Confidence-based prediction of antibiotic resistance at the patient level

Downloaded from: https://research.chalmers.se, 2026-02-27 11:10 UTC

Citation for the original published paper (version of record):
Inda Diaz, J., Johnning, A., Hessel, M. et al (2026). Confidence-based prediction of antibiotic resistance at the patient level. mBio, 17(2): e0343125-. http://dx.doi.org/10.1128/mbio.03431-25

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio

Check for updates

| Computational Biology | Research Article

Confidence-based prediction of antibiotic resistance at the patient level

Juan S. Inda-Díaz,[1,2,3] Anna Johnning,[1,2,4] Magnus Hessel,[2,5] Anders Sjöberg,[4,6] Anna Lokrantz,[4] Lisa Helldal,[5] Mats Jirstrand,[4,6] Lennart Svensson,[6] Erik Kristiansson[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 14.

ABSTRACT Rapid and accurate diagnostics of bacterial infections are necessary for efficient treatment of antibiotic-resistant pathogens. Cultivation-based methods, such as antibiotic susceptibility testing (AST), are limited by bacterial growth rates and seldom yield results before treatment needs to start, increasing patient risk and contributing to antibiotic overprescription. Here, we present a deep-learning method that leverages patient data and available AST results to predict antibiotic susceptibilities that have not yet been measured. After training on three million AST results from 30 European countries, the method achieved an average accuracy of 93% across bacterial species and antibiotics. It predicted susceptibility with an average major error rate below 5% for quinolones, cephalosporins, and carbapenems, and below 8% and 14% for aminoglycosides and penicillins, respectively. Furthermore, the model predicted resistance with an average very major error rate below 10% for cephalosporins, carbapenems, and aminoglycosides, but with higher very major error rates for penicillins and quinolones. We combined the method with conformal prediction and demonstrated accurate estimation of the predictive uncertainty at the patient level. Our results suggest that artificial intelligence-based decision support may offer new means to meet the growing burden of antibiotic resistance.

IMPORTANCE Improved diagnostic tools are vital for maintaining efficient treatment of antibiotic-resistant bacteria and for reducing antibiotic overconsumption. In our research, we introduce a new deep learning-based method capable of predicting untested antibiotic resistance phenotypes. The method uses transformers, a powerful artificial intelligence (AI) technique that efficiently leverages both antibiotic susceptibility tests (AST) and patient data simultaneously. The model produces predictions that can be used as time- and cost-efficient alternatives to results from cultivation-based diagnostic assays. Significantly, our study highlights the potential of AI technologies to address the increasing prevalence of antibiotic-resistant bacterial infections.

KEYWORDS antibiotic resistance, diagnostics, antibiotic susceptibility testing, transformers, conformal prediction

The global rise of antibiotic-resistant bacterial infections threatens human health globally (1). In 2019, almost five million yearly deaths were attributed to antibiotic-resistant bacteria (2), a number that is expected to continue to grow in the coming decades (3). Reduced antibiotic efficacy in treatment increases the risk of performing vital healthcare procedures—including surgery, chemotherapy, and organ transplantation (4)—and, thereby, jeopardizes modern medicine as a whole.

Accurate and fast diagnostics are necessary for effective treatment of antibiotic-resistant bacteria. A central method is antibiotic susceptibility testing (AST), a laboratory test in which a bacterium isolated from a patient sample is cultivated, and its resistance

Editor Alejandro J. Vila, Instituto de Biologia Molecular y Celular de Rosario, Rosario, Santa Fe, Argentina

Address correspondence to Juan S. Inda-Díaz, inda@chalmers.se, or Erik Kristiansson, erik.kristiansson@chalmers.se.

The authors declare no conflict of interest.

See the funding table on p. 15.

Received 9 November 2025
Accepted 16 December 2025
Published 23 January 2026

Copyright © 2026 Inda-Díaz et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

phenotype is assessed in the presence of antibiotics (5). However, AST can be time consuming due to the often low bacterial growth rate and the large number of antibiotics that may need to be tested for highly multidrug-resistant isolates. Tier-based testing strategies and time constraints usually yield incomplete AST results initially. For life-threatening infections, treatment needs to start as early as possible, often before AST results are available (6). Under these circumstances, the choice of treatment is reduced to educated guesses based on limited diagnostic information (7). This form of "empirical" treatment is associated with increased patient risks and overprescription of antibiotics (8–10).

Antibiotic resistance is commonly caused by resistance genes encoding various defense mechanisms. These genes are often co-localized on mobile genetic elements, such as plasmids and transposons (11, 12). Multiple resistance genes can thus be transferred simultaneously between bacterial cells, giving rise to strong correlations in susceptibility to different antibiotics. The infecting bacterium and, therefore, its susceptibility profile are, furthermore, dependent on patient characteristics, including age, sex, and the geographical region where the infection was acquired (13–15). Indeed, patient data have previously been shown to contain valuable information for selecting suitable antibiotic therapy for bacterial urinary tract infections (15–17). There are, however, no artificial intelligence (AI)-based methods that can combine patient data with the initial, often incomplete AST results and thus guide treatment choice based on all available diagnostic information. Indeed, integrating patient data with available AST results could enable more accurate prediction of susceptibility to antibiotics that have not been tested, thereby providing physicians with comprehensive diagnostic information at an earlier point in time.

