semisupervised' OR 'bayesian' OR 'kernel' AND 'maldi-tof'.

B. Study selection

The study selection process was guided by predefined inclusion criteria, as outlined in the search strategy and further refined by applying exclusion criteria to ensure the relevance and quality of the studies. The following exclusion criteria were applied:

- **Duplicate records:** Duplicates across databases (*PubMed/Medline*, *Scopus*, and *Web of Science*) were removed during the initial screening phase.
- **Non-full text and irrelevant types:** Records such as review articles, letters, conference proceedings, or those unavailable as full-text articles in English were excluded.
- **Off-topic studies:** Studies unrelated to MALDI-TOF MS and AI-based methods, such as research focused on genetics, proteomics, cancer, or other unrelated fields, were excluded.
- **Non-human bacteria:** Articles focusing on non-human bacteria, such as those related to foodborne pathogens, animal studies, viruses, or fungi, were removed.
- **Off-task focus:** Studies that did not directly address ID, AMR, or TYP were excluded. Examples include studies assessing risk factors, mortality, morbidity prediction or case studies.

This exclusion process ensured that the selected studies aligned with the objectives of this review. The PRISMA flow diagram (Figure 2) summarizes the screening process and the final set of included studies. A total of 1,888 documents were retrieved from *PubMed/Medline*, *Scopus*, and *Web of Science*. For the period before 2020, we incorporated the 36 studies from *Weis et al.* [5]. After removing 871 duplicates, 1,053 unique articles remained. Of these, 128 were excluded for being non-full-text or irrelevant document types, 378 were off-topic, and 347 focused on non-human bacteria. An additional 107 studies were removed for being off-task. Ultimately, 93 studies published between 2004 and 2024 met our inclusion criteria.

Notably, from the reviews by *Weis et al.*, *López-Cortés et al.*, and *Pennisi et al.*, only 25, 30, and 8 studies, respectively, are included in our review. This is due to our ML-specific search criteria, which cover a broader range of ML techniques as outlined in 1, lead to the inclusion of additional studies. Conversely, some studies analyzed in previous reviews were excluded from ours due to not analyzing MALDI-TOF data from human bacteria. Consequently, the overlap with more generalist reviews, such as *Pennisi et al.*, remains minimal, as they include other approaches beyond MALDI-TOF.

C. Data extraction

To ensure a comprehensive analysis of methodologies and results, the following information was extracted from each study: (i) **publication details** including first author's name, year of publication, and article title; (ii) **research task** categorized by bacterial species ID, AMR detection, or strain TYP; (iii) **bacterial scope** captured the genera and species investigated; (iv) **antibiotics** analyzed for AMR profiling, if applicable; (v) **sample size** recorded the number of bacterial isolates or mass spectra used in the analysis.

For computational methodologies, we documented (vi) **ML algorithms** used, detailing models, architectures, and techniques; (vii) **pre-processing pipelines**, outlining steps like normalization, feature extraction, and dimensionality reduction; (viii) **analysis platforms and tools**, identifying the programming languages used for ML implementation; and (ix) **performance metrics and methods**.

To ensure robustness and generalizability, (x) **validation strategies** were assessed, including inner CV methods for hyperparameter tuning, outer CV, and Out-of-Distribution (OOD) evaluation. Finally, (xi) **open-source practices** were evaluated to analyze transparency and reproducibility. If available, supplementary materials were reviewed to gather additional information.

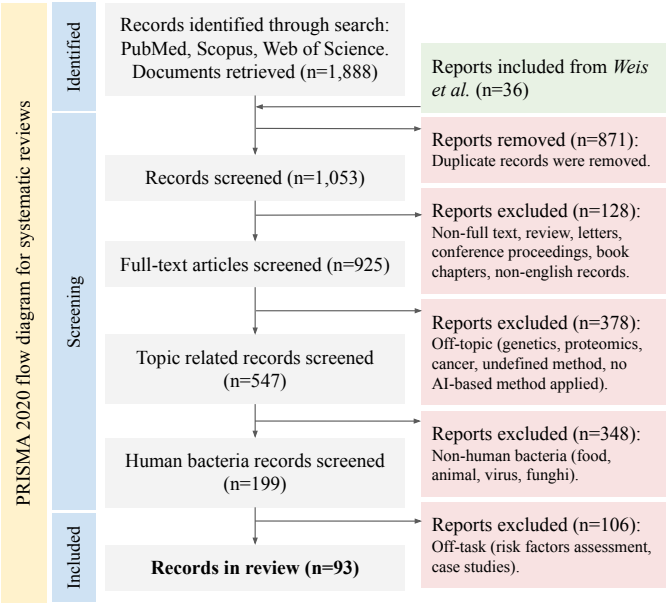

Fig. 2: PRISMA flowchart illustrating the literature search pipeline.

This systematic approach ensured consistency in the literature analysis and facilitated robust comparative analysis across all included studies. Table I summarizes the extracted data.

D. Quality assessment

Based on the data extracted from each study following Section II-C, we assess its quality according to the following key aspects:

- **ML methodology:** Clear description of the algorithms used, including model architectures, hyperparameter validation ranges, and training procedures.
- **Model evaluation:** Adequate evaluation metrics and validation strategies to ensure robustness and generalization, such as k-fold inner CV or the use of external datasets for OOD evaluation.
- **Data transparency:** Availability of datasets and/or code to enable reproducibility.
- **Clinical relevance:** The practical application of the findings, particularly in identifying and typing bacterial species and predicting antimicrobial resistance in clinical settings.

Each of these aspects will be explored in detail in the different sections of Section III, while broader implications and key challenges will be discussed in Section IV.

III. Results

This section presents the main findings from the n=93[2] studies under review, examining the tasks, data characteristics, and methodologies employed. The results offer an overview of the current state of ML applications in MALDI-TOF MS data analysis, highlighting strengths, limitations, and key aspects outlined in Section II-D.

A. Key research task

A total of n=19 papers focus on bacterial species ID, revealing that, while MALDI-TOF MS is a gold-standard technique for ID, it struggles to differentiate certain species. For instance, those within the *Streptococcus* genus or between closely related genera like *Escherichia coli* and *Shigella* [30], [116]. These limitations

[2]To avoid confusion with cites, the number of studies is denoted as n=X.

TABLE I: **Overview of Included Studies (2004–2024).** This table summarizes 93 studies categorized by year of publication. A single quotation mark (’) in the *Isolates* column indicates spectra; otherwise, it refers to bacterial isolates or samples. Algorithms in **bold** represent the primary models used or those achieving the best performance. An asterisk (*) in the *Pre-processing* column indicates steps subjected to ablation studies. Notation regarding bacteria, antibiotics, preprocessing steps, and algorithms can be seen in II and III.

Year	Paper	Task	Bacteria	Antibiotic	Isolates	Pre-processing	Peak Selection	Algorithm
2004	Zhang [38]	ID	Bacillus, S.au	-	69	N	-	RBF-**MLP**
	Satten [39]	ID	E.co, S.py, B.an, S.pn	-	120	S	I	**RF**
2008	Hsieh [40]	ID	S.au, E.co, K.pn, S.en, P.ae, GBS	-	132	N	-	QC, SVM, **GA**
2012	Schaumann [41]	AMR	E.co, K.pn, P.ae	ESBL, MBL	76	-	-	**SVM**
2013	Khot [42]	ID	Shigella, E.co	-	138	BaNA	O	**GA**
2014	Nakano [43]	AMR	E.fa	VAN	132	-	GA	**GA**, MLP, QC
	Li [44]	AMR	E.co, K.pn	CTX	103	T	SNR	GA, **MLP**, QC
	Xiao [45]	TYP	M.pn	-	68	BaNA	AUC	**GA**
2015	Zhang [46]	TYP	S.au	-	290	BaSNTA	AUC, SNR	**GA**
	Nakano [47]	TYP	S.pn	-	574	-	-	GA, **MLP**, QC
	Chen [48]	ID	S.pn, S.mi, S.or	-	75	-	S	GA, **SVM**, MLP, QC
2016	Camoez [49]	TYP	S.au	-	95	BaA	-	GA, **MLP**, QC
	Mather [50]	AMR	S.au	VAN	80	MQ	SNR	**SVM**
2017	Ho [51]	TYP	B.fr	-	424	-	AUC	GA, **SVM**, MLP, QC
	Sogawa [52]	AMR	S.au	MET	160	-	-	**SVM**
2018	Wang [53]	TYP	S.au	-	125	BaST	SNR, O	DT, **SVM**, KNN
	Wang [54]	TYP	S.au	-	306	BaS	S, GA	QC, MLP, **KNN**
	Schaumann [41]	ID	Clostridium	-	123	-	-	**SVM**
	Rodrigues [55]	TYP	Klebsiella	-	95	BaSTA	SNR	**SVM**
	Asakura [56]	AMR	S.au	VAN	129	BaSNBi	RF	**RF**
	Wang [57]	AMR	S.au	VAN	125	BaST	RF, SNR	**SVM**, KNN, DT, RF
2019	Ling [58]	ID	Shigella, E.co	-	23	-	-	**MLP**
	Chung [59]	TYP	S.ha	-	254	BaS	RF, AUC	LR, SVM, DT, **RF**
	Tang [60]	AMR	S.au	MET	20	MQ*	RF, O	BI, **RF**
	Wang [61]	TYP	GBS	-	787	BaSTBi*	SNR, O	RF, **SVM**
2020	Desaire [14]	ID	20	-	571	N	O	**AC**
	Pena [27]	TYP	K.pn	-	1,056	BaSTP	SNR, O	**SVM**
	Huang [19]	AMR	K.pn	IPM, MEM	95 ’	Ba	S	SVM, **RF**, KNN, NB, LR
	Ling [62]	ID	14	-	98,000	BaNT	SNR	**MLP**
	Kann [63]	TYP	S.pn	-	69	MQ	O	RF, **CART**
	Wei [64]	ID	B.an, B.ce	-	178	BaNTA	GA, SNR	**GA**
	Weis [25]	AMR	E.co, K.pn, S.au	AMC, CRO, CIP, TZP, PEN	2,676	MQN*	-	LR, PIKE-**GP**
2021	Wang [65]	AMR	S.au	MET	4,858	NBi	S, SNR	DT, **RF**, KNN, SVM
	Gato [66]	AMR	K.pn	MEM, ETP	1,458	BaSA	-	PLS-DA, **RF**
	Mortier [15]	ID	1,000	-	89,767	BaNBi*	-	RF, LR, SVM, **KNN**, **MLP**
	Wang [67]	AMR	E.fa	VAN	7,997	BaSTBi	S	**RF**, SVM, KNN
	Fiamanya [68]	TYP	B.co	-	55	BaSNTP	SNR	**GA**, MLP, **QC**
	Tang [69]	AMR	S.au	MET	960	-	GA	**GA**, MLP, QC
	Chung [70]	AMR	S.au	OXA, CLI, ERY	25,217 ’	BaT*	RF, SNR	DT, SVM, **RF**
	Calderaro [28]	TYP	C.di	-	950	BaSTA	GA, SNR	**GA**, QC, MLP
	Huang [71]	TYP	GBS	-	1,410	BaS	GA	**KNN**, MLP, QC
	Liu [72]	AMR	S.au	MET	452 ’	MQBi	O	RBF-**SVM**, RF
	Calderaro [20]	AMR	E.co	COL	1,040	BaSTAP	S, SNR	GA, QC, **MLP**

