Review Article

Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review

Author(s):
Weis, Caroline V.; Jutzeler, Catherine R.; Borgwardt, Karsten

Publication Date:
2020-10

Permanent Link:
https://doi.org/10.3929/ethz-b-000442204 →

Originally published in:
Clinical Microbiology and Infection 26(10), http://doi.org/10.1016/j.cmi.2020.03.014 →

Clinical Microbiology and Infection 26 (2020) 1310–1317

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Review

Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review

C.V. Weis [1, 2, *, †], C.R. Jutzeler [1, 2, †], K. Borgwardt [1, 2]

[1)] *Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland*
[2)] *SIB Swiss Institute of Bioinformatics, Switzerland*

ARTICLE INFO

Article history:
Received 3 December 2019
Received in revised form
5 March 2020
Accepted 13 March 2020
Available online 23 March 2020

Editor: L. Leibovici

Keywords:
Antimicrobial resistance
Antimicrobial susceptibility testing
Antimicrobial treatment
Machine learning
MALDI-TOF MS
Microbial identification

ABSTRACT

Background: The matrix assisted laser desorption/ionization and time-of-flight mass spectrometry (MALDI-TOF MS) technology has revolutionized the field of microbiology by facilitating precise and rapid species identification. Recently, machine learning techniques have been leveraged to maximally exploit the information contained in MALDI-TOF MS, with the ultimate goal to refine species identification and streamline antimicrobial resistance determination.
Objectives: The aim was to systematically review and evaluate studies employing machine learning for the analysis of MALDI-TOF mass spectra.
Data sources: Using PubMed/Medline, Scopus and Web of Science, we searched the existing literature for machine learning-supported applications of MALDI-TOF mass spectra for microbial species and antimicrobial susceptibility identification.
Study eligibility criteria: Original research studies using machine learning to exploit MALDI-TOF mass spectra for microbial specie and antimicrobial susceptibility identification were included. Studies focusing on single proteins and peptides, case studies and review articles were excluded.
Methods: A systematic review according to the PRISMA guidelines was performed and a quality assessment of the machine learning models conducted.
Results: From the 36 studies that met our inclusion criteria, 27 employed machine learning for species identification and nine for antimicrobial susceptibility testing. Support Vector Machines, Genetic Algorithms, Artificial Neural Networks and Quick Classifiers were the most frequently used machine learning algorithms. The quality of the studies ranged between poor and very good. The majority of the studies reported how to interpret the predictors (88.89%) and suggested possible clinical applications of the developed algorithm (100%), but only four studies (11.11%) validated machine learning algorithms on external datasets.
Conclusions: A growing number of studies utilize machine learning to optimize the analysis of MALDI-TOF mass spectra. This review, however, demonstrates that there are certain shortcomings of current machine learning-supported approaches that have to be addressed to make them widely available and incorporated them in the clinical routine. **C.V. Weis, Clin Microbiol Infect 2020;26:1310**

Introduction

Antimicrobial resistance (AMR)—the ability of microorganisms to resist antimicrobial treatment—constitutes a global challenge that threatens the effective prevention and treatment of an ever-increasing range of infections caused by bacteria and other pathogens. In a clinical setting, the rapid and reliable identification of potential pathogens is of utmost importance for a timely initiation of appropriate antimicrobial treatment. Over the last decades, matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) has revolutionized clinical diagnostics with its capability of rapid, reliable and cost-effective species identification as well as its potential to streamline antimicrobial susceptibility testing [1–4]. Numerous studies have demonstrated that MALDI-TOF MS surpasses conventional

* Corresponding author. C.V. Weis, Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, Basel, 4058, Switzerland.
E-mail address: caroline.weis@bsse.ethz.ch (C.V. Weis).
† These authors contributed equally.

https://doi.org/10.1016/j.cmi.2020.03.014
1198-743X/

diagnostic methods in terms of cost, speed and accuracy in microbial species identification [5–9]. Conventional phenotypic diagnostic methods, such as antimicrobial gradient method and disc diffusion test [10,11], require up to 48 hr and cost $100, while MALDI-TOF MS can provide information on the inflicting pathogen within minutes from cultures grown overnight at the cost of a few dollars.

Conventionally, the analysis of MALDI-TOF mass spectra relies on a small number of attributes, such as peak height and area under the peak, that have been empirically linked to microbial species. While this is a valid approach and works fairly well at species level, there is a wealth of information contained in these spectra that remains unused. To fully exploit the information contained in MALDI-TOF mass spectra, researchers have been implementing machine learning algorithms in their efforts to refine species identification [12–14] and streamline antimicrobial resistance determination [15,16]. Machine learning methods are capable of finding statistical dependencies in the data, also considering non-linear and interaction effects between features. Thereby, machine learning techniques can uncover novel or unknown information that is embedded in the MALDI-TOF mass spectra. This information has proven useful for identification and differentiation of species, particularly those that are phylogenetically proximal as well as sublineages of species [17–19]. Moreover, it has been recently recognized that information contained in MALDI-TOF mass spectra can also aid antibiotic resistance profiling [17,18]. Specifically, the MALDI-TOF technique is primarily based on fingerprinting analyses of ribosomal proteins, which are naturally suited to identify bacteria and systematically derive their relatedness to each other. Assuming that the AMR characteristics in closely related strains are similar, MALDI-TOF spectra of bacteria with known AMRs serve as a reference to determine the AMR profile of future analytes.

In this systematic review, we have compiled the current literature on machine learning-supported applications of MALDI-TOF mass spectra for species identification and antimicrobial susceptibility testing. The first aim was to gather information regarding the species investigated, machine learning algorithms employed and model performance. The second aim was to assess the quality of the studies in terms of reproducibility, robustness, generalizability and clinical significance of the machine learning models employed.

Methods

The study was registered with PROSPERO. We followed the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) statement [19,20].