AI and deep learning have been successfully applied to diagnostics (18, 19), but the focus has primarily been on image-based data commonly used in radiology and pathology (20). In contrast, methods for non-image multimodal data, which are more prevalent in the diagnosis of infectious diseases, have received less attention (21). Several methods for predicting phenotypes from genotypic data have also been proposed (22), as have machine-learning methods using electronic health records and patient data (23). There are, however, few AI-based decision support systems for antibiotic treatment selection in use in clinical settings (24). A major culprit in the development of such methods is the complexity of the diagnostic data, which is typically categorical (stratified test results and patient data) and contains dependencies and redundancies. Also, when applied to antibiotic-resistant bacteria, the incompleteness of the diagnostic data, particularly AST results, requires approaches that can efficiently handle missing observations. Furthermore, since model accuracy depends on the amount of available information, any prediction must be accompanied by estimates of its uncertainty. Indeed, the possibility of disregarding predictions that are insufficiently confident is vital for critical decision-making and, thus, essential for the adoption of AI-based methodologies in healthcare settings (25). However, today, most AI-based diagnostic methods are primarily evaluated in populations and do not provide uncertainty information for patient-level predictions (26). Existing methods have, so far, been unable to adequately address these challenges and are either limited to a few specific antibiotics, cannot incorporate missing AST data, or do not provide any estimates of the predictive uncertainty (27, 28).

In recent years, transformer-based models, such as BERT (bidirectional encoder representations from transformers) (29) and GPT (generative pre-trained transformer) (30), have transformed natural language processing. These models operate on categorical input data, often structured into sentences of words, and subject them to multi-head self-attention (31). This process enables inference of complex word dependencies directly from data and thereby predicts missing parts. Therefore, we hypothesized that transformers may be suitable for predicting antibiotic susceptibility results from a combination of incomplete diagnostic information and patient data. Multimodal self-attention would enable the identification of complex dependencies among

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

diagnostic data types and facilitate extrapolation to susceptibilities that have not been tested. Transformers have previously been shown to be highly useful beyond natural language processing (32, 33) but are rarely used for diagnosing infectious diseases.

In this study, we present a novel transformer-based method that accurately predicts antibiotic susceptibilities using patient data and incomplete diagnostic information for three common bacterial pathogens, *Escherichia coli*, *Klebsiella pneumoniae*, and *Pseudomonas aeruginosa*. We combine the model with conditional inductive conformal prediction (34) to estimate the uncertainty of each prediction at the patient level, thereby enabling dismissal of predictions with too low certainty. The model was trained and evaluated on a large, heterogeneous data set containing AST results from blood infections collected during routine care across 30 European countries. Our results showed that the model could accurately predict susceptibility to a wide range of antibiotics, even when only a few AST results were included in the input. We also show that the model handles increasing information well, providing more accurate predictions as additional AST results become available. Finally, we demonstrate that predictions can be made within predefined confidence levels for most bacterial species and antibiotics, which allows control of both the major error (ME) and very major error (VME) rates. We conclude that the combination of transformers and conditional inductive conformal predictions constitutes an appealing class of models for integrating and predicting diagnostic information.

RESULTS

A transformer-based model for antibiotic susceptibility prediction

We developed a transformer-based model to predict unavailable diagnostic information using multiple classifications. The input to the model is a set of words containing the available diagnostic information (results of AST) and patient data (Fig. 1). The AST results are assumed to be incomplete, where only a subset of the possible tests has available results. The model uses two transformers, which summarize the patient and AST information in two sentence-specific classification (CLS) vectors, one representing patient-specific information and one representing bacterial isolate-specific information. These vectors then serve as input to multiple antibiotic-specific feed-forward neural networks, each predicting the probability of susceptibility to the corresponding antibiotic. The uncertainty is estimated using inductive conformal prediction, conditioned on the susceptibility, thereby controlling the false-positive (ME) and false-negative (VME) rates for each antibiotic (34, 35). The full details of the model architecture and the uncertainty estimation are provided in Materials and Methods.

The model was trained and evaluated on data from the European Surveillance System (TESSy; https://www.ecdc.europa.eu/en/publications-data/european-surveillance-system-tessy), which contains AST results for 1,161,303 *E. coli*, 301,506 *K. pneumoniae*, and 166,448 *P. aeruginosa* isolates collected from blood infections of patients in 30 European countries. The training and evaluation were performed on 20 (16, 18, and 11 for *E. coli*, *K. pneumoniae*, and *P. aeruginosa*, respectively) commonly used antibiotics that belong to five large and clinically relevant antibiotic classes: aminoglycosides, carbapenems, cephalosporins, penicillins, and quinolones (Table 1). A bacterial isolate was, on average, tested for susceptibility to 7.3, 8.8, and 7.8 antibiotics, with standard deviations of 1.9, 2.4, and 1.7 for *E. coli*, *K. pneumoniae*, and *P. aeruginosa*, respectively (Fig. S1 and S2). The most commonly tested antibiotics were ceftazidime and ciprofloxacin, for which at least 93% of isolates were tested, regardless of pathogen, country of origin, or patient gender. In contrast, less than 10% of the *E. coli* and *K. pneumoniae* isolates were tested for piperacillin (PIP), moxifloxacin, and nalidixic acid (Fig. S3). The susceptibility rate for *E. coli* isolates was lowest for the penicillins amoxicillin (AMX), ampicillin (AMP), and PIP (43%–50%), whereas for *K. pneumoniae* isolates, it was lowest for PIP (20%). For other antibiotics and *P. aeruginosa* isolates, the susceptibility rates were higher, representing a more unbalanced data set. A slightly higher susceptibility rate was observed among isolates collected from female patients (Fig. S4).