Year	Paper	Task	Bacteria	Antibiotic	Isolates	Pre-processing	Peak Selection	Algorithm
2022	Li [16]	ID	Listeria	-	1,429 ’	BaSN*	O	KNN, NB, RF, SVM, MLP, **DAE-SVM**
	Feucherolles [73]	AMR	C.je, C.co	CIP, ERY, TCY, GEN, KAN, STR, AMP	340	BaNTA	RF, SNR	**RF**, LR, NB
	Weis [21]	AMR	803	42	300,000	MQBi	-	LR, **LGBM**, **MLP**
	Kong [74]	AMR	S.au	MET	548	NBi*	O	GA, **SVM**, DT, RF, PR
	Wang [75]	AMR	E.fa	VAN	7,997	BaST	SNR	**MLP**, RF, XGB
	Wang [76]	AMR	K.pn	CIP	15,782	BaSTBi	S, SNR, O	LR, **SVM**, RF, **XGB**
	Dematheis [77]	ID	B.me, B.ab, B.su	-	44	MQA	RF	**RF**, SVM, MLP, **MNR**
	van den Beld [78]	ID	Shigella, E.co	-	559	BaSA	O	**SVM**
	Jeon [79]	AMR	S.au	MET	553	BaSNABi	RF	**RF**
	Candela [80]	AMR	E.fa	VAN	178	BaSNA*	-	**SVM**, RF, PLS-DA
	Zhang [81]	AMR	S.au	OXA, CLI	31,815	BaSTBi	SNR, O	LR, **XGB**
	Wang [82]	AMR	K.pn	IPM, MEM, ETP	171	-	RF	RF, SVM, **SVM-k**
	Yu [83]	AMR	S.au	MET	20,359	BaSNBi*	-	**LGBM**, GB, LR, XGB, ET, RF, SVM, DT, KNN, NB
	R.-Villodres [84]	AMR	E.co	TZP	248	BaSNTA*	SNR	PLS-DA, SVM, **RF**
2023	Wang [85]	AMR	S.au	PEN, OXA, ERY, CLI, SXT, FUS	31,807	BaNTBi	SNR, O	**XGB**
	G.-López [86]	AMR	K.pn	15	402	MQ*	O	KSSHIBA-**FA**, RF, SVM, PIKE-GP
	Candela [17]	ID	E.cl	-	337	BaSNA	-	PLS-DA, SVM-L, **SVM-R**, **RF**
	Chung [87]	AMR	E.co	AMX, CAZ, CIP, CRO, CXM	37,918	BaST	RF, SNR	LR, SVM, RF, **XGB**
	Gato [88]	AMR	K.pn	MEM, ETP	715	BaSA*	-	SVM, PLS-DA, KNN, **RF**
	Yu [89]	AMR	K.pn	IPM, MEM, ETP, COL	3,938	BaSNBi	-	**LGBM**
	Abdrabou [29]	TYP	C.di	-	240	BaSNA*	I	SVM, **PLS-DA**, KNN, **RF**
	Zintgraf [90]	TYP	S.pn	-	238	MQAPBi*	O	GLMNET, **RF**, NB, KNN, LDA, MLP, SVM, DT
	R.-Temporal [91]	ID	M.ab	-	325	BaSN	RF	PLS-DA, SVM, **RF**, KNN
	Zeng [22]	AMR	K.pn	IPM	174	BaSNBi	-	**LASSO**, LR, SVM, MLP
	Yu [92]	AMR	K.pn, S.au	IPM, MEM, ETP, MET	170	BaSNBi	-	**LGBM**
	Wang [93]	TYP	S.si	-	205	MQ	-	**RF**, SVM
	Fu [94]	AMR	P.ae	TOB, FEP, MEM	4,039 ’	MQBi	RF	**MLP**
	Manfredi [30]	TYP	E.co	-	202	BaS	S	MLP, **GA-KNN**, QC
	Wang [95]	ID	M.ab	-	102	Bi	RF	**RF**, SVM, DT, LR, KNN
	Iskender [96]	AMR	K.pn	COL	260	BaN	GA	**LDA**, SVM
	Liu [97]	ID	B.lo, B.in	-	100	SN	O	**RF**, SVM, LR
	Candela [98]	TYP	P.ae	-	67	BaSNA	O	PLS-DA, SV, **RF**, NCA-KNN
	Zhang [99]	AMR	K.pn	IPM, MEM, DOR, ETP	2,683	MQBi	O	**MLP**
	Chung [100]	AMR	A.ba, A.no, E.fa, GBS	CIP, VAN, CLI	22,101	MQBaS*	RF	LR, NB, **RF**, SVM
2024	Wang [101]	AMR	S.au	OXA, ERY, CLI, TCY, SXT, MET	5,290 ’	BaSNBi	O	**LGBM**, MLP, RF, KNN, AdaB, LR, MLP
	de Waele [102]	AMR	25	74	55,773 ’	BaSNBi	-	**MLP**, LR, XGB, MLP
	Candela [103]	TYP	C.di	-	69	BaSNA	-	KNN, SVM, PLS-DA, **RF**, **LGBM**
	Lin [104]	AMR	K.pn	CPA	675	NBi	RF	LR, LDA, RF, GB, AdaB, XGB, **LGBM**
	Jian [105]	AMR	K.pn	COL, DOR, IPM	4,307	NBi	O	LDA, **RF**, GB, AdaB, XGB, LGBM, SVM
	Huang [106]	TYP	GBS	-	167	A	GA	GA, MLP, QC, **KNN**
	Godmer [31]	TYP	C.di	-	201	MQ	SNR	MNR, MLP, **RF**, SVM
	Mao [18]	ID	E.co	-	38,400 ’	T	SNR	**LSTM**
	Lin [23]	AMR	S.ma	LVX, SXT	7,266	NBi	RF	LR, LDA, AdaB, **LGBM**, XGB, **RF**, GB
	Xu [107]	AMR	K.pn	ETP, IPM, MEM	240	-	O	**RF**, SVM, LR, XGB
	L.-Cortés [108]	AMR	E.co, K.pn, S.au	CIP, CRO, FEP, TZP, TOB, MEM, FUS, OXA	11,600 ’	BaSBi	-	**MLP**
	Astudillo [109]	AMR	E.co, K.pn, S.au, P.ae	OXA, CLI, FUS, CIP, CRO, TZP, FEP, IPM, MEM	13,157 ’	Bi	-	**MLP**, **SVM**, RF, XGB
	Jian [110]	AMR	K.pn	LVX, CIP	11,996	NBi	O	LR, LDA, **RF**, GB, AdaB, XGB, LGBM
	Macaya [111]	AMR	E.co, K.pn, S.au	CRO, OXA	30,069	NBi	-	**MLP**
	Nguyen [112]	AMR	P.au	11	360	MQBi*	-	LR, RF, **SVM**, LGBM, MLP, **T**
	Ren [113]	AMR	S.ep	20	75,042 ’	MQBi	-	RF, LR, NB, **SVM**, **LGBM**, MLP

TABLE II: List of used bacteria and antibiotic codes [114].

Code	Bacterial Species	Code	Bacterial Species
A.ba	*Acinetobacter baumannii*	S.py	*Streptococcus pyogenes*
A.no	*Acinetobacter nosocomialis*	S.si	*Streptococcus suis*
B.an	*Bacillus anthracis*	S.pn	*Streptococcus pneumoniae*
B.ce	*Bacillus cereus*	S.or	*Streptococcus oralis*
B.fr	*Bacteroides fragilis*	S.mi	*Streptococcus mitis*
B.in	*Bifidobacterium infantis*	S.ma	*Stenotrophomonas maltophilia*
B.lo	*Bifidobacterium longum*	S.ha	*Staphylococcus haemolyticus*
B.ab	*Brucella abortus*	S.ep	*Staphylococcus epidermidis*
B.me	*Brucella melitensis*	S.au	*Staphylococcus aureus*
B.su	*Brucella suis*	S.en	*Salmonella enterica*
B.co	*Burkholderia contaminans*	P.ae	*Pseudomonas aeruginosa*
C.co	*Campylobacter coli*	M.pn	*Mycoplasma pneumoniae*
C.je	*Campylobacter jejuni*	M.ab	*Mycobacterium abscessus*
C.di	*Clostridium difficile*	K.pn	*Klebsiella pneumoniae*
E.cl	*Enterobacter cloacae*	GBS	*Group B Streptococcus*
E.fa	*Enterococcus faecium*	E.co	*Escherichia coli*
Code	**Antibiotic**	**Code**	**Antibiotic**
AMC	Amoxicillin/Clavulanic acid	FUS	Fusidic acid
AMP	Ampicillin	GEN	Gentamicin
AMX	Amoxicillin	IPM	Imipenem
CAZ	Ceftazidime	KAN	Kanamycin
CIP	Ciprofloxacin	LVX	Levofloxacin
CLI	Clindamycin	MBL	Metallo β-lactamases
COL	Colistin	MEM	Meropenem
CPA	Ceftazidime-avibactam	MET	Methicillin
CRO	Ceftriaxone	OXA	Oxacillin
CTX	Cefotaxime	PEN	Penicillin
CXM	Cefuroxime	STR	Streptomycin
DOR	Doripenem	SXT	Trimethoprim/Sulfamethoxazole
ESBL	Extended spectrum β-lactamases	TCY	Tetracycline
ERY	Erythromycin	TOB	Tobramycin
ETP	Ertapenem	TZP	Piperacillin/Tazobactam
FEP	Cefepime	VAN	Vancomycin

TABLE III: List of abbreviates.

Code	Pre-processing	Code	Peak Selection
MQ	MALDIquant [115]	AUC	Area Under Curve
Ba	Baseline Subtraction	GA	Genetic Algorithm
S	Smoothing	I	Intensity Thresholding
N	Normalization	RF	Random Forest
T	SNR Thresholding	S	Statistical Tests
A	Averaging	SNR	Signal-to-Noise Ratio
Bi	Binning	O	Other
Code	**Algorithm**	**Code**	**Algorithm**
AC	Aristotle Classifier	LGBM	Light GB Machine
AdaB	AdaBoost	LR	Logistic Regression
BI	Binary Discriminant Analysis	LSTM	Long Short-Term Memory
DAE	Denoising Autoencoder	MLP	Multi-Layer Perceptron
DT	Decision Tree	MNR	Multinomial LR
CART	Classification Regression Tree	NB	Naïve Bayes
ET	Extra Trees	PLS-DA	Partial Least Squares Discriminant Analysis
FA	Factor Analysis	PR	Polynomial Regression
GA	Genetic Algorithm	QC	Quick Classifier
GB	Gradient Boosting	RBF	Radial Basis Function
GLM	Generalized Linear Model	RF	Random Forest
GP	Gaussian Process	SVM	Support Vector Machine
KNN	k-Nearest Neighbours	T	Transformer
LASSO	Least Absolute Shrinkage and Selection Operator	XGB	Extreme GB
LDA	Linear Discriminant Analysis		

underscore the need for enhanced techniques. Some studies (n=10) explore bacterial ID across multiple genera [17], [38]–[40], [42], [61], [62], [78]. Notably, *Desaire et al.* [14] identified 20 species, while *Mortier et al.* [15] classified over 1,000, demonstrating ML ability to distinguish a wide range of bacterial classes.

AMR is the most common task across the studies, totaling n=51 studies focused on predicting resistance profiles to specific antibiotics. Among those handling multiple AMR profiles for one bacteria species (n=10), most do not adopt a multitask approach [21], [25], [41], [44], [92], [100], [108], [111], instead training separate AMR models for each species. Only one simultaneously predicts both bacterial ID and AMR [102]. Likewise, AMR prediction for multiple antibiotics is typically performed with separate models per bacteria-antibiotic pair, with only n=3 studies integrating species-specific AMR predictions for all tested antibiotics [81], [102], [109].

The remaining studies (n=23), focus on TYP. This remarks that research on TYP has declined in favor of AMR, with studies on bacterial strain prediction dropping from 35% before 2022 to 16% after. This shift is likely due to TYP being limited to genetic studies rather than routine clinical practice, making large datasets for training ML models scarce. In contrast, AMR labeling is routinely performed, providing abundant data and driving research toward AMR prediction.