Search methods for identification of studies

A systematic literature search of original research articles was performed using PubMed/Medline, Scopus and Web of Science using the time range from their respective inception dates to 31 January 2020. The search term string was constructed as ('machine learning' OR 'classification algorithm' OR 'support vector machine' OR 'random forest' OR 'logistic regression' OR 'neural network') AND 'maldi-tof', to include publications analysing MALDI-TOF MS data with different machine learning methods. Manual searching was also performed, reviewing reference lists of relevant trials and comprehensive review articles.

Selection of studies

One author (C.V.W.) carried out an initial screening of retrieved articles and applied inclusion criteria. Subsequently, a second reviewer (C.R.J.) independently reviewed all the studies in order to ensure that the publications met all inclusion criteria. All disagreements were discussed and resolved at a consensus meeting.

Inclusion and exclusion criteria

All original studies using machine learning-supported applications of MALDI-TOF mass spectra for microbial species and antimicrobial susceptibility identification were included. Included studies must have provided information on the machine learning algorithms used for the analysis as well as on the species of interest (genera, sample size). Studies employing MALDI-TOF mass spectra analysis of single proteins and peptides were excluded. Also excluded were studies stating a non-microbial research topic, such as cancer or genomics, paediatric studies, case studies and review articles.

Data extraction and synthesis

The following information was extracted from all studies: (a) publication characteristics (first author's last name, publication time), (b) study objectives (species discrimination, identification or antimicrobial susceptibility testing), (c) cohort selection (genera, sample size), (d) specifics on MALDI-TOF instrument used, (e) model selection (applied machine learning algorithm, platforms (software) and packages, parameters), (f) statistics for model performance (methods to evaluated the model, mean and measure of variance) and (g) methods to avoid overfitting as well as external validation strategies. Supplementary material for each study was also reviewed if available.

Quality assessment of machine learning studies

Based on nine criteria relevant to the objectives of the review (adapted from Qiao [21]), the quality of the included machine learning studies was assessed. The quality assessment comprised five categories: unmet needs (rationale for machine learning algorithm), reproducibility (feature engineering, software and hardware, hyperparameters), robustness (valid methods to overcome overfitting, stability of results), generalizability (external data validation) and clinical significance (interpretation of predictors and suggested clinical use). A quality assessment table was provided by listing 'yes' or 'no' of corresponding items in each category. C.V.W. and C.R.J. independently performed the quality assessment by providing a 'yes' or 'no' of corresponding items in each category. Disagreements of ratings were discussed and final scores for each publication were determined.

Results

Study selection and study characteristics

The results of the literature search, including the numbers of studies screened, assessments for eligibility and articles reviewed (with reasons for exclusions at each stage) are presented in Fig. 1. A total of 36 published studies met the inclusion criteria: 27 for microbial species identification [12,14,22–49] and nine for antibiotic susceptibility testing [15,16,44,47,48,50–53]. The majority of excluded studies ($n = 924$) did not meet one or multiple inclusion criteria, such as the applied data analysis (e.g. statistical approaches, including linear and logistic regression), research topic beyond the current review and study design (e.g. case reports and reviews). Detailed information on all included studies are provided in Table S1.