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

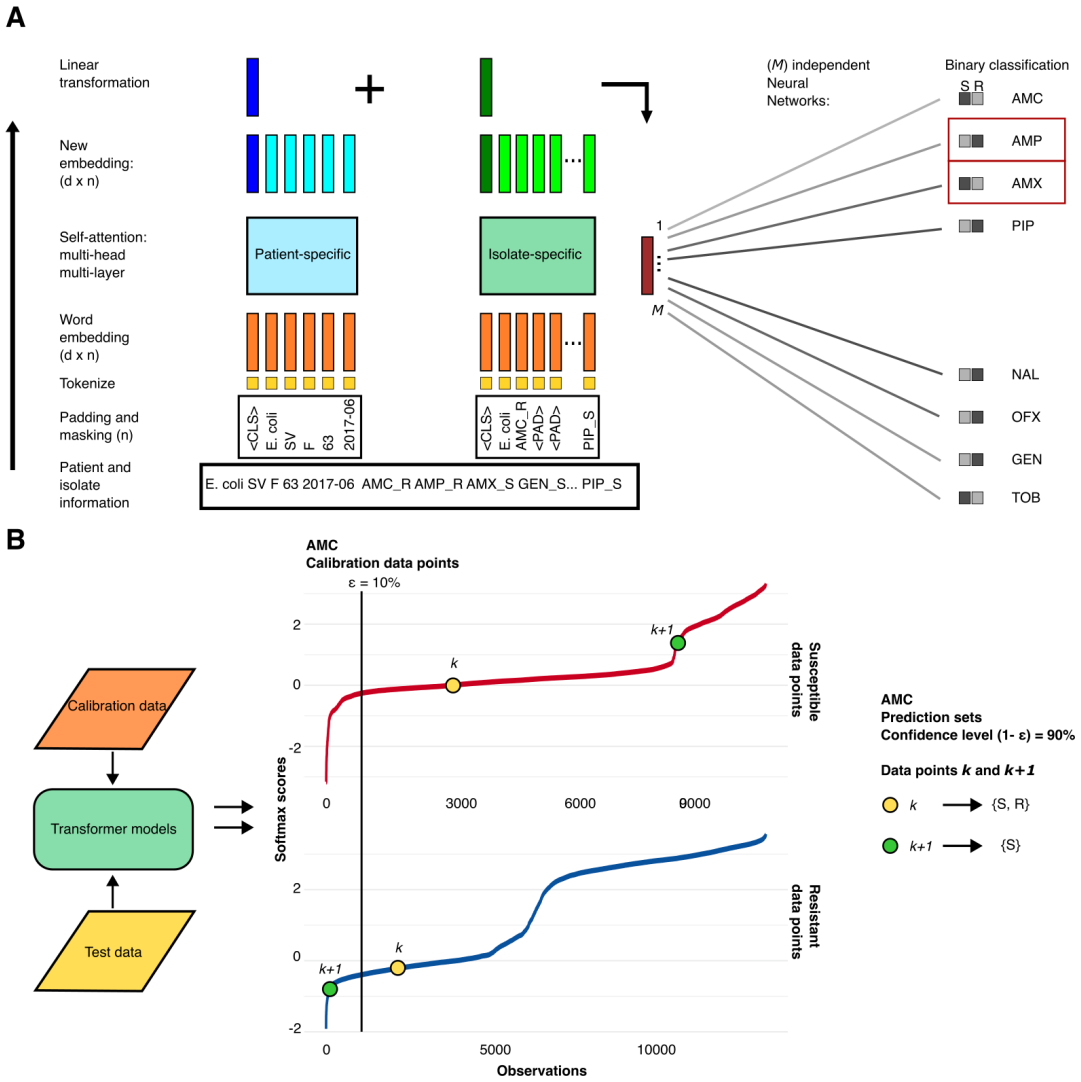

FIG 1 The proposed model architecture and uncertainty control. (A) The architecture of the proposed model. Both input sequences start with a classification word, *CLS*, followed by the bacterial species name. One of the input sequences also contains patient information (country, gender, age, and sampling date), while the other includes available AST data. Both input sequences are fixed to lengths $L = 5$ for the patient information and $L = 21$ for the AST data, padded with the *PAD* word. We use a linear embedding to represent the input words numerically, which are fed into the transformer encoders. The first vectors (*CLS*) from the outputs of each encoder are combined into a linear model and fed to M independent neural networks, each representing one antibiotic. The neural network outputs are two-dimensional vectors indicating susceptibility and resistance, respectively. (B) Uncertainty control. The neural network outputs undergo a softmax rescaling. A calibration data set is used to build empirical distributions of conformity for resistant and susceptible predictions for each pathogen and antibiotic. The prediction regions for the data points k and $k + 1$ are built based on the deviation of the observed softmax score from the empirical distribution and the confidence interval threshold. See Materials and Methods for full details.