Summary. The scarcity of large datasets for TYP has hindered research in this area, contributing to a decline in TYP-focused studies and a shift toward AMR prediction. Although multi-task learning could enhance efficiency by addressing multiple related tasks simultaneously, its potential remains largely unexplored. To date, only one study has tackled bacterial ID and AMR prediction together, highlighting an opportunity for improving automated screening of multidrug-resistant profiles.

B. Data acquisition device

The reviewed studies employ various MALDI-TOF MS data acquisition devices. The two main manufacturers are Bruker Daltonics [117] and BioMérieux [118]. Both provide a full pipeline, from performing MALDI-TOF MS measurements of the cultured isolate to species identification based on database lookup. Nevertheless, species ID from MALDI-TOF MS remains an active field of research, as several bacteria, such as *Burkholderia cepacia* complex or *Citrobacter freundii* complex, are still not identifiable at the species level [25].

The most common instrument is the Bruker Microflex Biotyper (n=74), which is favored due to its high throughput capabilities and well-established software tools, such as flexAnalysis [119] and ClinProTools [120], which facilitate data analysis and integration with ML algorithms. The VITEK MS system by BioMérieux (n=10) appears less frequently in the reviewed literature. VITEK MS is valued for its ease of use in clinical settings but lacks the widespread adoption of research seen with Bruker systems. One study [101] uses data acquired from Bruker as well as bioMérieux.

In addition, n=8 make use of less common MALDI-TOF instruments or customized systems for specific research purposes, such as Zybio Inc., Applied Biosystems, PerSeptive Biosystems, Absciex, ASTA MicroIDSys, or Voyager DE-STR.

Summary. Bruker and BioMérieux dominate the industry, with the majority of studies relying on their MALDI-TOF MS systems. While new manufacturers have entered the market in recent years, their adoption remains regionally limited.

C. Data description

Before applying any ML model, it is crucial to thoroughly understand the dataset. In MALDI-TOF MS studies, this begins with clearly distinguishing between bacterial samples (isolates) and mass spectra, as this distinction impacts data processing and analysis. A bacterial sample or isolate refers to a cultured strain obtained from a clinical or environmental source, representing a unique biological entity [121]. In contrast, a spectrum represents the protein fingerprint of an isolate, generated through MALDI-TOF MS analysis [122].

Depending on the experimental design, a single isolate may yield multiple spectra due to biological or technical replicates or varying acquisition or pre-processing conditions. This distinction is crucial for dataset sizing, as some studies report isolates [72], others only spectra [16], and some specify both, including technical replicates [29], ensuring a clearer count of available spectra. In general, isolates are the most relevant unit of measurement for ML applications, but some studies (n=13) only report the number of MALDI-TOF spectra (see Table I), leaving the number of isolates unspecified. This lack of standardization in reporting data sizes complicates comparisons across studies.

Dataset sizes in the reviewed studies vary significantly, ranging from just 44 isolates [27] to over 300,000 [21]. The majority (n=57) rely on small datasets with fewer than 500 isolates, while n=14 use medium-sized datasets (500–10,000 isolates), and only n=9 incorporate more than 10,000 samples. The median dataset size is 209 isolates and 1,443.5 spectra.

Data imbalance is a frequent issue, with only one-third of studies (n=30) working with balanced datasets [18]–[20], [22], [39], [40], [43], [44], [49], [51], [52], [60], [62], [65], [67], [69]–[71], [75], [79], [80], [82]–[84], [88], [90], [95]–[97], [100], while n=53 report imbalances, particularly in AMR and TYP tasks. Notably, n=10 studies fail to disclose balance status. To address this, n=16 studies apply resampling methods [17], [23], [25], [31], [61], [76], [77], [81], [86], [91], [94], [104], [105], [109]–[111], including SMOTE [123] (n=5) [31], [76], [94], [109], [111]. Others (n=14) use imbalance-aware evaluation metrics such as AUPRC or balanced accuracy [16], [21], [25], [31], [73], [86], [94], [98], [99], [101], [108], [109], [111], [113], while only n=3 adjust loss functions [25], [107], [112]. It is concerning that n=27 acknowledge working with imbalanced datasets without applying corrective techniques.

Summary. Establishing systematic data handling and reporting practices—explicitly stating dataset size, biological and technical replicates, and their management—is crucial to prevent bias, misinterpretation, and irreproducibility, ensuring scientific rigour. The lack of medium-large datasets remains a major challenge, though large repositories are beginning to mitigate this issue. Smaller datasets, while easier to handle, often hinder model reliability and generalizability, especially without robust external validation. ML model performance is further constrained by class imbalance, yet standardized methodologies to address these challenges remain underutilized.

D. Bacteria

The reviewed studies primarily focus on a few bacterial species that are clinically significant due to their association with hospital-acquired infections and AMR. The most prevalent species studied is *Staphylococcus aureus* (n=26), which is extensively studied particularly for the detection of *methicillin-resistant S. aureus* (MRSA), a notorious pathogen responsible for numerous healthcare-associated infections [124], [125]. Following closely, *Klebsiella pneumoniae* (n=21), reflecting its critical role in carbapenem-resistant infections, which pose a major challenge in healthcare settings [126]. *Escherichia coli* (n=16), known for its role in urinary tract infections and rising resistance to antibiotics such as *fluoroquinolones* [127] and *carbapenems* [128], is also a frequent subject of analysis.

The majority of studies (n=73) focus on bacteria from a single genus, such as *Streptococcus* [63], [71], [90], [93]. In contrast, a smaller subset (n=20) examines bacteria across multiple genera, of which only n=7 address more than three genera. This tendency likely stems from the focus of most AMR-related studies on single-species bacterial analysis for susceptibility to one or multiple antibiotics (see Section III-E).

Summary. The range of studied bacteria is broad, covering 37 specified species, yet over 67% of studies focus on just three: *S. aureus*, *K. pneumoniae*, and *E. coli*. This concentration highlights a gap in research on lesser-studied pathogens, limiting the generalizability of ML models in clinical microbiology.

E. Antibiotics under study

The reviewed studies examine a diverse range of antibiotics to predict AMR across various bacterial species, encompassing a total of 30 different antibiotics and 3 antibiotic groups—namely carbapenems, Extended spectrum β-lactamases, and Metallo β-lactamases. The scope of antibiotics analyzed per study varies significantly; while n=22 studies focus on a single antibiotic, others (n=29) investigate resistance across a broader spectrum. Notably, the largest number of antibiotics examined in a single study reaches 74 [102]. As shown in Table I, the number of antibiotics analyzed per study as well as the number of studies about AMR has increased over time, underscoring the growing importance of addressing multiresistant pathogens.

Among the antibiotics studied, *carbapenems* (CP) are the most frequently analyzed (n=18), including *meropenem* (n=11), *imipenem* (n=9), *ertapenem* (n=7), and *doripenem* (n=2). Their critical role in treating infections caused by multidrug-resistant pathogens makes them a primary focus of AMR research. However, the growing threat of CP-resistant bacteria poses a significant public health threat [129] and has led many studies to classify resistance broadly as CP-resistant vs. susceptible (n=10), and others differentiating between specific CP drugs (n=8).

Beyond CP, other commonly studied antibiotics include *methicillin* (n=10), primarily in the context of MRSA detection [52], [60], [65], [69], [72], [74], [79], [83], [92], [101], and *vancomycin* (n=8), widely analyzed for its role in *Enterococcus faecium* and *vancomycin*-resistant *Enterococci* (VRE) [43], [67], [75], [80], [100]. Additionally, *ciprofloxacin* (CIP) (n=8) is frequently studied in *E. coli* and *K. pneumoniae* for urinary tract and systemic infections [25], [76], [87], [108]–[110]. *Oxacillin* (n=7) is consistently examined for its relevance in detecting *S. aureus* resistance, while *clindamycin* (n=6) is another frequently assessed antibiotic. Lastly, *colistin* (n=4), a last-resort antibiotic, is commonly investigated for resistance in *K. pneumoniae* and other Gram-negative bacteria [20], [89], [96], [105].

However, studying multiple antibiotics in a single paper does not mean that AMR prediction is performed for all drugs simultaneously within one model. In fact, only n=3 studies [81], [102], [109] employ multilabel classification to predict AMR profiles. *Zhang et al.* [81] predict *S. aureus* resistance to two antibiotics, while *Astudillo et al.* [109] perform species-level multilabel classification for nine antibiotics across four bacterial species. *De Waele et al.* [102] explore three approaches: a single model for all antibiotics and species, individual models per species, and species-antibiotic pair models.

The World Health Organization Bacterial Priority Pathogens List (WHO BPPL) has served as a tool for guiding research and development (R&D) investments, as well as shaping surveillance and control strategies for AMR [130]. The 2024 BPPL update categorizes antibiotic-resistant bacteria into three priority levels: critical, high, and medium. Analyzing how the literature aligns with these categories, it can be observed that for the critical-priority group where it is included CP-resistant *Acinetobacter baumannii* and *Enterobacterales*, such as *K. pneumoniae* and *E. coli*, in cases of CP and third-generation *cephalosporin* resistance (e.g., CTX, CAZ, CPA, CRO), we find n=0, n=12 and n=7, respectively. In the high-priority category there are studies for CP-resistant *Pseudomonas aeruginosa* (n=3), MRSA (n=10), VRE (n=5), and *fluoroquinolone*-resistant *Salmonella* and *Shigella* species (e.g., CIP, LVX) (n=0), all of which pose

increasing threats due to their resistance profiles and prevalence in healthcare settings. While for the medium-priority group, there is no study for *macrolide*-resistant *Streptococcus pneumoniae* and Group A *Streptococci* (e.g., ERY resistance) and *penicillin*-resistant GBS, which represents growing risks, particularly in low- and middle-income countries. These gaps highlight the need for further ML-driven research tailored to high-burden-resistant bacteria, aligning with global AMR mitigation efforts.

Summary. In recent years, AMR prediction has become increasingly dominant, drawing research focus away from other tasks, particularly TYP. Since 2022, it has been the primary focus of 68% of the reviewed studies. However, AMR predictions are predominantly limited to binary classifications (resistant/susceptible), overlooking distinctions between individual drugs. Despite the growing number of antibiotics analyzed, only three studies employ multilabel classification to predict multiple resistance profiles simultaneously, highlighting a key research gap. While many analyzed antibiotics align with the 2024 WHO BPPL, critical threats like CP-resistant *A. baumannii* remain unaddressed. Expanding multilabel approaches and broadening pathogen coverage could improve AMR prediction and better support global health priorities.

F. Exploratory data analysis and feature importance analysis

Exploratory Data Analysis (EDA) plays a crucial role in understanding the structure, distribution, and variability of the data and helps identify patterns, inconsistencies, and outliers. In fact, it can help us select appropriate data pre-processing and adequate ML algorithms. Visualization is an essential aspect of EDA, as it enables a more intuitive interpretation of complex data structures, making it easier to uncover meaningful relationships and trends.

A total of n=59 studies incorporate some form of EDA, with dimensionality reduction being the most common approach. Principal Component Analysis (PCA) is the most frequently used technique (n=25) [15], [17], [18], [20], [25], [27], [29], [30], [45], [50], [53], [64], [68], [78], [80], [81], [84], [88], [90], [91], [93], [97], [98], [103], [109], followed by t-Distributed Stochastic Neighbor Embedding (t-SNE) (n=5) [17], [86], [87], [95], [112]. While these techniques are often used for dimensionality reduction to 2 or 3 dimensions, their visual representations can also provide valuable insights for data interpretation. Clustering methods are used to group bacterial strains or resistance profiles based on similarity, providing valuable insights into potential subgroups within the dataset. Such techniques include Unsupervised Hierarchical Clustering Analysis (UHCA) (n=15) [17], [27], [40], [60], [61], [63], [67], [77], [78], [88], [90], [91], [93], [98], [131], dendrograms (n=9) [28], [41], [43], [46], [50], [68], [78], [90], [93], and k-means clustering (n=2) [77], [107], aiding in pattern recognition and data exploration.