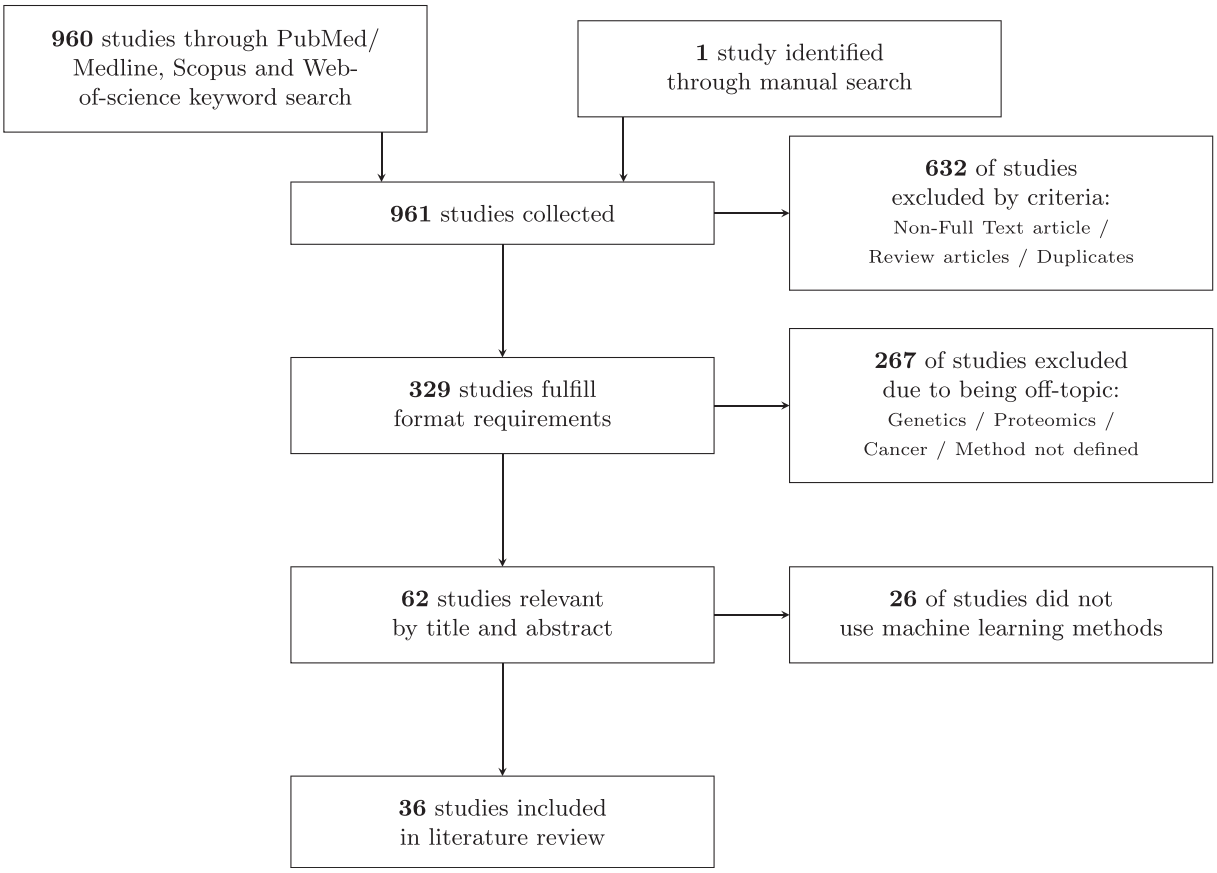

Fig. 1. PRISMA flowchart illustrating the complete literature search pipeline. A total of 960 studies were found using the search phrase ('machine learning' OR 'classification algorithm' OR 'support vector machine' OR 'random forest' OR 'logistic regression' OR 'neural net') AND 'maldi-tof'. One additional publication was identified by manual search of references. We excluded 632 studies based on article type (reviews or non-full-text research articles) or because of duplicate entry. The title and abstract of the remaining 329 studies were screened. We excluded 267 studies based on their research topic that was beyond the scope of our review, such as studying proteins instead of whole microbial organisms. Another 26 studies were excluded after reading the articles because no machine learning methods were not employed for the analysis. The remaining 36 studies were reviewed in detail to be included in this review.

Species genera and antibiotic drugs investigated

Table S1 provides an overview of the wide range of microbial genera that were investigated in the studies reviewed. The most frequently investigated genera were *Staphylococcus* ($n = 14$), *Streptococcus* ($n = 6$), *Escherichia* ($n = 4$), *Bacillus* ($n = 4$) and *Klebsiella* ($n = 3$). In terms of susceptibility testing, included studies focused on the broad-spectrum antibiotics vancomycin ($n = 3$) and carbapenems ($n = 1$), narrow-spectrum antibiotic methicillin ($n = 3$) and the antifungal drug fluconazole ($n = 1$). Lastly, the number of isolates analysed varied quite considerably between studies. Wang et al. [25] included 787 isolates in their analysis, while others had less than 50 isolates [32,36,49,52].

Overview of algorithms and software

As highlighted in Table 1, the most widely used machine learning algorithms were Support Vector Machines (SVM, 18 studies), Genetic Algorithms (GA, 15 studies), Artificial/Supervised Neural Networks (ANN, 13 studies) and Quick Classifiers (QC, 11 studies). A detailed description of these four algorithms can be found in Table 2. Less frequently chosen algorithms included clustering/hierarchical cluster analysis (UHCA), Random Forests (RF), decision trees (DT), *k*-Nearest Neighbors (kNN), Multiple Logistic Regression (MLR), Naïve Bayes and Aristotle Classifiers. A large number of the reviewed studies ($n = 17$) used ClinProTools Software (i.e. Brucker Daltonik) to perform the analysis of the MALDI-TOF mass spectra. The remaining studies analysed their data with R or R Studio ($n = 9$), MATLAB ($n = 7$), Python ($n = 1$), MALDI Biotools 3.0 ($n = 1$), Statistics Program for Social Sciences ($n = 1$), Mathematica ($n = 1$) or a combination thereof (Table S1).

Model validation

All reviewed studies employed a type of cross-validation (e.g. fivefold, tenfold or leave-one-out cross-validation) to avoid

Table 1
Algorithms used in the studies (n = 36 studies). One study can use multiple algorithms

Type of algorithm	*n* (%)
Support Vector Machine (SVM)	18 (50)
Genetic Algorithm (GA)	15 (41.7)
Artificial/Supervised Neural Network (ANN)	13 (36.1)
Quick Classifier (QC)	11 (30.5)
Random Forest (RF)	9 (5)
Clustering/hierarchical cluster analysis (UHCA)	8 (22.2)
k-Nearest Neighbors (kNN)	5 (13.9)
Decision Tree (DT)	4 (11.1)
Logistic regression (single and multi)	3 (8.3)
Aristotle Classifier	1 (2.8)
Linear Discriminant Analysis	1 (2.8)
Naïve Bayes	1 (2.8)

Table 2
Description of frequently used machine learning algorithms

Algorithm	Description	References
Genetic Algorithm (GA)	Genetic algorithms are optimization algorithms belonging to the larger class for evolutionary algorithms. They rely on operations inspired by biological operations, such as mutations, crossover and selection by evolving a collection of candidate solutions towards a better performing solution. The genetic algorithm is used to select a combination of peaks that separate the classes best, using a cost function that measures the variance between classes. When the genetic algorithm is selected in ClinProTools, the genetic algorithm is only used as a peak (feature) selection algorithm. The classification on unseen instances is performed using a kNN algorithm based on the selected peaks.	[a]
Artificial/Supervised Neural Network (ANN)	Artificial neural networks are algorithms structurally modelled after mammalian brain neural networks. The network consists of several stacked layers of relatively simple mathematical units, which take as input information of several neurons of the previous layers and pass on the output to several neurons in the next layer. Most commonly, the learning takes place by taking the gradient of the loss function and adjusting the model weights to decrease the loss, leading those networks also to be called 'back propagation neural networks' (BPNNs). While unsupervised versions of artificial neural networks exist, the more common supervised setting is used for classification with neural networks.	[c]
Support Vector Machine (SVM)	Support vector machines are a supervised learning algorithm that finds the best separating maximum margin hyperplane between the classes in a higher dimensional representation of the instances. Hyperplane is the general term for planes with one dimension less than the dimension they are in. During optimization the hyperplane maximizing the gap between the plane and the instances is determined. The data representation in higher dimensional space is performed by kernel functions. Commonly used kernel functions include the radial basis function kernel and the polynomial kernel. While SVM is a classification algorithm, when selected in ClinProTools, the SVM algorithm is only used as a peak (feature) selection algorithm. The classification on unseen instances is performed using a kNN algorithm based on the selected peaks.	[b,d,a]
Quick Classifier (QC)	The Quick Classifier calculates the average area of each peak together and provides a p value per class. During classification, the peak areas are sorted by the univariate sorting algorithm and an average over all peaks is calculated which indicates class membership.	[a]
K Nearest Neighbor (kNN)	K-nearest neighbor algorithm is a classification based on the similarity between the instance to classify and known training instances. For classification of an unseen instance, the similarity between the instance and each training data point. The assigned class is determined as the majority class of the closest k training data points. A commonly used similarity measure is the Euclidean distance.	[e]