Predictions of antibiotic susceptibility have high performance

To assess how well the model predicts the susceptibility of antibiotics not included as inputs, we used 70% of the bacterial isolates for fivefold cross-validation and model training. We used 15% of the isolates for testing and the remaining 15% for calibration of conformal prediction. During training, calibration, and testing, we selected a random number of antibiotics as input to the model (distribution available in Fig. S5; mean 5.7, 6.3, and 5.7 with standard deviations of 1.3, 1.5, and 1.1 for *E. coli*, *K. pneumoniae*, and *P. aeruginosa* isolates, respectively; see Materials and Methods for details) together with patient data. The susceptibility of the remaining tested antibiotics was assumed to be unknown and, therefore, masked from the input data. The model's predictions were

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

TABLE 1 Bacterial isolate summary[a]

Antibiotic	*E. coli*		*K. pneumoniae*		*P. aeruginosa*		Loss function weights (S/R)[c]
	Isolates	%R[b]	Isolates	%R[b]	Isolates	%R[b]	
Amoxicillin/clavulanic acid	550,937	33%	129,537	42%	–	–	0.45/0.55
Ampicillin	800,283	57%	–	–	–	–	0.45/0.55
Amoxicillin	293,369	55%	–	–	–	–	0.45/0.55
Piperacillin	93,919	50%	21,028	80%	62,218	22%	0.45/0.55
Piperacillin/tazobactam	646,892	8%	164,551	29%	152,233	16%	0.15/85
Ceftazidime	1,080,891	10%	280,916	30%	159,118	15%	0.3/0.7
Ceftriaxone	350,947	11%	90,851	30%	–	–	0.3/0.7
Cefotaxime	990,234	11%	253,971	30%	–	–	0.3/0.7
Cefepime	336,526	10%	102,213	33%	65,619	16%	0.3/0.7
Ciprofloxacin	1,117,333	22%	286,899	31%	160,669	20%	0.3/0.7
Levofloxacin	291,403	23%	73,574	29%	45,636	24%	0.3/0.7
Moxifloxacin	105,648	26%	23,231	26%	–	–	0.3/0.7
Nalidixic acid	63,468	27%	16,532	36%	–	–	0.45/0.55
Ofloxacin	136,704	19%	33,764	36%	–	–	0.3/0.7
Amikacin	–	–	202,968	9%	130,264	9%	0.15/85
Gentamicin	1,073,614	9%	278,400	20%	143,645	14%	0.15/85
Tobramycin	564,295	10%	145,914	27%	119,954	13%	0.15/85
Ertapenem	–	–	109,997	14%	–	–	0.15/85
Imipenem	–	–	189,733	10%	121,852	20%	0.15/85
Meropenem	–	–	242,732	9%	130,352	15%	0.15/85

[a]The included antibiotics, the number of isolates from each species tested for each antibiotic, their resistance rates, and the weights used in the loss function. Dash (–) indicates combinations of bacterial species and antibiotics that the model does not cover.
[b]Proportion of bacterial isolates resistant to each of the antibiotics.
[c]Weight on the cross-entropy loss function for the two different labels: susceptible/resistant.

then compared with the antibiotics masked from the input to evaluate performance. In this setting, the model had an overall high performance that did not differ substantially between folds in the cross-validation, with a standard error (se) below 0.015 for all antibiotics and bacterial species except the ME for the antibiotic PIP in *K. pneumoniae* (se = 0.016 for both training and test data sets). The performance was largely consistent across the training and test processes (Fig. 2; Fig. S6). There were, however, apparent differences in performance between the three bacterial species and between antibiotics. The F1 score (the harmonic mean of precision and recall) in the test data set for *E. coli and K. pneumoniae* isolates was highest for cephalosporins (91% and 96% on average, respectively), and quinolones (86% and 93% on average, respectively), though lower for penicillins (74% and 88% on average, respectively, Fig. 2A). For *P. aeruginosa* isolates, the F1 score was 84% for penicillins and 82% for quinolones and cephalosporins. Carbapenems had an average F1 score of 90% and 76% for *K. pneumoniae* and *P. aeruginosa* isolates, respectively, while aminoglycosides had the lowest F1 scores: 65% for *E. coli* and 76% for *K. pneumoniae* and *P. aeruginosa* isolates.