When working with very high-dimensional data, feature importance analysis becomes essential for understanding how different variables contribute to predictions and overall model performance. Identifying the most influential features helps refine models, improve interpretability, and optimize computational efficiency. Model-agnostic SHapley Additive exPlanations (SHAP) values are frequently used (n=13) [21], [83], [85], [88], [89], [92], [97], [99], [101], [108], [109], [112], [113] to interpret the contribution of different features, offering transparency in model decision-making.

Summary. EDA is crucial for understanding data structure, identifying patterns, and informing pre-processing and model selection. Many reviewed studies apply EDA using dimensionality reduction techniques like PCA and t-SNE, as well as clustering methods such as UHCA and dendrograms, to reveal relationships within bacterial strains and resistance profiles. Feature importance analysis, particularly through SHAP values is usually applied to enhance interpretability. However, EDA remains underutilized in many studies, and its interpretation is often not straightforward. Adopting a more structured and standardized approach to EDA and feature importance analysis, along with clearer result interpretation, would improve data comprehension and support more informed decision-making in ML workflow design.

G. Pre-processing

Pre-processing is a critical step in MALDI-TOF MS data analysis, directly influencing ML model performance. The principle of "garbage in, garbage out" [132] strongly applies here, emphasizing the need for standardized workflows to ensure high-quality data input. Spectrometry data typically contains noise, such as baseline shifts and fluctuations, which require correction through baseline subtraction and smoothing techniques. Additionally, not all regions of the mass spectrum are informative; peak selection methods help isolate relevant signals (see Section III-H). Normalization is essential due to the wide intensity range in spectra, often spanning thousands of units. Furthermore, spectra acquired from different instruments (see Section III-B) exhibit variability, necessitating standardization to ensure robust, generalizable models. Since MALDI-TOF MS data can vary in length, binning techniques are often applied to create fixed-size inputs for ML models. Lastly, technical replicates—multiple spectra from the same sample—must be carefully handled to prevent duplicated data and potential bias.

Two widely used pre-processing tools are the open-source R package MALDIquant [115] (n=16), and ClinProTools [120] (n=20), a commercial software integrated into Bruker Daltonics' [117] systems. Both offer extensive pre-processing options, but the lack of standardization in parameter selection and processing order presents a major challenge, hindering reproducibility and comparability across studies. This variability results in diverse pre-processing pipelines, for example, *Weis et al.* [25] applied normalization, Savitzky-Golay smoothing, baseline correction, Total Ion Current (TIC) normalization, signal-to-noise ratio (SNR) thresholding, warping, and trimming, whereas *Dematheis et al.* [77] employed spectral cleaning, transformation, smoothing, baseline correction, and normalization.

Despite methodological discrepancies, several pre-processing steps are consistently applied across studies, including baseline subtraction (n=51), smoothing (n=41; e.g., Savitzky-Golay), normalization (n=41; e.g., TIC), SNR thresholding (n=23), spectra averaging (n=54), and binning (n=32). Additional techniques, such as scaling, spectra and peak alignment, recalibration, binarization, and transformations (e.g., wavelet or helix matrix), are less commonly reported. Table I provides a comprehensive overview of the principal steps applied and whether a study includes ablation or comparative studies, indicated by an asterisk (*).

Summary. The lack of standardization in MALDI-TOF MS pre-processing impacts data quality, model performance, and reproducibility. Studies vary widely in method selection, ordering, and hyperparameter choices, often without clear justification, making replication and optimization difficult. While tools like MALDIquant and ClinProTools provide automated workflows, their customization options introduce inconsistencies that hinder comparability. Some studies explore different pre-processing pipelines through ablation studies, but these remain too limited to establish universal best practices. Addressing these challenges through systematic evaluations and standardized pipelines is crucial for enhancing reproducibility and a key gap in the research that must be filled in to advance ML applications in MALDI-TOF MS analysis.

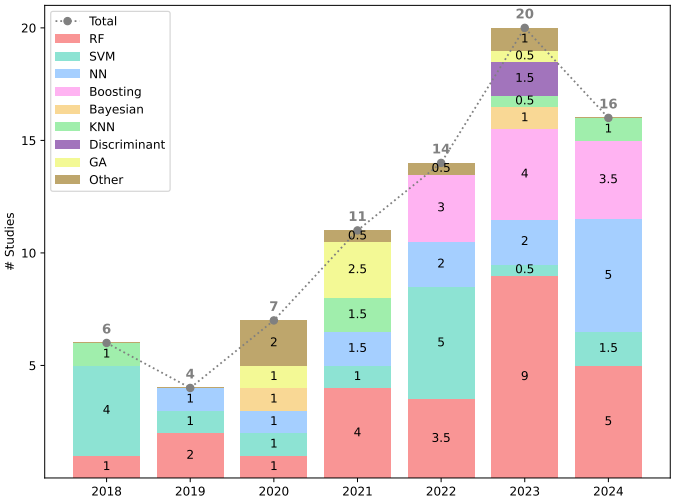

Fig. 3: Trend of the best-performing models across studies over time. Papers published before 2018 are not included, as the mean number of studies per year was below 2. In studies (n=14) reporting two models as equally best-performing (see **bold** in *Algorithm* in Table I), each model contributes 0.5 to the counts in this visualization.

H. Feature selection

Feature selection, or peak selection (PS) in this context, is a crucial step in handling high-dimensional datasets. In fact, most reviewed studies (n=68) employ PS techniques to enhance model interpretability, improve computational efficiency, and mitigate overfitting. The most common approach involves applying cutoffs, either based on SNR (n=23) or intensity thresholds (n=2) to eliminate minimally informative variables. Additionally, some studies use statistical methods (n=10), such as Pearson Correlation Coefficient, t-tests, or Analysis of Variance (ANOVA), while AUC-based selection (n=3) prioritizes features with the highest discriminative power.

ML-based PS is also widely adopted, with RF-based importance ranking being the most frequently applied technique (n=19). Genetic Algorithms (GA) (n=8) are explored as well, leveraging evolutionary strategies to optimize feature subsets. Moreover, several studies (n=22) utilize alternative PS methods, such as LASSO (n=2) and Bayesian Automatic Relevance Determination (ARD) analysis (n=1). For a comprehensive overview of the PS techniques used in each study, please refer to Table I.

Summary. PS is very helpful for optimizing MALDI-TOF MS-based ML models, improving interpretability, efficiency, and generalization. Advanced ML-driven approaches, such as RF-based importance ranking, Genetic Algorithms, and LASSO, are highly recommended. However, a key unresolved issue is whether MALDI-TOF data should be processed with a fixed input size, as spectra vary in length, and predefined PS may lead to information loss. This challenge is particularly relevant for ML models requiring fixed inputs, though DL techniques, especially Transformer models, offer new possibilities for handling variable-length spectra.

I. Algorithms

A wide range of algorithms (n=30) is applied across the reviewed studies, with most (n=67) comparing multiple models to assess performance. However, algorithm selection is generally driven by dataset size, quality, and computational constraints rather than the specific task (ID, AMR, or TYP), with minimal correlation between model choice and study objective.

Figure 3 illustrates the evolution of best-performing algorithms over time. The increasing use of ensemble models, such as RF and Boosting, along with the emergence of DL approaches—particularly after the release of the DRIAMS database—reflects a shift toward more complex, data-driven methodologies. Despite this, clinical applications often prioritize interpretability, making classical ML models like RF (n=29) and SVM (n=22) the most commonly used due to their robustness and ability to handle high-dimensional data. RF is favored for its ensemble structure, which enhances predictive accuracy and reduces overfitting, while SVMs remain popular in high-dimensional datasets where the number of features exceeds the number of samples [70], [72], [74], [76], [80], [82], [112], [113].

Boosting methods (n=13), including GB, AdaB [133], LGBM, and XGB [134], also demonstrate strong performance. Less common techniques include genetic algorithms (GA) (n=10), KNN (n=5), discriminant analysis (PLS-DA, LDA) (n=2), and probabilistic approaches (Gaussian processes, Factor Analysis) (n=2). Overall, classical ML models (n=59) outnumber DL approaches (n=19)—such as Multi-Layer Perceptrons (MLP), Convolutional Neural Networks (CNN), AutoEncoders (AE), and Long Short-Term Memory (LSTM) models. DL models are advantageous when applied to raw spectra, as they excel at extracting local patterns from mass spectrometry data. However, their effectiveness depends on large datasets [15], [62], [75], [108], with classical ML models often outperforming them when data is limited [28], [69], [77].

Only n=1 study [112] utilizes transformers, specifically for transfer learning, where a pre-trained bacterial ID model is fine-tuned for AMR prediction using a meta-learner. This approach integrates five base models (LR, RF, SVM, LGBM, and MLP), leveraging the latent embeddings of the transformer to achieve the highest performance. Additionally, n=2 studies use two-step models that combine different algorithms to enhance classification [16], [30]. For example, *Li et al.* [16] applied a Denoising AE for dimensionality reduction, feeding its latent space into an SVM.

Summary. Model selection in ML-based studies is influenced by dataset size, computational constraints, and study objectives, yet reviewed studies show little correlation between algorithm choice and task specificity. Given the high dimensionality of MALDI-TOF data ($\sim$20,000 features) and typically low data-to-feature ratios, classical ML models remain dominant due to their robust performance in small datasets. In contrast, DL models, which require large datasets for effective generalization, are less common but are expected to gain traction as data availability increases, particularly sequential models suited for complex spectral data [112], [135]. Future research should focus on refining algorithm selection strategies and integrating more sophisticated models to improve adaptability and performance as dataset sizes grow.

J. Software

Python is the predominant software for MALDI-TOF MS analysis, used in most reviewed studies (n=36) [136]. Its widespread adoption stems from its open-source nature, versatility, and extensive ML libraries, including scikit-learn [137], TensorFlow [138], and XGBoost [134]. Additionally, Python is more commonly associated with recent and complex studies, with approximately 64% of publications in the last two years employing this programming language.

Manufacturer-specific software is widely used (n=23), with Bruker's ClinProTools [117], [120] being the most prevalent (n=20) [20], [28], [30], [40], [42]–[49], [51], [54], [61], [64], [68], [69], [71], [106]. Other proprietary tools include Bionumerics [139] (n=2) [55], [78] and AMRQuest [140] (n=1) [79], both designed specifically for MALDI-TOF MS data analysis.

R [141] is used in a smaller subset of studies (n=15), particularly for studies utilizing MALDIquant for pre-processing (n=7) [31], [50], [60], [63], [72], [77], [90]. R is favored for its robust statistical analysis capabilities and specialized mass spectrometry tools.

Clover [142], a web-based MALDI-TOF MS analysis platform, is featured in n=9 studies [17], [29], [66], [80], [84], [88], [91], [98], [103]. It is particularly popular in clinical settings due to its ease of use, requiring minimal data science expertise. Clover provides a streamlined workflow with built-in pre-processing and ML; however, its proprietary nature limits flexibility compared to Python-based solutions, which offer greater customization and adaptability.

MATLAB [143], a proprietary software, is employed in some studies (n=6). It is valued for its strong matrix computation and visualization capabilities, though its high cost and closed-source nature limit broader adoption in the field.

Of the reviewed articles, only n=3 do not specify the software used, which hinders reproducibility.

Summary. More than half of the reviewed studies utilize Python and R, the most common programming languages in the field, facilitating flexible model deployment and open sharing through code repositories. Approximately 34% of studies rely on proprietary tools like ClinProTools and Clover as these are simple and easy to use for the clinical commmunity, which function as black-boxes with limited fine-tuning, implementation, and no support for DL models. Their rigidity restricts innovation and significantly hinders research progress. To advance the field, we encourage interdisciplinary collaboration, facilitating the adoption of more complex and powerful models by the clinical community.