[a] https://medschool.vanderbilt.edu/msrc/wp-content/uploads/sitwes/41/public_files/Forms/clinprotoolsmanual.pdf.
[b] https://www.nature.com/articles/nbt1206-1565.
[c] http://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html.
[d] https://deepai.org/machine-learning-glossary-and-terms/hyperplane.
[e] https://medium.com/@chiragsehra42/k-nearest-neighbors-explained-easily-c26706aa5c7f.

overfitting. Additional validation of the models on out-of-distribution MALDI-TOF data (i.e. external validation) was only performed in four studies [14,25,36,49]. Specifically, Wang and colleagues [25] collected data for the training and test set from the bacterial bank of a teaching hospital in northern Taiwan. The external validation data set comprised data that were independently obtained from a bacterial biobank and two teaching hospitals in middle and southern Taiwan. Aiming at discriminating contagious from environmental strains of *Streptococcus uberis* in dairy herds, Esener et al. collected data from 29 farms of which the data of 19 farms were used to train and test the algorithms and the remaining ten were hold-out for external validation [36]. Moreover, in many studies [14] machine learning models of 40 *Mycobacterium abscessus* isolates that were collected across France were trained and tested. The subsequent external validation was conducted on 40 *M. abscessus* isolates from the French National Reference Centre for Mycobacteria and Resistance of Mycobacteria to Antituberculosis. Rodrigues et al. [49] aimed at precisely identifying different species of the genus *Klebsiella* based on 46 strains collected from different sources around the globe (e.g. human, environment, water, plant). To validate their SVM-based models, Rodriguez et al. [49] analysed isolates belonging to *K. pneumoniae* phylogroups derived from 49 faecal samples of humans in Madagascar.

Quality of included studies

The results from the quality assessment are shown in Table S1. The quality of the studies ranged from poor (<60%) to very good (100%). Only one study fulfilled all nine quality criteria and, thus, reached 100% in the rating [36]. Out of the nine criteria, four were met by more than 97% of the studies (35 out of 36). Namely, all studies highlighted the limits in current non-machine learning approaches in the introduction, provided information on hardware and software utilized in the study, employed valid methods to avoid model overfitting, and provided information on the clinical relevance. Only ~11% of the studies validated the machine learning models on an external data set.

Discussion

The primary step of this review entailed a systematic search for studies employing machine learning algorithms to advance species identification and susceptibility testing based on MALDI-TOF mass spectra. A total of 36 studies met our inclusion criteria and applied a broad range of machine learning algorithms to identify one or multiple species as well as to support AMR profiling. There was a substantial overlap between studies in terms of inflicting pathogens (e.g. *Staphylococcus*, *Streptococcus* and *Escherichia*) and types of machine learning algorithms applied (e.g. SVM, RF and GA). This is because a large proportion of the analyses were performed on manufacturer-provided software, such as flexAnalysis and ClinProTools from Bruker Daltonics, with preprogrammed machine learning algorithms. Four machine learning algorithms are implemented in ClinProTools, namely GA, SVM, SNN and QC (for more details on the algorithms refer to Table 2) [39]. In a second step, we conducted a systematic assessment of the studies to evaluate the quality of the machine learning models employed. This quality assessment unequivocally points out several shortcomings of current machine learning implementations that have to be addressed in order to incorporate them in the routine diagnostics.