Next, we evaluated the model based on the ME rate, defined as the proportion of susceptible bacterial isolates erroneously predicted as resistant, and the VME rate, defined as the proportion of resistant isolates erroneously predicted as susceptible. The ME and VME rates are standard performance measures in AST and are frequently used to evaluate and compare diagnostic methods (36). Based on the test data set, cephalosporins had, on average over all predictions in the test data, an ME rate of 1.4% (1%–1.8%), 2.3% (1.5%–3.9%), and 5.2% (5.1%–5.2%) for *E. coli*, *K. pneumoniae*, and *P. aeruginosa* isolates, respectively, while the corresponding VME rate was, on average, 6.7% (5.6%–9.4%), 2.3% (1.2%–3.2%), and 12.2% (9.7%–12.8%; Fig. 2B). For quinolones, the average ME rate was 2.2% (2%–5%), 2.9% (1.6%–4.7%), and 3.3% (2.5%–4.2%), while the VME rate was, on average, 19.2% (8.9%–28.6%), 8.3% (4.2%–13.4%), and 21.9% (20.3%–23.5%) for *E. coli*, *K. pneumoniae*, and *P. aeruginosa* isolates, respectively. For *E. coli* isolates, the penicillins had a higher overall average ME rate of 12.5% (4.1%–16.6%). Here, PIP had the

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

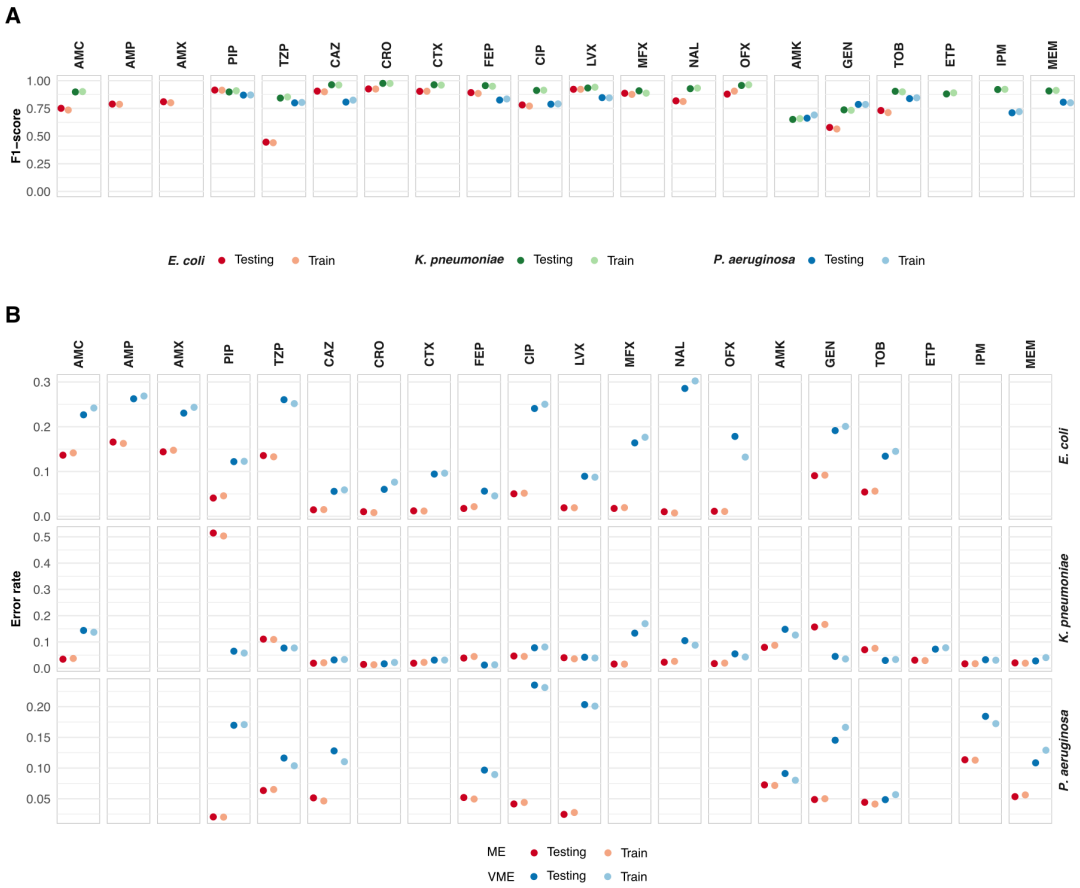

FIG 2 Performance of the transformer model. Results from the train and test data sets: (A) F1-score, (B) ME rate, and VME rate for the transformer model, shown for each antibiotic and pathogen. Note that, due to data restrictions, not all antibiotics are assessed for all three pathogens.

lowest VME rate (12.2%), while the other penicillins had rates between 23% and 26%. For *K. pneumoniae* and *P. aeruginosa* isolates, the average ME for penicillins, excluding PIP, was 7.3% and 6.3%, respectively, while the VME was 11%. We noted that the model struggled to predict PIP resistance in *K. pneumoniae* with an ME as high as 51%. This may be a consequence of data imbalance, where *K. pneumoniae* had a resistance rate of 80% to PIP (compared to 50% and 22% for *E. coli* and *P. aeruginosa*, respectively) but accounted for only 11% of the data for this antibiotic. Aminoglycosides had average ME/VME rates of 7.3/16.3, 10.2/7.4, and 5.5/9.5 for *E. coli*, *K. pneumoniae*, and *P. aeruginosa* isolates, respectively. Finally, for carbapenems, the ME and VME rates were 2.3% (1.7%–3.1%) and 4.4% (2.8%–7.3%) for *K. pneumoniae* isolates, respectively, and 8.4% (5.3%–11.4%) and 14.6% (10.9%–18.4%) for *P. aeruginosa* isolates, respectively.