K. Evaluation metrics

The reviewed studies employ various evaluation metrics to assess model performance. Accuracy is the most frequently reported metric (n=77), providing a general measure of correctly classified instances. However, given the potential class imbalances in AMR prediction and bacterial ID and TYP tasks, relying solely on accuracy can be misleading. To address this, sensitivity (recall) (n=58) and specificity (n=52) are commonly reported, measuring the model's ability to correctly identify positive and negative cases, respectively. Positive/negative predictive value (n=19)/(n=16) assesses the reliability of predictions, while precision (n=5)—which quantifies the proportion of correctly identified positives out of all predicted positives—is less frequently used but remains valuable when false positives are costly. These metrics are particularly critical in AMR tasks, where misclassifying resistant bacteria as susceptible (false negatives) can have serious clinical implications.

F1-score (n=16), which balances precision and recall, is especially useful for imbalanced datasets. AUROC (n=51) is also widely used, offering an aggregate measure of a model's discriminative ability across different thresholds. To further address class imbalance, the area under the precision-recall curve (AUPRC) (n=12) has gained popularity [15], [21], [25], [29], [73], [85], [88], [99], [101], [108], [112], [113], as it provides a more reliable assessment when the minority class is of primary interest (see Section III-C).

Less frequently reported metrics include confusion matrix analysis (n=6), Kappa (n=4), Youden's index (n=3), Matthews correlation coefficient (n=3), mean error (n=2), Jaccard index (n=1), Hamming loss (n=1), and G-measure (n=1).

Summary. While traditional metrics such as accuracy, sensitivity, and specificity remain dominant, the increasing adoption of AUROC, AUPRC, and F1-score reflects a growing awareness of the challenges posed by imbalanced datasets.

L. Model proceedings

1) Hyperparameter tuning: Hyperparameter tuning is crucial for ML model optimization, directly influencing predictive performance and generalizability. However, the extent of tuning varies significantly across the reviewed studies. A substantial number (n=23) do not disclose any details on hyperparameter selection, leaving it unclear whether parameters were optimized or left at default values. Additionally, n=30 rely entirely on default settings.

Among studies that perform hyperparameter optimization, CV is the most common approach (n=33). The CV method optimizes hyperparameters by creating training and validation partitions within the training set, helping to mitigate overfitting and ensure robust model selection. Only n=2 studies apply Bayesian hyperparameter selection [109], [112].

A smaller subset (n=4) employs manual ablation studies [14], [38], [54], [100], though this approach is often inefficient and lacks reproducibility.

2) Evaluation strategy: Evaluating model performance is a crucial step in ML workflows, ensuring that predictive models generalize well beyond their training data. The reviewed studies employ a variety of evaluation strategies, with only one study not specifying its approach (n=1) [58].

Predefined train-test splits are used in n=48 studies. In these cases, a fixed portion of the dataset is allocated for training, while the remaining data is reserved for posterior evaluation. This approach is computationally efficient but may introduce bias if the split is not representative of the full data distribution. N=21 also apply CV on the predefined training set to improve model robustness.

Outer CV is the most common evaluation method, applied in n=60 studies. This technique repeatedly splits the dataset into training and validation folds, reducing overfitting and providing a more reliable estimate of model performance. Within this group, a subset of studies (n=4) employs stratified CV, which ensures class distribution is preserved across folds—a particularly valuable approach for imbalanced datasets [25], [59], [84], [86]. In contrast, n=6 studies rely on Leave-One-Out (LOO) CV for model validation [19], [38], [40], [50], [56], [98]. This method iteratively trains the model on all but one sample, testing it on the excluded instance until every sample has been used for validation. While LOO-CV provides an almost unbiased estimate of model performance, it comes at a high computational cost and suffers from increased variance due to strong correlations between training sets. Conversely, lower-fold CV methods reduce variance by training on more distinct subsets but may introduce higher bias when training sets are too small. Given this trade-off, LOO-CV is only justified for very small datasets.

OOD validation, or external validation, refers to the process of evaluating a ML model on an independent dataset that was not used during training or outer CV. This step is crucial for assessing a model's ability to generalize beyond the data on which it was developed, ensuring its robustness when applied to new or unseen distributions. However, despite its importance, OOD is often overlooked or misunderstood. Many studies conflate OOD validation with standard evaluation methods such as a predefined test split, which only assesses performance within the same dataset. Among the reviewed studies, only n=25 actually conduct OOD [15], [17], [21], [40], [55], [61], [65]–[68], [70], [75], [81], [83], [85], [87], [88], [90], [91], [97], [100], [108], [109], [112], [113]. The remaining studies rely solely on internal validation.

Summary. Hyperparameter tuning is essential for optimizing model performance, yet many studies either omit details or rely on default values. Regarding evaluation strategy, while fixed train-test splits remain common, relying solely on this approach is discouraged. More rigorous evaluation methods, particularly outer CV, are increasingly

TABLE IV: Summary of studies that provide open-source datasets and/or code. The table lists the available datasets, their sizes, formats (raw or preprocessed, pp), and whether the corresponding code is publicly accessible. Additionally, it includes the software used in each study and relevant repository links (e.g., GitHub, Zenodo, DRYAD).

Paper	Database	Size	Format	Code	Software	Links
Asakura (2018) [56]	✓	N/A	pp	✓	C++/R	GitHub
Desaire (2020) [14]	✓	571	raw			UCI
Ling (2020) [62]				✓	Python	GitHub
Weis (2020) [25]				✓	Python	GitHub
Weis (2022) [21]	✓	132,097	raw, pp	✓	Python	DRYAD, Github
Wang (2022) [75]				✓	Python	GitHub
Yu (2022) [83]	✓	2,456	pp			Zenodo
Wang (2023) [85]	✓	31,807	pp	✓	R	Github
G.-López (2023) [86]	✓	2,000	raw, pp	✓	Python	Github
Candela (2023) [17]	✓	1,566	raw			Clover
Chung (2023) [87]	✓	37,918	pp			Github
Yu (2023) [89]	✓	1,853	pp			Zenodo
Zintgraff (2023) [90]	✓	193	raw	✓	R	Mendeley
Wang (2023) [93]	✓	205	pp	✓	Python	Article
De Waele (2024) [102]				✓	Python	Github
Mao (2024) [18]	✓	N/A	pp			SimTK
L.-Cortés (2024) [108]				✓	Python	Github
Astudillo (2024) [109]				✓	Python	Github
Nguyen (2024) [112]	✓	30	raw	✓	Python	Github
Ren (2024) [113]				✓	Python	Github
RKI dataset [35]	✓	11,038	raw			Zenodo
Sum:	**14**			**14**		

adopted to improve reliability and reduce overfitting, though LOO-CV remains impractical for larger datasets due to its computational cost. A major limitation is the frequent misinterpretation of external validation, with many studies failing to perform true OOD validation, which is critical for assessing a model's generalizability beyond its training dataset. Without rigorous OOD testing, models risk overfitting and failing in real-world applications. To promote the implementation of generalizable models useful for the clinical community, researchers should prioritize standardized evaluation strategies, transparent reporting, and proper OOD evaluation practices.

M. Open-source data and code

The availability of open-source data and code is essential for advancing research, promoting transparency, and ensuring reproducibility in ML studies. However, relatively few studies in our review provide public access to their datasets or make their code available on platforms like GitHub [144] or other public repositories.

Among the reviewed studies, only a small subset publicly shares their datasets (n=13) and (n=14) share their ML implementation, which is crucial for full reproducibility. Despite the growing availability of MALDI-TOF MS datasets, only two contain data labeled for TYP [90], [93]. The absence of open-access typed data restricts researchers' ability to develop and validate models for hypervirulence prediction across bacterial strains.

Among the open datasets, file formats vary considerably, with few studies (n=2) providing raw spectral data (*BrukerRaw* format) [17], [90]. In contrast, others release raw or pre-processed data in the non-original format (*csv*, *txt*, *npy*), limiting its usability for other research purposes. As shown in Table IV, only a few datasets retain the original metadata, such as acquisition time or device information, which is crucial for standardization and reproducibility.

Two large public MALDI-TOF MS datasets are available to the research community: the RKI [35] and the DRIAMS [34] dataset published by *Weis et al.* [21]. Both provide a comprehensive collection of spectra from various bacterial species, enabling large-scale studies on bacterial ID and AMR. These repositories have facilitated replicating and validating existing models and spurred new research. A total of n=7 studies use the DRIAMS dataset for their research [94], [101], [102], [108], [109], [111], [113], illustrating the transformative impact of open-access data on advancing the field.

Summary. Despite increasing efforts toward open science, most datasets remain inaccessible due to privacy concerns and institutional restrictions. This lack of open-source resources hinders reproducibility, limits collaboration, and slows the advancement of ML applications in MALDI-TOF MS. The absence of large, standardized repositories remains a major gap. Moreover, open-source code should be well-documented and structured to facilitate accessibility for both clinical and technical researchers. Prioritizing transparency and open access will drive progress, accelerate discovery, and strengthen the clinical applicability of ML-driven microbiology.

IV. Challenges and Future Directions

A crucial step in advancing ML applications in microbiology is the establishment of standardized, open-access databases through collaboration between research groups. MALDI-TOF mass spectra are inherently anonymizable, making their publication more feasible compared to other clinical data types. Routine procedures like disk diffusion already provide AMR labels with minimal additional effort for ML training. In contrast, TYP approaches require genetic sequencing, which is labor-intensive and costly, making dataset availability a major bottleneck. Sharing labeled datasets could accelerate research, particularly in hypervirulent and nosocomial strains. Additionally, publicly releasing outbreak-related pathogen data would enhance early detection and rapid response efforts, aligning with global health initiatives such as the United Nations' 2030 Agenda goal of ensuring good health and well-being [145].

Large-scale shared databases could drive the development of multi-task ML models, enabling simultaneous bacterial ID and AMR prediction or even AMR and TYP analysis. These advanced solutions would be highly valuable in clinical settings. A major limitation in current research is the lack of multi-task models and large-scale antibiotic screening, as no study has yet successfully integrated bacterial identification with comprehensive AMR profiling. Developing such methodologies would significantly improve infection management, supporting the transition to personalized medicine.

Additionally, collaborative databases would enhance spectral diversity, since by incorporating data from different machines, geographical locations, and laboratory protocols, models could be trained to generalize better and make predictions independent of data origin. Additionally, access to multi-source data would facilitate robust model evaluation using OOD data, ensuring fairer assessments.

To achieve a real transference of ML techniques to microbiology close interdisciplinary collaboration between ML experts and clinical specialists is required, including biologists and clinicians. By working together, they can identify existing gaps, assess the feasibility of technological solutions, and define the necessary requirements for their implementation in clinical practice. Beyond database creation, interdisciplinary collaboration between ML and clinical experts is crucial for ensuring practical and impactful ML applications. A significant number of studies are conducted solely by clinical research groups using basic tools or proprietary software, such as Clover or ClinProTools, which limit flexibility, fine-tuning, and innovation. The lack of ML expertise in such studies prevents the use of cutting-edge

technologies and constrains the development of more sophisticated and scalable solutions.

We advocate for research initiatives that integrate both clinical and ML expertise to foster innovation. From our experience, we next propose a workflow to help improve research quality, reproducibility, and an adequate application of ML-driven solutions in microbiology.

A. Guidelines

After reviewing n=93 SOTA papers, we identified a lack of standardization in ML research on MALDI-TOF MS data. To address this, we propose a set of guidelines to establish a well-structured research pipeline that can effectively advance ML applications in this field. A clear definition of the research task is crucial, specifying how data will be utilized and how ML techniques will be applied. Once this is defined, the pipeline should follow key stages: (i) EDA, (ii) spectra pre-processing and peak selection (PS), (iii) model selection, (iv) training, and (v) evaluation, as in Figure 4.