Species identification and antibiotic resistance profiling

Out of 36 identified studies, 27 leveraged the power of machine learning to improve MALDI-TOF-based species identification [12,14,22–49]. From a data analysis perspective, species identification can be formalized as a multiclass (or binary) classification task with special challenges, including high dimensionality, the sheer number of classes and unbalanced class ratios [13,54]. Machine learning methods can be used to modify representations of MALDI-TOF spectra data by removing uninformative variables and to determine which specific piece of (unknown) information will be most valuable to obtain and optimize species identification. By recognizing patterns in the data, machine learning techniques maximally exploit the information that is embedded in the MALDI-TOF spectra. Most studies reported superior performance of machine learning algorithms in the classification task than current approaches. For instance, employing ANN on MALDI-TOF spectra allowed the rapid and accurate identification of *Bacillus fragilis* and some subgroups thereof [23,33]. Similarly, De Bruyne and colleagues demonstrated how well machine learning techniques can discriminate between MALDI-TOF spectra of different species or classes [12]. Specifically, they applied SVM and RF to binarized MALDI-TOF spectra of species within the genera *Leuconostoc*, *Fructobacillus* and *Lactococcus* and discriminatory performance of these methods was excellent. In addition to species identification, our literature search yielded nine studies that made use of machine learning for AMR prediction [15,16,44,47,48,50–53]. Six out of these nine studies tested the antimicrobial susceptibility of *Staphylococcus aureus* [15,16,44,50,52,53], a pathogen that is associated with a high mortality rate in both hospital and community settings due to the dearth of effective treatments [55,56]. Sogawa and colleagues used SVM to discriminate methicillin-resistant (MRSA) from methicillin-sensitive *S. aureus* (MSSA) based on features derived from MALDI-TOF mass spectra [15]. Their model reached prediction accuracies of over 85% and significantly reduced the time to initiation of targeted antibiotic treatment in comparison with phenotypic resistance profiling. Against the backdrop of resistance, vancomycin emerged as the first-line treatment for MRSA infections. Unfortunately, MRSA strains have also become resistant to vancomycin as a result of the overuse/misuse that caused an accumulation of mutations in genes involved in cell wall synthesis [57]. The resulting vancomycin-intermediate *S. aureus* (VISA) strains are associated with enduring infections and treatment failure [57]. Both, SVM and RF have been applied to VISA and vancomycin-susceptible *S. aureus* (VSSA) MALDI-TOF mass spectra for separation of strains and phenotypic prediction [16,53], with reported accuracies of over 97%. Comparable successes of machine learning based AMR prediction were also reported for β-lactam antibiotic-resistant *B. fragilis*. Briefly, SVM-based prediction algorithms were able to predict the presence of the *cfi* gene in *B. fragilis* with an accuracy close to 100% [51]. The *cfi* gene is suspected to govern the production of metallo-β-lactamase, which hydrolyses carbapenems and other β-lactam antibiotics and, thereby, renders these treatments ineffective against *B. fragilis* strains [58,59]. Huang and colleagues employed five different machine learning algorithms (KNN, RF, SVM, naïve Bayes and logistic regression), of which RF outperformed the other four algorithms in discriminate carbapenem-resistant from carbapenem-susceptible *Klebsiella pneumoniae* with accuracy rates of over 93% [48]. Lastly, a recent proof-of-principle study in the field of mycology demonstrated the potential of using machine learning approaches to predict resistance to antifungals (e.g. fluconazole) [47].

Model evaluation is key

The success of the clinical applicability of machine learning frameworks hinges on their robustness, reliability and validity. Our quality assessment revealed that the majority of the reviewed studies (Table S1) undertook valid steps to assess the robustness of their models by means of standard *k*-fold cross-validation and stating the stability of the results (e.g. standard deviation, confidence intervals). In the cross-validation, the data set is divided into *k* subsets, called folds. Subsequently, the algorithm is iteratively trained on *k*-1 folds while using the remaining fold as the test set (called the 'holdout fold'). During the cross-validation procedure, hyperparameters can be tuned only using the original training set. Then, the model selection and performance assessment are performed using the test set (i.e. data that has never been seen before). Somewhat surprisingly, only 60.71% of the reviewed studies provided details on the hyperparameters, which are indispensable for replication studies. To gauge the clinical applicability of the models, further types of validation are required. Firstly, the generalizability (or interoperability) of models is essential to allow between-hospital comparisons and applications. MALDI-TOF mass spectra measured at different locations suffer from batch effects [60], which are likely to stem from differences in laboratory routine or machine settings. Hence models trained on spectra collected at a specific hospital are likely have worse predictions on out-of-hospital data (also called 'out-of-distribution data'). Conventionally, the generalizability of a model is assessed through the so-called external validation, during which the model is presented with unseen, out-of-distribution data (e.g. data from different hospitals or strains). Considering that only ~11% of all reviewed studies have performed an external validation, our review points out a major limitation in current machine learning frameworks for MALDI-TOF MS data analysis. A possible explanation, although speculative, is that the collection and characterization of the AMR profiles of isolates from multiple sites is expensive and requires a large-scale effort to collect, annotate and combine the datasets while checking for consistency.

Secondly, the interpretability of machine learning models is crucial for applying predictors in patient care. Confidence in the model is built by understanding the MALDI-TOF MS peaks from which information is derived. Almost all reviewed studies provided an explanation (biological or quantification) of the importance of the predictive peaks. Esener et al. [36], for instance, cross-referenced the peaks identified with the NCBI protein database to find the corresponding protein. Based on this analysis they found that their peaks correspond to bacteriocin and ribosomal proteins providing a biological plausible explanation. In another study, Ho and colleagues used a CabaNP assay to characterize the *Bacteroides fragilis* strains that were classified *cfiA*-positive and, thus, likely to be resistant to the antibiotic carbapenem [51]. This characterization revealed that over a third of these strains harbour an upstream insertion sequence of the gene *cfiA*, which is suspected to cause the carbapenem resistance of *Bacteroides fragilis*. Moreover, MALDI-TOF MS peaks can provide insights to whether known resistance properties or unrelated phylogenetic differences lead to predictive performance. In this manner, Sogawa et al. [15] compared predictive peaks picked by their algorithm with known fragments of the methicillin resistance-causing penicillin-binding protein PBP in MRSA. However, biological interpretation of feature peaks is rare as a consequence of limited prior knowledge, as most resistance mechanisms have not been previously studied by means of MALDI-TOF mass spectra.