The number of available AST results influenced the model's performance. When the number of AST results used as input to the model increased from four to eight antibiotics, significant reductions were seen in the ME rate for penicillins, where AMX and AMP dropped from 34% and 22% to 3% and 7%, respectively, and the VME of CIP, which fell from 36% to 7% in *E. coli* isolates (Fig. 3D). Interestingly, the drop for other antibiotics and pathogens was more modest or non-existent. A substantial reduction in the VME rate was also observed across most antibiotics and pathogens.

When eight AST results were included in the input, 14 of 16 antibiotics in *E. coli* isolates, 16 of 20 antibiotics in *K. pneumoniae*, and all antibiotics in *P. aeruginosa* had an ME rate below 10%, and below 5% for 11, 13, and 10 antibiotics in each pathogen, respectively. Similarly, the VME rate was below 10% for 9, 16, and 10, and below 5% for 6, 12, and 3 of the antibiotics corresponding to each pathogen. Going from four to eight input AST results increased the F1 scores from 75% to 87% for *E. coli*, from 85% to 92% for

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

K. pneumoniae, and from 75% to 84% for *P. aeruginosa*, on average. The model's overall high performance, as indicated by its F1 score, ME, and VME rates, suggests that it can be used to predict complete susceptibility patterns for bacterial isolates accurately.

Control of the ME and VME rates at the patient level

In clinical practice, decisions are taken at the patient level. It is thus vital that decision support based on predictions also conveys information about the confidence in the prediction. Indeed, if the uncertainty is too high, the prediction may need to be considered with care or completely dismissed until more diagnostic information becomes available. Therefore, we implemented an algorithm that assigns a quantitative measure of uncertainty to each individual prediction (34). The algorithm was based on conditional inductive conformal prediction, where the certainty of each prediction was derived from a conformity measure, defined in our case as the softmax score from each antibiotic-specific neural network (Fig. 1B). The distributions of softmax scores for both susceptible and resistant bacterial isolates were empirically calculated for each antibiotic and each of the three pathogens, using a dedicated calibration data set. These distributions were used to determine whether a new prediction aligns with the susceptible and/or resistant isolates in the calibration data set and, based on a predefined confidence level, is deemed sufficiently certain. The output of a final prediction for one antibiotic is a prediction set, with either (i) a single label, that is, either susceptible or resistant, when there is enough conformity to only one of the softmax distributions; (ii) multiple labels, that is, both susceptible and resistant, if there is enough conformity for both softmax distributions; or (iii) no label if there is no conformity to either group. The proportion of predictions that will, on average, have the correct label in their prediction set is governed by the confidence level $1 - \varepsilon$. Note that in this setting, ε corresponds to the average ME and the VME rates for susceptible and resistant bacterial isolates for each pathogen, respectively (see Materials and Methods for full details).

The empirically derived ME and VME rates were close to the prespecified values of ε for all antibiotics (Fig. S7 to S11). For example, at a confidence level of 90% ($\varepsilon = 0.1$), the observed ME rates were, on average, 10.1%, 9.9%, and 9.9% (standard deviation 0.4%, 1.3%, and 0.3%), and the observed VME rates were, on average, 9.7%, 9.4%, and 9.3% (standard deviation 1.43%, 0.95%, and 1.7%) for *E. coli*, *K. pneumoniae*, and *P. aeruginosa*