During the EDA, a comprehensive dataset analysis is crucial for making informed pre-processing and model selection decisions. One of the first aspects to consider is the number of unique bacterial isolates, as this directly influences model selection. Additionally, handling spectral replicates is essential, and there are two primary approaches: (1) either peak alignment and averaging spectra to create a single representation per isolate or (2) treating them as a multi-instance problem. The latter approach is preferable, as it maximizes data usage and improves reproducibility, but care must be taken to ensure that replicates remain within the same split in both inner and outer CV to prevent data leakage. Another key factor is identifying data imbalance, which has a direct impact on loss functions, and evaluation strategies. Finally, data visualization using unsupervised learning techniques, e.g. PCA or UHCA, can enhance interpretability and pattern recognition. However, these methods should be reserved for exploratory analysis rather than classification tasks, which require supervised learning models.

Pre-processing plays a fundamental role in standardizing datasets for better model learning. Although current SOTA practices lack consistency, we recommend several essential steps for raw spectral data: (i) variance stabilization using log2 or square root transformation, (ii) smoothing with a Savitzky-Golay filter, (iii) baseline removal with techniques such as the Statistics-sensitive Non-linear Iterative Peak-clipping algorithm (SNIP) or top-hat transformation, (iv) normalization of the intensities, and (v) trimming within the range of 2,000 to 20,000 Da to eliminate noise. After this, spectra can either be used in their original varying lengths or standardized. A straightforward approach is binning with a 3 Da bin size, converting spectra into 6,000-element vectors by summing intensities within each bin. Alternatively, PS techniques, e.g. based on RF, can be employed to retain the most relevant features. For more advanced implementations, models such as Transformers [112] can directly handle variable-length spectra.

Choosing the right model depends on the research question and the characteristics of the dataset. For small to medium-sized datasets, classical ML models such as RF and LGBM generally provide strong performance. Bayesian models, such as GPs [25] or FA [86], are particularly useful for combining multimodal data, as they generalize well and do not require an inner CV. In contrast, larger datasets allow for the use of DL models [21], [112], which are highly effective at capturing complex representations. Before model training, data imbalance must be addressed, which can be done through either (1) oversampling techniques like SMOTE [111], [112] or by (2) incorporating imbalance-aware loss functions, such as weighted Log-Loss for logistic regression, weighted Hinge Loss for SVMs, or class-weight adjustments in RF models. Hyperparameter tuning is also a key step and should be performed using inner CV, calculating the mean and standard deviation of evaluation metrics to identify the most robust model. As in previous steps, spectral replicates must remain within the same split to prevent bias.

Model evaluation should follow training using outer CV, where the dataset is split into k folds (typically 3, 5, or 10), allowing the model to be trained and tested k times. Once the best-performing model is identified, it should be retrained using the entire internal dataset. To assess generalizability, an OOD evaluation should be performed using a completely independent dataset, ideally from a different machine or hospital. Evaluation metrics should address class imbalance, with balanced accuracy as a priority, complemented by AUC-ROC, precision, recall, and AUPRC for a comprehensive performance assessment.

V. Conclusions

The integration of ML into MALDI-TOF MS analysis has significantly advanced bacterial ID, AMR prediction, and strain TYP. However, several challenges remain, including data scarcity, insufficient transparency in sharing study details, non-standardized pre-processing and training workflows, inconsistencies in model evaluation protocols, and limited access to open-source datasets and code. While classical ML models remain the predominant approach, there is growing interest in DL methods but they require larger datasets to provide accurate solutions. Despite these advancements, issues related to reproducibility and generalizability continue to hinder clinical use.

We hope this review not only provides a comprehensive overview of the current landscape but also serves as a guide toward more standardized practices in ML-based MALDI-TOF research. By promoting open science and facilitating external validation across research groups, we aim to support the development of robust and clinically relevant models. Furthermore, we seek to inspire future efforts to fully harness the potential of ML in MALDI-TOF analysis, promote interdisciplinary work between data scientists and clinical experts, and ultimately drive its broader adoption in clinical microbiology and enhance diagnostic accuracy and patient outcomes.

References

[1] S. S. Arbefeville *et al.*, "Evolving strategies in microbe identification—a comprehensive review of biochemical, maldi-tof ms and molecular testing methods," *J of Antimicrobial Chemotherapy*, vol. 79, no. Supplement_1, pp. i2–i8, 2024.

[2] L. M. Weiner-Lastinger *et al.*, "Antimicrobial-resistant pathogens associated with adult healthcare-associated infections: summary of data reported to the National Healthcare Safety Network, 2015-2017," *Infection Control & Hospital Epidemiology*, vol. 41, no. 1, pp. 1–18, 2020.

[3] WHO, *Antimicrobial Resistance*, Accessed: February 10, 2025, 2023.

[4] L. B. Reller *et al.*, "Antimicrobial susceptibility testing: a review of general principles and contemporary practices," *Clinical infectious diseases*, vol. 49, no. 11, pp. 1749–1755, 2009.

[5] C. V. Weis *et al.*, "Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review," *Clin Microbiol Infect*, vol. 26, no. 10, pp. 1310–1317, 2020.

[6] F. Hillenkamp *et al.*, "Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers," *Anal Chem*, vol. 57, no. 14, pp. 2299–2304, 1985.

Fig. 4: Overview of key steps in ML-based MALDI-TOF MS analysis, covering exploratory data analysis (EDA), preprocessing, model selection, training, and evaluation. The diagram highlights critical considerations such as data imbalance handling, hyperparameter tuning, and cross-validation (CV). Recommended steps are indicated in **black**, steps with multiple valid approaches are shown in blue, and optional steps are marked in green.

[7] A. Tran *et al.*, "Cost savings realized by implementation of routine microbiological identification by matrix-assisted laser desorption ionization-time of flight mass spectrometry," *J Clin Microbiol*, vol. 53, no. 8, pp. 2473–2479, 2015.

[8] F. Nomura, "Proteome-based bacterial identification using matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS): A revolutionary shift in clinical diagnostic microbiology," *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, vol. 1854, no. 6, pp. 528–537, 2015.

[9] N. M. B. Rodrigues *et al.*, "The Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS) identification versus biochemical tests: a study with enterobacteria from a dairy cattle environment," *Brazilian J of Microbiol*, vol. 48, no. 1, pp. 132–138, 2017.

[10] N. Dhiman *et al.*, "Performance and cost analysis of matrix-assisted laser desorption ionization-time of flight mass spectrometry for routine identification of yeast," *J Clin Microbiol*, vol. 49, no. 4, pp. 1614–1616, 2011.

[11] R. Holland *et al.*, "Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 10, no. 10, pp. 1227–1232, 1996.

[12] C. Fenselau and P. A. Demirev, "Characterization of intact microorganisms by MALDI mass spectrometry," *Mass Spectrometry Reviews*, vol. 20, no. 4, pp. 157–171, 2001.

[13] R. Giebel *et al.*, "Microbial fingerprinting using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS): applications and challenges," *Advances in Applied Microbiol*, vol. 71, pp. 149–184, 2010.

[14] H. Desaire and D. Hua, "The Aristotle Classifier Provides Accurate Identifications of Highly Similar Bacteria Analyzed by MALDI-TOF MS," *Anal Chem*, Nov. 2020.

[15] T. Mortier *et al.*, "Bacterial species identification using MALDI-TOF mass spectrometry and machine learning techniques: A large-scale benchmarking study," *Comput. Struct. Biotechnol. J*, vol. 19, Nov. 2021.

[16] Y. Li *et al.*, "Accurate classification of Listeria species by MALDI-TOF mass spectrometry incorporating denoising autoencoder and machine learning," *J Microbiol Methods*, vol. 192, 2022.

[17] A. Candela *et al.*, "Automatic Discrimination of Species within the Enterobacter cloacae Complex Using Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry and Supervised Algorithms," *J Clin Microbiol*, vol. 61, no. 4, 2023.

[18] Q. Mao *et al.*, "Identification of Escherichia coli strains using MALDI-TOF MS combined with long short-term memory neural networks," *Aging*, vol. 16, Jun. 2024.

[19] T.-S. Huang *et al.*, "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*, vol. 15, Feb. 2020.

[20] A. Calderaro *et al.*, "Rapid Identification of Escherichia coli Colistin-Resistant Strains by MALDI-TOF Mass Spectrometry," *Microorganisms*, vol. 9, Oct. 2021.

[21] C. Weis *et al.*, "Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning," *Nat Med*, vol. 28, Jan. 2022.

[22] Y. Zeng *et al.*, "Machine learning model of imipenem-resistant Klebsiella pneumoniae based on MALDI-TOF-MS

platform: An observational study," *Health Sci Rep*, vol. 6, no. 9, e1108, 2023.

[23] T.-H. Lin *et al.*, "Innovative strategies against superbugs: Developing an AI-CDSS for precise Stenotrophomonas maltophilia treatment," *J Glob Antimicrob Resist*, vol. 38, pp. 173–180, 2024.

[24] Y. Huang and H. Desaire, "Calibration-free quantification of antibiotic resistance markers using machine learning on MALDI-TOF data," *Anal Chem*, vol. 92, no. 2, pp. 1050–1057, 2020.

[25] C. Weis *et al.*, "Topological and kernel-based microbial phenotype prediction from MALDI-TOF mass spectra," *Bioinformatics*, vol. 36, no. Supplement1, Jul. 2020.

[26] L. Qiao, "MALDI-TOF MS for pathogenic bacteria analysis," *Int J Mass Spectrom*, vol. 482, p. 116 935, 2022.

[27] I. Pena *et al.*, "Comparison of performance of MALDI-TOF MS and MLST for biotyping carbapenemase-producing Klebsiella pneumoniae sequence types ST11 and ST101 isolates," *Enferm. Infecc. Microbiol Clin.*, vol. 40, Dec. 2020.

[28] A. Calderaro *et al.*, "Rapid Classification of Clostridioides difficile Strains Using MALDI-TOF MS Peak-Based Assay in Comparison with PCR-Ribotyping," *Microorganisms*, vol. 9, Mar. 2021.

[29] A. Abdrabou *et al.*, "Discrimination between hypervirulent and non-hypervirulent ribotypes of Clostridioides difficile by MALDI-TOF mass spectrometry and machine learning," *Eur. J Clin. Microbiol Infect. Dis.*, Sep. 2023.

[30] E. Manfredi *et al.*, "Rapid and accurate detection of Shiga toxin-producing Escherichia coli (STEC) serotype O157 : H7 by mass spectrometry directly from the isolate, using 10 potential biomarker peaks and machine learning predictive models," *J of medical microbiology*, vol. 72, May 2023.

[31] A. Godmer *et al.*, "Contribution of MALDI-TOF mass spectrometry and machine learning including deep learning techniques for the detection of virulence factors of Clostridioides difficile strains," *Microb Biotechnol*, vol. 17, no. 6, 2024.

[32] D. Sharma *et al.*, "phylaGAN: data augmentation through conditional GANs and autoencoders for improving disease prediction accuracy using microbiome data," *Bioinformatics*, vol. 40, no. 4, btae161, 2024.

[33] J. Schneider, "Explainable generative ai (genxai): A survey, conceptualization, and research agenda," *Artificial Intelligence Review*, vol. 57, no. 11, p. 289, 2024.

[34] C. Weis *et al.*, *DRIAMS: Database of Resistance Information on Antimicrobials and MALDI-TOF Mass Spectra*, Dryad, 2022.

[35] P. Lasch *et al.*, *A MALDI-TOF Mass Spectrometry Database for Identification and Classification of Highly Pathogenic Microorganisms from the Robert Koch-Institute (RKI)*, Zenodo, Oct. 2016.

[36] X. A. López-Cortés *et al.*, "Machine learning and matrix-assisted laser desorption/ionization time-of-flight mass spectra for antimicrobial resistance prediction: A systematic review of recent advancements and future development," *J Chromatogr A*, p. 465 262, 2024.

[37] F. Pennisi *et al.*, "Artificial intelligence in antimicrobial stewardship: a systematic review and meta-analysis of predictive performance and diagnostic accuracy," *European J Clin Microbiol & Infectious Diseases*, pp. 1–51, 2025.