Limitations of machine learning applications

Machine learning applications on MALDI-TOF mass spectra data are still in their infancy, also evidenced by the low number of publications available (Table S1). We identified three key obstacles to swift progress: small sample size, lack of external validation and poor reproducibility. All reviewed studies were limited to small sample sizes, ranging between dozens and hundreds of isolates for training and testing, with the largest study including 787 isolates. Isolate collections of this size hardly reflect the microbial diversity that would occur in clinical routine. Training machine learning predictors is hampered by small and microbial non-diverse samples, likely to be causing inflated false discovery rates and low generalizability [61]. The generalizability of models is further impeded by the lack of external validation. Both limitations could be addressed by large publicly available data sets. For instance, the MassIVE repository [62] is a community database for mass spectrometry mainly focused on proteomics and containing MALDI-TOF datasets. A larger number of spectra is provided by the Robert Koch Institute, including 6264 MALDI-TOF mass spectra of highly pathogenic microbes [63]. In addition to a large number of samples, these data repositories should ideally provide AMR profiles along with the spectra and crucial information to create a well-defined problem statement (e.g. annotation of sublineages and AMR profiles). Lastly, the reproducibility of the reviewed studies is low. MALDI-TOF MS datasets (in our review only nine studies [25,28,37,38,46–48,50,53]) and machine learning codes (in our review only five studies [12,34,37,50]) are rarely made public after publication. Additionally, the information on hyperparameter choices for models is often insufficient [14,15,26,28–31,33,35,40,51]. The lack of this information makes it virtually impossible to understand, reimplement, and most importantly reproduce the described analysis pipelines. Compounding the problem is the fact that the majority of analyses are conventionally performed on manufacturer-provided software, such as flexAnalysis and ClinProTools from Bruker Daltonics [14,15,22,24,25,27,28,30,31,34–36,39–41,51]. As a consequence, applied workflows are not accessible, thereby obstructing any external attempts to refine the pipelines or implement new ones. To tackle these challenges and unleash the full potential of machine learning applications, joint efforts of clinical and computational researchers are required. In addition to publicly accessible MALDI-TOF MS datasets that are suitable for machine learning, analytical codes should be made public on repositories, such as GitHub [64]. This will foster the replication of published experiments, provide a foundation to new ones, and allow the application of advanced machine learning algorithms [65–68].

Limitations of this systematic review

A major limitation of the current review was that literature search was limited to articles listed in PubMed/Medline, Scopus and Web of Science, or identified by hand searches. Considering the pace at which the research in this area is moving forward, it is likely that the findings of the publications described in this paper will be quickly complemented by further research. The literature search also excluded grey literature (e.g. preprints, reports, conference proceedings), the importance of which to this topic is unknown, and thus might have introduced another source of search bias. The lack of studies reporting poor performance of machine learning algorithms regarding species identification and susceptibility testing speaks to a high probability of publication bias, as well as potential for a search bias. Publication bias is likely to result in studies with more positive results being preferentially submitted and accepted for publication.

Conclusion and future directions

This the first systematic review of studies employing machine learning to refine the identification of microbial species and streamline the AMR profiling. In contrast to other biomedical fields, microbiology is lagging behind in terms of reaping the benefits of big data and machine learning. While the overall quality of the majority of the studies is fair to good, three major limitations, namely small sample size, poor reproducibility and lack of external validation, impede swift progress. To unleash the full potential of machine learning guided species identification and AMR profiling, it is important that computational and clinical scientists tackle the abovementioned challenges in joint efforts.

Transparency Declaration

This study was supported by the Alfried Krupp Prize for Young University Teachers of the Alfried Krupp von Bohlen und Halbach-Stiftung (Borgwardt), Germany, an ETH “Personalised Medicine Basel” grant (PMB-03-17, Borgwardt, Weis), and the Swiss National Science Foundation, Bern, Switzerland (Ambizione Grant, PZ00P3_186101, Jutzeler). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors do not report any (financial or otherwise) conflict of interest.

Author contributions

Caroline Weis and Catherine Jutzeler contributed substantially to the data acquisition, analysis (i.e. quality assessment), and interpretation. Furthermore, they drafted the review article. Karsten Borgwardt made substantial contributions to data interpretation (i.e. quality assessment) and participated in revising the review article critically for important intellectual content.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cmi.2020.03.014.

References

[1] Novak SM, Marlowe EM. Automation in the clinical microbiology laboratory. Clin Lab Med 2013;33:567–88.
[2] Croxatto A, Prod'hom G, Faverjon F, Rochais Y, Greub G. Laboratory automation in clinical bacteriology: what system to choose? Clin Microbiol Infect 2016;22:217–35.
[3] Moreno-Camacho JL, Calva-Espinosa DY, Leal-Leyva YY, Elizalde-Olivas DC, Campos-Romero A, Alcántar-Fernández J. Transformation from a conventional clinical microbiology laboratory to full automation. Lab Med 2017;49:e1–8.
[4] Vrioni G, Tsiamis C, Oikonomidis G, Theodoridou K, Kapsimali V, Tsakris A. MALDI-TOF mass spectrometry technology for detecting biomarkers of antimicrobial resistance: current achievements and future perspectives. Ann Transl Med 2018;6:240.
[5] Scola BL, La Scola B, Raoult D. Direct identification of bacteria in positive blood culture bottles by matrix-assisted laser desorption ionisation time-of-flight mass spectrometry. PLoS ONE 2009;4:e8041.
[6] Stevenson LG, Drake SK, Murray PR. Rapid identification of bacteria in positive blood culture broths by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 2010;48:444–7.
[7] Foster AGW. Rapid identification of microbes in positive blood cultures by use of the Vitek MS matrix-assisted laser desorption ionization–time of flight mass spectrometry system. J Clin Microbiol 2013;51:3717–9.
[8] Haigh JD, Green IM, Ball D, Eydmann M, Millar M, Wilks M. Rapid identification of bacteria from bioMérieux BacT/ALERT blood culture bottles by MALDI-TOF MS. Br J Biomed Sci 2013;70:149–55.
[9] Tadros M, Petrich A. Evaluation of MALDI-TOF mass spectrometry and Sepsityper Kit™ for the direct identification of organisms from sterile body fluids in a Canadian pediatric hospital. Can J Infect Dis Med Microbiol 2013;24:191–4.

[10] Reller LB, Weinstein M, Jorgensen JH, Ferraro MJ. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. Clin Infect Dis 2009;49:1749–55.

[11] Stratton CW. Susceptibility testing revisited. Prog Clin Pathol 1984;9:65–100.

[12] De Bruyne K, Slabbinck B, Waegeman W, Vauterin P, De Baets B, Vandamme P. Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. Syst Appl Microbiol 2011;34:20–9.

[13] Vervier K, Mahé P, Veyrieras J-B, Vert J-P. Benchmark of structured machine learning methods for microbial identification from mass-spectrometry data. 2015. arXiv:1506.07251 [statML], https://arxiv.org/abs/1506.07251.