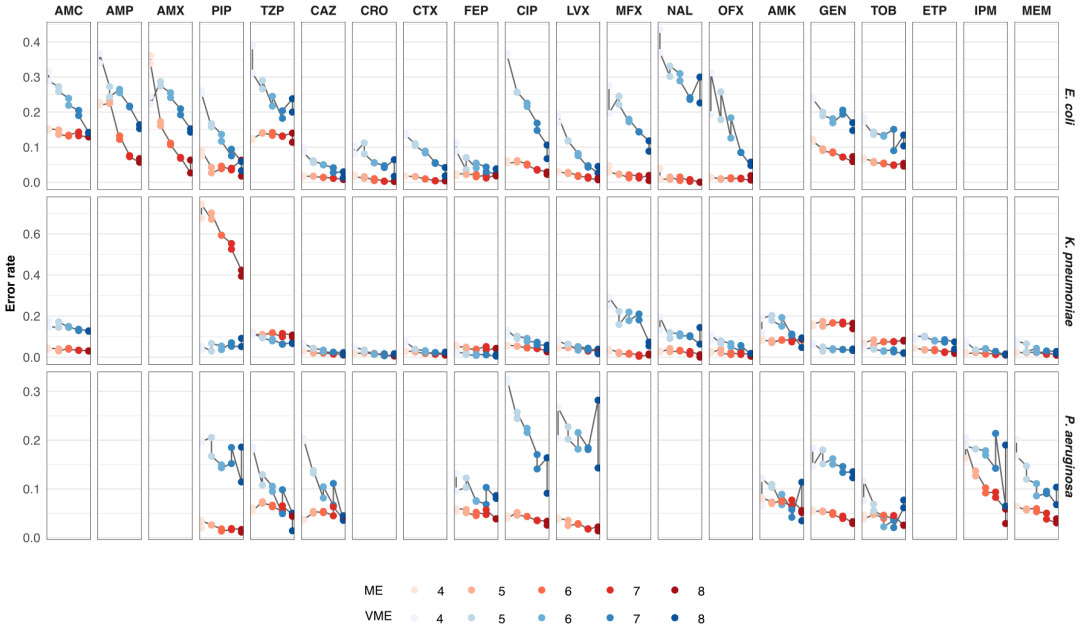

FIG 3 The predictive performance of the model as a function of the number of AST results included in the input. ME and VME rates from light to dark: 4–8 AST results are shown for each antibiotic and pathogen. Note that, due to data restrictions, not all antibiotics are assessed for all three pathogens.

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

isolates, respectively. The concordance between pre-specified and observed error rates was sustained at higher confidence levels where, for a confidence level of 95%, the average ME rates were 5.1%, 5.1%, and 5% (standard deviation 0.5%, 1%, and 0.3%), and the average VME rates were 5%, 4.7%, and 5% (standard deviation 0.8%, 0.6%, and 1.4%) for *E. coli*, *K. pneumoniae*, and *P. aeruginosa* isolates, respectively. Thus, the results showed that the specified confidence levels calculated from empirical distributions were sufficiently stable between data sets.

The model could confidently predict the phenotype for a large majority of bacterial isolates and antibiotics

At a confidence level of 90%, as many as 82%, 89%, and 86% of the predictions were unambiguous and correct (only the correct label was in the prediction set) for *E. coli*, *K. pneumoniae*, and *P. aeruginosa* isolates, respectively; however, this number varied between antibiotics (Fig. 4A, 5A and 6A). Cephalosporins had a high proportion of correct unambiguous predictions (90%) across the three pathogens, and quinolones had 90% for *K. pneumoniae*. For the remaining antibiotics, between 82% and 89% of the predictions were correct and unambiguous, except for penicillins for *E. coli* (71%) and quinolones for *P. aeruginosa*, which showed higher uncertainty. The proportion of unambiguous predictions was, as expected, reduced when the confidence level was increased to 78%, 89%, and 77% at a 95% confidence level and, finally, to 72%, 85%, and 63% at a 97.5% confidence level for *E. coli*, *K. pneumoniae*, and *P. aeruginosa*, respectively. On the other hand, the number of unambiguous predictions increased as more diagnostic information was provided to the model. At a 90% confidence level, the proportion of correct unambiguous predictions increased from, on average, 76%, 84%, and 83% when four AST results were included in the input to 89%, 91%, and 91% when eight AST results were included in the input for *E. coli*, *K. pneumoniae*, and *P. aeruginosa* isolates, respectively (Fig. 4B, 5B and 6B). The increase was substantial for predicting susceptibility to the penicillins AMP and AMX in *E. coli* isolates, to PIP and ofloxacin in *K. pneumoniae* isolates, and for predicting susceptibility and resistance to CIP, as well as susceptibility to levofloxacin (LVX) and IPM in *P. aeruginosa* isolates. With eight input AST results, the model achieved 90% correct, unambiguous predictions for most of the remaining masked antibiotics. This was also true for higher confidence levels, where the proportion of unambiguous and correct predictions with eight input AST results was 89% and 84% for 95% and 97.5% confidence, respectively.

DISCUSSION

In this study, we present a method that uses a transformer model to predict unavailable diagnostic information. When combined with conditional inductive conformal prediction to estimate the predictive uncertainty at the patient level, the method can abstain from making decisions unless the confidence is deemed sufficiently high. The model was applied to the diagnostics of infectious bacteria, a field facing rapidly rising societal and economic costs due to the growing challenges related to antibiotic resistance (37, 38). We showed that only a few AST results are sufficient to accurately derive a more complete resistance profile. This has the potential to guide antibiotic therapy for serious infections, where treatment must begin before all diagnostic tests are completed.