[38] Z. Zhang *et al.*, "Forward selection radial basis function networks applied to bacterial classification based on MALDI-TOF-MS," *Talanta*, vol. 63, no. 3, pp. 527–532, 2004.

[39] G. A. Satten *et al.*, "Standardization and denoising algorithms for mass spectra to classify whole-organism bacterial specimens," *Bioinformatics*, vol. 20, no. 17, pp. 3128–3136, 2004.

[40] S.-Y. Hsieh *et al.*, "Highly efficient classification and identification of human pathogenic bacteria by MALDI-TOF MS," *Molecular & cellular proteomics*, vol. 7, no. 2, pp. 448–456, 2008.

[41] R. Schaumann *et al.*, "A step towards the discrimination of beta-lactamase-producing clinical isolates of Enterobacteriaceae and Pseudomonas aeruginosa by MALDI-TOF mass spectrometry," *Medical science monitor: international medical journal of experimental and clinical research*, vol. 18, no. 9, MT71, 2012.

[42] P. D. Khot and M. A. Fisher, "Novel approach for differentiating Shigella species and Escherichia coli by matrix-assisted laser desorption ionization-time of flight mass spectrometry," *J Clin Microbiol*, vol. 51, no. 11, pp. 3711–3716, 2013.

[43] S. Nakano *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*, vol. 44, no. 3, pp. 256–259, 2014.

[44] B. Li *et al.*, "Matrix-assisted laser desorption ionization: time of flight mass spectrometry-identified models for detection of ESBL-producing bacterial strains," *Medical Science Monitor Basic Research*, vol. 20, p. 176, 2014.

[45] D. Xiao *et al.*, "Novel strategy for typing Mycoplasma pneumoniae isolates by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry coupled with ClinProTools," *J Clin Microbiol*, vol. 52, no. 8, pp. 3038–3043, 2014.

[46] T. Zhang *et al.*, "Analysis of methicillin-resistant Staphylococcus aureus major clonal lineages by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI–TOF MS)," *J Microbiol Methods*, vol. 117, pp. 122–127, 2015.

[47] S. Nakano *et al.*, "Development and evaluation of MALDI-TOF MS-based serotyping for Streptococcus pneumoniae," *Eur J Clin Microbiol Infect Dis*, vol. 34, pp. 2191–2198, 2015.

[48] J. H. Chen *et al.*, "Use of MALDI Biotyper plus ClinProTools mass spectra analysis for correct identification of Streptococcus pneumoniae and Streptococcus mitis/oralis," *J of Clinical Pathology*, vol. 68, no. 8, pp. 652–656, 2015.

[49] M. Camoez *et al.*, "Automated categorization of methicillin-resistant Staphylococcus aureus clinical isolates into different clonal complexes by MALDI-TOF mass spectrometry," *Clin Microbiol Infect*, vol. 22, no. 2, 161–e1, 2016.

[50] C. A. Mather *et al.*, "Rapid detection of vancomycin-intermediate Staphylococcus aureus by matrix-assisted laser desorption ionization-time of flight mass spectrometry," *J Clin Microbiol*, vol. 54, no. 4, pp. 883–890, 2016.

[51] P.-L. Ho *et al.*, "Rapid detection of cfiA metallo-β-lactamase-producing Bacteroides fragilis by the combination of MALDI-TOF MS and CarbaNP," *J of Clinical Pathology*, vol. 70, no. 10, pp. 868–873, 2017.

[52] K. Sogoawa *et al.*, "Rapid Staphylococcus aureus discrimination between methicillin-sensitive and methicillin-resistant using MALDI-TOF mass spectrometry," *Biocontrol Sci*, vol. 22, no. 163, e9, 2017.

[53] H.-Y. Wang *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*, vol. 13, no. 3, e0194289, 2018.

[54] H.-Y. Wang *et al.*, "Application of a MALDI-TOF analysis platform (ClinProTools) for rapid and preliminary report of MRSA sequence types in Taiwan," *PeerJ*, vol. 6, e5784, 2018.

[55] C. Rodrigues *et al.*, "Identification of Klebsiella pneumoniae, Klebsiella quasipneumoniae, Klebsiella variicola and related phylogroups by MALDI-TOF mass spectrometry," *Frontiers in microbiology*, vol. 9, p. 3000, 2018.

[56] K. Asakura *et al.*, "Rapid and easy detection of low-level resistance to vancomycin in methicillin-resistant Staphylococcus aureus by matrix-assisted laser desorption ionization time-of-flight mass spectrometry," *PloS one*, vol. 13, no. 3, e0194212, 2018.

[57] H.-Y. Wang *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," *Frontiers in Microbiol*, vol. 9, p. 2393, 2018.

[58] J. Ling *et al.*, "A novel short-term high-lactose culture approach combined with a matrix-assisted laser desorption ionization-time of flight mass spectrometry assay for differentiating Escherichia coli and Shigella species using artificial neural networks," *PLoS One*, vol. 14, no. 10, e0222636, 2019.

[59] C.-R. Chung *et al.*, "Incorporating statistical test and machine intelligence into strain typing of staphylococcus haemolyticus based on matrix-assisted laser desorption ionization-time of flight mass spectrometry," *Frontiers in microbiology*, vol. 10, p. 2120, 2019.

[60] W. Tang *et al.*, "MALDI-TOF mass spectrometry on intact bacteria combined with a refined analysis framework allows accurate classification of MSSA and MRSA," *PloS one*, vol. 14, no. 6, e0218951, 2019.

[61] H.-Y. Wang *et al.*, "Rapid classification of group B Streptococcus serotypes based on matrix-assisted laser desorption ionization-time of flight mass spectrometry and machine learning techniques," *BMC bioinformatics*, vol. 20, pp. 1–17, 2019.

[62] J. Ling *et al.*, "Helix Matrix Transformation Combined With Convolutional Neural Network Algorithm for Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry-Based Bacterial Identification," *Frontiers in Microbiol*, vol. 11, Nov. 2020.

[63] S. Kann *et al.*, "MALDI-TOF mass spectrometry for subtyping of Streptococcus pneumoniae," *BMC microbiology*, vol. 20, pp. 1–9, 2020.

[64] J. Wei *et al.*, "Novel Strategy for Rapidly and Safely Distinguishing Bacillus anthracis and Bacillus cereus by Use of Peptide Mass Fingerprints Based on Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry," *J Clin Microbiol*, vol. 59, no. 1, 2020.

[65] H.-Y. Wang *et al.*, "A large-scale investigation and identification of methicillin-resistant Staphylococcus aureus based on peaks binning of matrix-assisted laser desorption ionization-time of flight MS spectra," *Brief Bioinform*, vol. 22, no. 3, Jul. 2021.

[66] E. Gato *et al.*, "An improved matrix-assisted laser desorption ionization-time of flight mass spectrometry data analysis pipeline for the identification of carbapenemase-producing Klebsiella pneumoniae," *J Clin Microbiol*, vol. 59, no. 7, pp. 10–1128, 2021.

[67] H.-Y. Wang *et al.*, "Clinically Applicable System for Rapidly Predicting Enterococcus faecium Susceptibility to Vancomycin," *Microbiol Spectr*, vol. 9, no. 3, 2021.

[68] S. Fiamanya *et al.*, "Exploring the value of MALDI-TOF MS for the detection of clonal outbreaks of Burkholderia contaminans," *J Microbiol Methods*, vol. 181, p. 106 130, 2021.

[69] K. Tang *et al.*, "MALDI-TOF MS platform combined with machine learning to establish a model for rapid identification of methicillin-resistant Staphylococcus aureus," *J Microbiol Methods*, vol. 180, 2021.

[70] C.-R. Chung *et al.*, "MDRSA: A Web Based-Tool for Rapid Identification of Multidrug Resistant Staphylococcus aureus Based on Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry," *Frontiers in Microbiol*, vol. 12, Dec. 2021.

[71] L. Huang *et al.*, "Rapid Classification of Multilocus Sequence Subtype for Group B Streptococcus Based on MALDI-TOF Mass Spectrometry and Statistical Models," *Frontiers in Cellular and Infection Microbiol*, vol. 10, Jan. 2021.

[72] X. Liu *et al.*, "Rapid identification and discrimination of methicillin-resistant strains via matrix-assisted laser desorption/ionization time-of-flight mass spectrometry," *Rapid Commun. Mass Spectrom.*, vol. 35, no. 2, 2021.

[73] M. Feucherolles *et al.*, "Combination of MALDI-TOF Mass Spectrometry and Machine Learning for Rapid Antimicrobial Resistance Screening: The Case of Campylobacter spp," *Frontiers in Microbiol*, vol. 12, Feb. 2022.

[74] P.-H. Kong *et al.*, "Discrimination of Methicillin-resistant Staphylococcus aureus by MALDI-TOF Mass Spectrometry with Machine Learning Techniques in Patients with Staphylococcus aureus Bacteremia," *Pathogens*, vol. 11, May 2022.

[75] H.-Y. Wang *et al.*, "Efficiently Predicting Vancomycin Resistance of Enterococcus Faecium From MALDI-TOF MS Spectra Using a Deep Learning-Based Approach," *Frontiers in Microbiol*, vol. 13, Jun. 2022.

[76] C. Wang *et al.*, "Large-Scale Samples Based Rapid Detection of Ciprofloxacin Resistance in Klebsiella pneumoniae Using Machine Learning Methods," *Frontiers in Microbiol*, vol. 13, Mar. 2022.

[77] F. Dematheis *et al.*, "Machine Learning Algorithms for Classification of MALDI-TOF MS Spectra from Phylogenetically Closely Related Species Brucella melitensis, Brucella abortus and Brucella suis," *Microorganisms*, vol. 10, Aug. 2022.

[78] M. J. van den Beld *et al.*, "MALDI-TOF MS using a custom-made database, biomarker assignment, or mathematical classifiers does not differentiate Shigella spp. and Escherichia coli," *Microorganisms*, vol. 10, no. 2, p. 435, 2022.

[79] K. Jeon *et al.*, "Performance of a Machine Learning-Based Methicillin Resistance of Staphylococcus aureus Identification System using MALDI-TOF MS and Comparison of the Accuracy According to SCCmec Types," *Microorganisms*, vol. 10, Sep. 2022.

[80] A. Candela *et al.*, "Rapid and Reproducible MALDI-TOF-Based Method for the Detection of Vancomycin-Resistant Enterococcus faecium Using Classifying Algorithms," *Diagnostics*, vol. 12, Jan. 2022.

[81] J. Zhang *et al.*, "Rapid Antibiotic Resistance Serial Prediction in Staphylococcus aureus Based on Large-Scale MALDI-TOF Data by Applying XGBoost in Multi-Label Learning," *Frontiers in Microbiol*, vol. 13, Apr. 2022.

[82] J. Wang *et al.*, "Rapid Detection of Carbapenem-Resistant Klebsiella pneumoniae Using Machine Learning and

MALDI-TOF MS Platform," *Infection and Drug Resistance*, vol. Volume 15, Jul. 2022.

[83] J. Yu *et al.*, "Rapid Identification of Methicillin-Resistant Staphylococcus aureus Using MALDI-TOF MS and Machine Learning from over 20,000 Clinical Isolates," *Microbiol Spectr*, vol. 10, Mar. 2022.

[84] á. Rodríguez-Villodres *et al.*, "Ultrasensitive and rapid identification of ESRI developer- and piperacillin/tazobactam-resistant Escherichia coli by the MALDIpiptaz test," *Emerging Microbes & Infections*, vol. 11, Aug. 2022.

[85] Z. Wang *et al.*, "A risk assessment framework for multidrug-resistant Staphylococcus aureus using machine learning and mass spectrometry technology," *Brief Bioinform*, vol. 24, no. 6, Sep. 2023.

[86] A. Guerrero-López *et al.*, "Automatic antibiotic resistance prediction in Klebsiella pneumoniae based on MALDI-TOF mass spectra," *Engineering Applications of Artificial Intelligence*, vol. 118, p. 105 644, 2023.