[14] Fangous M-S, Mougari F, Gouriou S, Calvez E, Raskine L, Cambau E, et al. Classification algorithm for subspecies identification within the Mycobacterium abscessus species, based on matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 2014;52:3362–9.

[15] Sogawa K, Watanabe M, Ishige T, Segawa S, Miyabe A, Murata S, et al. Rapid Staphylococcus aureus discrimination between methicillin-sensitive and methicillin-resistant using MALDI-TOF MASS SPECTROMETRY. Biocontrol Sci 2017;22:163–9.

[16] Mather CA, Werth BJ, Sivagnanam S, SenGupta DJ, Butler-Wu SM. Rapid detection of vancomycin-intermediate Staphylococcus aureus by matrix-assisted laser desorption ionization–time of flight mass spectrometry. J Clin Microbiol 2016;54:883–90.

[17] Burckhardt I, Zimmermann S. Susceptibility testing of bacteria using MALDI-TOF mass spectrometry. Front Microbiol 2018;9:1744.

[18] Florio W, Tavanti A, Barnini S, Ghelardi E, Lupetti A. Recent advances and ongoing challenges in the diagnosis of microbial infections by MALDI-TOF mass spectrometry. Front Microbiol 2018;9:1097.

[19] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al., PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4. https://doi.org/10.1186/2046-4053-4-1.

[20] Leeflang MMG, Kraaijpoel N. Systematic reviews of diagnostic test accuracy in CMI. Clin Microbiol Infect 2018;24:1115–6.

[21] Qiao N. A systematic review on machine learning in sellar region diseases: quality and reporting items. Endocr Connect 2019;8:952–60.

[22] Zhang T, Ding J, Rao X, Yu J, Chu M, Ren W, 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 2015;117:122–7.

[23] Zhang Z, Wang D, Harrington P de B, Voorhees KJ, Rees J. Forward selection radial basis function networks applied to bacterial classification based on MALDI-TOF-MS. Talanta 2004;63:527–32.

[24] Xiao D, Zhao F, Zhang H, Meng F, Zhang J. 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 2014;52:3038–43.

[25] Wang H-Y, Li W-C, Huang K-Y, Chung C-R, Horng J-T, Hsu J-F, 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 Bioinform 2019;20:703.

[26] Wang H-Y, Lee T-Y, Tseng Y-J, Liu T-P, Huang K-Y, Chang Y-T, 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 2018;13:e0194289.

[27] Wang H-Y, Lien F, Liu T-P, Chen C-H, Chen C-J, Lu J-J. Application of a MALDI-TOF analysis platform (ClinProTools) for rapid and preliminary report of MRSA sequence types in Taiwan. PeerJ 2018;6:e5784.

[28] Sonthayanon P, Jaresitthikunchai J, Mangmee S, Thiangtrongjit T, Wuthiekanun V, Amornchai P, et al. Whole cell matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) for identification of Leptospira spp. in Thailand and Lao PDR. PLoS Negl Trop Dis 2019;13:e0007232.

[29] Schaumann R, Knoop N, Genzel GH, Losensky K, Rosenkranz C, Stîngu CS, et al. A step towards the discrimination of beta-lactamase-producing clinical isolates of Enterobacteriaceae and Pseudomonas aeruginosa by MALDI-TOF mass spectrometry. Med Sci Monit 2012;18. MT71–7.

[30] Nakano S, Matsumura Y, Ito Y, Fujisawa T, Chang B, Suga S, et al. Development and evaluation of MALDI-TOF MS-based serotyping for Streptococcus pneumoniae. Eur J Clin Microbiol Infect Dis 2015;34:2191–8.

[31] Nakano S, Matsumura Y, Kato K, Yunoki T, Hotta G, Noguchi T, 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 2014;44:256–9.

[32] Ling J, Wang H, Li G, Feng Z, Song Y, Wang P, 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 2019;14:e0222636.

[33] Lasch P, Beyer W, Nattermann H, Stämmler M, Siegbrecht E, Grunow R, et al. Identification of Bacillus anthracis by using matrix-assisted laser desorption ionization-time of flight mass spectrometry and artificial neural networks. Appl Environ Microbiol 2009;75:7229–42.

[34] Khot PD, Fisher MA. Novel approach for differentiating Shigella species and Escherichia coli by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 2013;51:3711–6.

[35] Hsieh S-Y, Tseng C-L, Lee Y-S, Kuo A-J, Sun C-F, Lin Y-H, et al. Highly efficient classification and identification of human pathogenic bacteria by MALDI-TOF MS. Mol Cell Proteomics 2008;7:448–56.

[36] Esener N, Green MJ, Emes RD, Jowett B, Davies PL, Bradley AJ, et al. Discrimination of contagious and environmental strains of Streptococcus uberis in dairy herds by means of mass spectrometry and machine-learning. Sci Rep 2018;8:17517.

[37] Desaire H, Hua D. Adaption of the Aristotle classifier for accurately identifying highly similar bacteria analyzed by MALDI-TOF MS. Anal Chem 2020;92:1050–7.

[38] Chung C-R, Wang H-Y, Lien F, Tseng Y-J, Chen C-H, Lee T-Y, 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. Front Microbiol 2019;10:2120.

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

[40] Camoez M, Sierra JM, Dominguez MA, Ferrer-Navarro M, Vila J, Roca I. Automated categorization of methicillin-resistant Staphylococcus aureus clinical isolates into different clonal complexes by MALDI-TOF mass spectrometry. Clin Microbiol Infect 2016;22:161.e1–7.

[41] Pérez-Sancho M, Vela AI, Horcajo P, Ugarte-Ruiz M, Domínguez L, Fernández-Garayzábal JF, et al. Rapid differentiation of Staphylococcus aureus subspecies based on MALDI-TOF MS profiles. J Vet Diagn Invest 2018;30:813–20.