[87] C.-R. Chung *et al.*, "Data-Driven Two-Stage Framework for Identification and Characterization of Different Antibiotic-Resistant Escherichia coli Isolates Based on Mass Spectrometry Data," *Microbiol Spectr*, vol. 11, Apr. 2023.

[88] E. Gato *et al.*, "Direct Detection of Carbapenemase-Producing Klebsiella pneumoniae by MALDI-TOF Analysis of Full Spectra Applying Machine Learning," *J Clin Microbiol*, vol. 61, May 2023.

[89] "Direct prediction of carbapenem-resistant, carbapenemase-producing, and colistin-resistant Klebsiella pneumoniae isolates from routine MALDI-TOF mass spectra using machine learning and outcome evaluation," *Int. J Antimicrob. Agents*, vol. 61, no. 6,

[90] J. Zintgraff *et al.*, "Exploring Streptococcus pneumoniae capsular typing through MALDI-TOF mass spectrometry and machine-learning algorithms in Argentina: Identifying prevalent NON PCV13 serotypes alongside PCV13 serotypes," *J Mass Spectrom Adv Clin Lab*, vol. 30, 2023.

[91] D. Rodríguez-Temporal *et al.*, "Identification of Mycobacterium abscessus Subspecies by MALDI-TOF Mass Spectrometry and Machine Learning," *J Clin Microbiol*, vol. 61, no. 1, 2023.

[92] J. Yu *et al.*, "Prediction of methicillin-resistant Staphylococcus aureus and carbapenem-resistant Klebsiella pneumoniae from flagged blood cultures by combining rapid Sepsityper MALDI-TOF mass spectra with machine learning," *Int J Antimicrob Agents*, vol. 62, Oct. 2023.

[93] Z. Wang *et al.*, "Promising potential of machine learning-assisted MALDI-TOF MS as an effective detector for Streptococcus suis serotype 2 and virulence thereof," *Appl Environ Microbiol*, vol. 89, no. 11, 2023.

[94] J. Fu *et al.*, "Rapid AMR prediction in Pseudomonas aeruginosa combining MALDI-TOF MS with DNN model," *J of applied microbiology*, vol. 134, Nov. 2023.

[95] H.-Y. Wang *et al.*, "Rapid and accurate discrimination of Mycobacterium abscessus subspecies based on matrix-assisted laser desorption ionization-time of flight Spectrum and machine learning algorithms," *Biomedicines*, vol. 11, no. 1, p. 45, 2023.

[96] S. Iskender *et al.*, "Rapid Determination of Colistin Resistance in Klebsiella pneumoniae by MALDI-TOF Peak Based Machine Learning Algorithm with MATLAB," *Diagn. Microbiol Infect. Dis.*, Aug. 2023.

[97] K. Liu *et al.*, "Rapid discrimination of Bifidobacterium longum subspecies based on MALDI-TOF MS and machine learning," *Frontiers in Microbiol*, vol. 14, Dec. 2023.

[98] A. Candela *et al.*, "Rapid Discrimination of Pseudomonas aeruginosa ST175 Isolates Involved in a Nosocomial Outbreak Using MALDI-TOF Mass Spectrometry and FTIR Spectroscopy Coupled with Machine Learning," *Transboundary and Emerging Diseases*, vol. 2023, no. 1, p. 8 649 429, 2023.

[99] Y.-M. Zhang *et al.*, "Rapid identification of carbapenem-resistant Klebsiella pneumoniae based on matrix-assisted laser desorption ionization time-of-flight mass spectrometry and an artificial neural network model," *J of Biomedical Science*, vol. 30, Apr. 2023.

[100] C.-R. Chung *et al.*, "Towards Accurate Identification of Antibiotic-Resistant Pathogens through the Ensemble of Multiple Preprocessing Methods Based on MALDI-TOF Spectra," *International J of Molecular Sciences*, vol. 24, Jan. 2023.

[101] W.-Y. Wang *et al.*, "Accurate prediction of antimicrobial resistance and genetic marker of Staphylococcus aureus clinical isolates using MALDI-TOF MS and machine learning–across DRIAMS and Taiwan database," *Int J Antimicrob Agents*, vol. 64, no. 5, p. 107 329, 2024.

[102] G. De Waele *et al.*, "An antimicrobial drug recommender system using MALDI-TOF MS and dual-branch neural networks," *eLife*, vol. 13, RP93242, 2024.

[103] A. Candela *et al.*, "Analysis of high-molecular-weight proteins using MALDI-TOF MS and machine learning for the differentiation of clinically relevant Clostridioides difficile ribotypes," *European J Clin Microbiol & Infectious Diseases*, pp. 1–9, 2024.

[104] T.-H. Lin *et al.*, "Artificial intelligence-clinical decision support system for enhanced infectious disease management: Accelerating ceftazidime-avibactam resistance detection in Klebsiella pneumoniae," *J of Infection and Public Health*, vol. 17, no. 10, p. 102 541, 2024.

[105] M.-J. Jian *et al.*, "Artificial intelligence-clinical decision support system in infectious disease control: combatting multidrug-resistant Klebsiella pneumoniae with machine learning," *Infection and Drug Resistance*, pp. 2899–2912, 2024.

[106] L. Huang *et al.*, "Automated classification of group B Streptococcus into different clonal complexes using MALDI-TOF mass spectrometry," *Frontiers in Molecular Biosciences*, vol. 11, p. 1 355 448, 2024.

[107] X. Xu, "Modelling the rapid detection of Carbapenemase-resistant Klebsiella pneumoniae based on machine learning and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry," *Diagnostic Microbiol and Infectious Disease*, vol. 110, no. 2, p. 116 467, 2024.

[108] X. López *et al.*, "MSDeepAMR: antimicrobial resistance prediction based on deep neural networks and transfer learning," *Frontiers in Microbiol*, vol. 15, Apr. 2024.

[109] C. A. Astudillo *et al.*, "Multi-label classification to predict antibiotic resistance from raw clinical MALDI-TOF mass spectrometry data," *Scientific Reports*, vol. 14, no. 1, p. 31 283, 2024.

[110] M.-J. Jian *et al.*, "Pioneering Klebsiella Pneumoniae Antibiotic Resistance Prediction With Artificial Intelligence-Clinical Decision Support System–Enhanced Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry: Retrospective Study," *J of Medical Internet Research*, vol. 26, e58039, 2024.

[111] V. Macaya Mejias *et al.*, "Predicting bacterial antibiotic resistance using MALDI-TOF mass spectrometry databases with ELM applications," *J of Computer Science & Technology*, vol. 24, 2024.

[112] H.-A. Nguyen *et al.*, "Predicting Pseudomonas aeruginosa drug resistance using artificial intelligence and clinical MALDI-TOF mass spectra," *MSystems*, vol. 9, no. 9, e00789–24, 2024.

[113] M. Ren *et al.*, "Repurposing MALDI-TOF MS for effective antibiotic resistance screening in Staphylococcus epidermidis using machine learning," *Scientific Reports*, vol. 14, no. 1, p. 24 139, 2024.

[114] World Health Organization, *WHONET Antimicrobial Codes*, Sep. 2021. [Online]. Available: `https://whonet.org/WebDocs/WHONET_Antimicrobial_Codes.pdf`.

[115] S. Gibb and K. Strimmer, "MALDIquant: a versatile R package for the analysis of mass spectrometry data," *Bioinformatics*, vol. 28, no. 17, pp. 2270–2271, 2012.

[116] K. Bilecen *et al.*, "Performances and reliability of Bruker Microflex LT and VITEK MS MALDI-TOF MS systems for the identification of clinical microorganisms," *BioMed research international*, vol. 2015, no. 1, p. 516 410, 2015.

[117] [Online]. Available: `https://www.bruker.com/`.

[118] [Online]. Available: `https://www.biomerieux.com/`.

[119] B. Daltonics, *flexAnalysis 3.4 user manual*, 2011.

[120] R. Ketterlinus *et al.*, "Fishing for biomarkers: analyzing mass spectrometry data with the new ClinProTools™ software," *Biotechniques*, vol. 38, no. S6, S37–S40, 2005.

[121] A. Lianou and K. P. Koutsoumanis, "Strain variability of the behavior of foodborne bacterial pathogens: a review," *International J of Food Microbiol*, vol. 167, no. 3, pp. 310–321, 2013.

[122] N. Singhal *et al.*, "MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis," *Frontiers in microbiology*, vol. 6, p. 791, 2015.

[123] N. V. Chawla *et al.*, "SMOTE: synthetic minority over-sampling technique," *J of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[124] A. Lee *et al.*, "Methicillin-resistant Staphylococcus aureus," *Nat. Rev. Dis. Primers*, vol. 4, no. 1, pp. 1–23, 2018.

[125] N. Turner *et al.*, "Methicillin-resistant Staphylococcus aureus: an overview of basic and clinical research," *Nat. Rev. Microbiol*, vol. 17, no. 4, pp. 203–218, 2019.

[126] G. Daikos and A. Markogiannakis, "Carbapenemase-producing Klebsiella pneumoniae: (when) might we still consider treating with carbapenems?" *Clin. Microbiol Infect.*, vol. 17, no. 8, pp. 1135–1141, 2011.

[127] A. Stapleton *et al.*, "Escherichia coli resistance to fluoroquinolones in community-acquired uncomplicated urinary tract infection in women: a systematic review," *Antimicrob. Agents Chemother.*, vol. 64, no. 10, e00862–20, 2020.

[128] B. Johnston *et al.*, "Global molecular epidemiology of carbapenem-resistant Escherichia coli (2002–2017)," *Eur. J Clin. Microbiol Infect. Dis.*, pp. 1–13, 2021.

[129] A. Elshamy and K. Aboshanab, "A Review on Bacterial Resistance to Carbapenems: epidemiology, Detection and Treatment Options," *Future Science OA*, vol. 6, Jan. 2020.

[130] World Health Organization, *WHO bacterial priority pathogens list, 2024: Bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance*. World Health Organization, 2024.

[131] R. Schaumann *et al.*, "Discrimination of human pathogen clostridium species especially of the heterogeneous C. sporogenes and C. botulinum by MALDI-TOF mass spectrometry," *Current microbiology*, vol. 75, pp. 1506–1515, 2018.

[132] M. F. Kilkenny and K. M. Robinson, *Data quality:"Garbage in–garbage out"*, 2018.

[133] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[134] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016, pp. 785–794.

[135] G. De Waele *et al.*, "Pre-trained Maldi Transformers improve MALDI-TOF MS-based prediction," *Computers in Biology and Medicine*, vol. 186, p. 109 695, 2025.

[136] Python Software Foundation, *Python Language Reference, version 3*, Available at `http://www.python.org`, 2023.

[137] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J Mach Learn Res*, vol. 12, pp. 2825–2830, 2011.

[138] Martín Abadi *et al.*, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, 2015. [Online]. Available: `https://www.tensorflow.org/`.

[139] Applied Maths NV, *Bionumerics v7.6.3*, Sint-Martens-Latem, Belgium, 2024. [Online]. Available: `https://www.applied-maths.com`.

[140] ASTA, *AMRQuest Software, v.2.1*, Suwon, Korea, 2024. [Online]. Available: `https://www.astasolutions.com`.

[141] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: `https://www.R-project.org/`.

[142] Clover Bioanalytical Software, *Clover: MALDI-TOF MS Data Analysis Software*, Available at `https://www.cloverms.com`, Clover Bioanalytical Software, Granada, Spain, 2023.

[143] T. M. Inc., *MATLAB version: 24.1.0 (R2024a)*, Natick, Massachusetts, United States, 2024. [Online]. Available: `https://www.mathworks.com`.

[144] GitHub, Inc., *GitHub*, `https://github.com`, 2023.

[145] United Nations, *The Sustainable Development Goals Report 2023: Special Edition*, [Accessed: February 17, 2025], 2023.