[42] Li B, Guo T, Qu F, Li B, Wang H, Sun Z, et al. Matrix-assisted laser desorption ionization: time of flight mass spectrometry-identified models for detection of ESBL-producing bacterial strains. Med Sci Monit Basic Res 2014;20:176–83.

[43] Lee J, Shin Y, Kim S, Rho K, Park KH. SVM classification model of similar bacteria species using negative marker: based on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. In: IEEE 17th international conference on bioinformatics and bioengineering (BIBE). Piscataway, NJ: IEEE; 2017. p. 145–50.

[44] Bai J, Fan ZC, Zhang LP, Xu XY, Zhang ZL. Classification of methicillin-resistant and methicillin-susceptible Staphylococcus aureus using an improved genetic algorithm for feature selection based on mass spectra. In: Proceedings of the 9th international conference on bioinformatics and biomedical technology – ICBBT '17. New York, NY: ACM; 2017. p. 7–63.

[45] Schaumann R, Dallacker-Losensky K, Rosenkranz C, Genzel GH, Stîngu CS, Schellenberger W, et al. Discrimination of human pathogen clostridium species especially of the heterogeneous C. sporogenes and C. botulinum by MALDI-TOF mass spectrometry. Curr Microbiol 2018;75:1506–15.

[46] Satten GA, Datta S, Moura H, Woolfitt AR, Carvalho MdG, Carlone GM, et al. Standardization and denoising algorithms for mass spectra to classify whole-organism bacterial specimens. Bioinformatics 2004;20:3128–36.

[47] Delavy M, Cerutti L, Croxatto A, Prod'hom G, Sanglard D, Greub G, et al. Machine learning approach for candida albicans fluconazole resistance detection using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Front Microbiol 2019;10:3000.

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

[49] Rodrigues C, Passet V, Rakotondrasoa A, Brisse S. Identification of Klebsiella pneumoniae, Klebsiella quasipneumoniae, Klebsiella variicola and related phylogroups by MALDI-TOF mass spectrometry. Front Microbiol 2018;9:3000.

[50] Asakura K, Azechi T, Sasano H, Matsui H, Hanaki H, Miyazaki M, 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 2018;13:e0194212.

[51] Ho P-L, Yau C-Y, Ho L-Y, Chen JHK, Lai ELY, Lo SWU, et al. Rapid detection of cfiA metallo-β-lactamase-producing Bacteroides fragilis by the combination of MALDI-TOF MS and CarbaNP. J Clin Pathol 2017;70:868–73.

[52] Tang W, Ranganathan N, Shahrezaei V, Larrouy-Maumus G. MALDI-TOF mass spectrometry on intact bacteria combined with a refined analysis framework allows accurate classification of MSSA and MRSA. PLoS One 2019;14:e0218951.

[53] Wang H-Y, Chen C-H, Lee T-Y, Horng J-T, Liu T-P, Tseng Y-J, et al. Rapid detection of heterogeneous vancomycin-intermediate Staphylococcus aureus based on matrix-assisted laser desorption ionization time-of-flight: using a machine learning approach and unbiased validation. Front Microbiol 2018;9:2393.

[54] Martiny D, Busson L, Wybo I, El Haj RA, Dediste A, Vandenberg O. Comparison of the Microflex LT and Vitek MS systems for routine identification of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 2012;50:1313–25.

[55] Lodise Jr TP, McKinnon PS. Burden of methicillin-resistant Staphylococcus aureus: focus on clinical and economic outcomes. Pharmacother J Hum Pharmacol Drug Ther 2007;27:1001–12.

[56] Naber CK. Staphylococcus aureus bacteremia: epidemiology, pathophysiology, and management strategies. Clin Infect Dis 2009;48:S231–7.

[57] McGuinness WA, Malachowa N, DeLeo FR. Vancomycin resistance in Staphylococcus aureus. Yale J Biol Med 2017;90:269–81.
[58] Ayala J, Quesada A, Vadillo S, Criado J, Píriz S. Penicillin-binding proteins of Bacteroides fragilis and their role in the resistance to imipenem of clinical isolates. J Med Microbiol 2005;54:1055–64.
[59] Yamazoe K, Kato N, Kato H, Tanaka K, Katagiri Y, Watanabe K. Distribution of the cfiA gene among Bacteroides fragilis strains in Japan and relatedness of cfiA to imipenem resistance. Antimicrob Agents Chemother 1999;43:2808–10.
[60] Oberle M, Wohlwend N, Jonas D, Maurer FP, Jost G, Tschudin-Sutter S, et al. The technical and biological reproducibility of matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) based typing: employment of bioinformatics in a multicenter study. PLoS One 2016;11:e0164260.
[61] Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013;14:365–76.
[62] Perez-Riverol Y, Alpi E, Wang R, Hermjakob H, Vizcaíno JA. Making proteomics data accessible and reusable: current state of proteomics databases and repositories. Proteomics 2015;15:930–49.
[63] Lasch P, Stämmler M, Schneider A. A MALDI-TOF Mass spectrometry database for identification and classification of highly pathogenic microorganisms from the Robert Koch-Institute (RKI) [Data set]. Zenodo; 2016. http://doi.org/10.5281/zenodo.163517.
[64] Build software better, together. Github. https://github.com/
[65] Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data 2016;3(9). https://doi.org/10.1186/s40537-016-0043-6.
[66] Ruder S. An overview of multi-task learning in deep neural networks. 2017. arXiv:1706.05098v1 [cs.LG], https://arxiv.org/abs/1706.05098.
[67] Caruana R. Multitask learning. Learning to learn. Boston, MA: Springer; 1998. https://doi.org/10.1007/978-1-4615-5529-2_5. 978-1-4615-5529-2.
[68] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. 2014. arXiv:1409.7495 [stat.ML], https://arxiv.org/abs/1409.7495.