The method was trained on a large, heterogeneous data set comprising over 12 million AST results from more than a million bacterial isolates of three pathogens collected from 30 European countries. This data set was compiled for surveillance purposes and contains AST results originally produced in routine diagnostics. Validation, in which AST results were randomly excluded, showed that the model had generally high accuracy in predicting antibiotic susceptibility. The performance improved further as more ASTs were included in the input data, demonstrating that the model could efficiently incorporate diagnostic information as it becomes available and, thus, over time, produce more certain predictions. Indeed, when AST for eight antibiotics was used as input, the model could predict most susceptibilities with a VME rate (false-negative

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

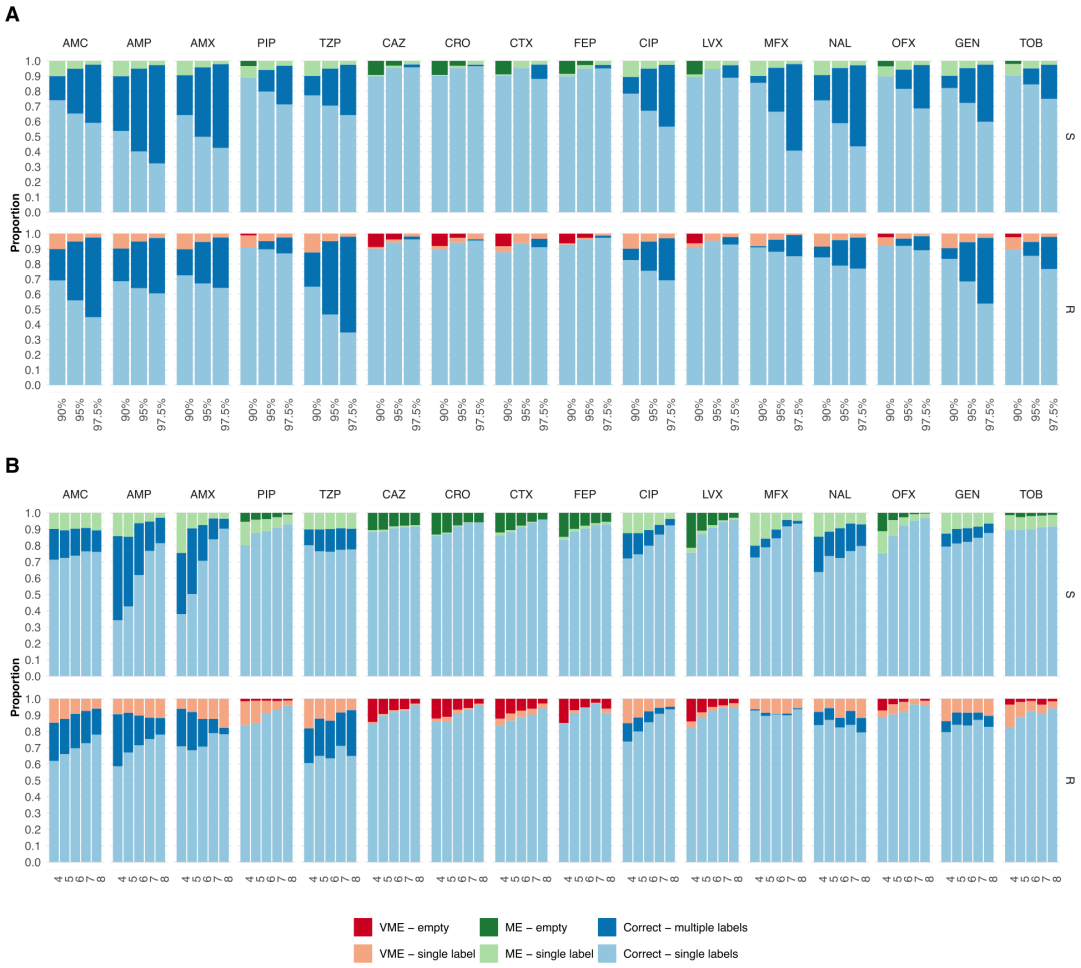

FIG 4 The proportion of correct predictions as a function of the number of labels in the prediction set for *E. coli* isolates. The proportion of correct predictions with a single label (light) and multiple labels (dark), MEs, and VMEs with a single label (light) and empty set (dark) predictions for resistant (R) and susceptible (S) isolates, shown for each antibiotic and pathogen using three different confidence levels: 90%, 95%, and 97.5%. (A) The proportions are shown for each antibiotic at three confidence levels (90%, 95%, and 97.5%), considering all possible numbers of input AST results. (B) The proportions are shown as a function of the number of input AST results (90% confidence level).

rate) as low as 10%. This indicates that AI prediction can serve as a viable complement to laboratory diagnostic tests and may be utilized to save time, alleviate suffering, and reduce economic costs.

Providing information about predictive uncertainty is crucial in diagnostics, where population-level knowledge is used to predict outcomes for individual patients. We addressed this challenge by implementing an algorithm based on conditional inductive conformal prediction (34, 39, 40) to provide each prediction with an accompanying confidence score. In our application, the predefined confidence corresponded to ME and VME rates, yielding predictions that limit false-positive and false-negative rates to the desired levels. During model testing, we found consistent error rates across data sets, and for cephalosporins and quinolones, the prediction sets mainly contained a single, correct label. However, a higher variation was observed for penicillins, resulting in a larger proportion of ambiguous predictions (multiple labels). Furthermore, our implementation separates the uncertainty for susceptible and resistant predictions. This is valuable in the diagnostics of bacterial infections, where a VME, that is, the incorrect identification of a resistant isolate, may lead to ineffective antibiotic treatment. Therefore, VMEs are often considered to be the most serious errors, especially for life-threatening infections. The

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

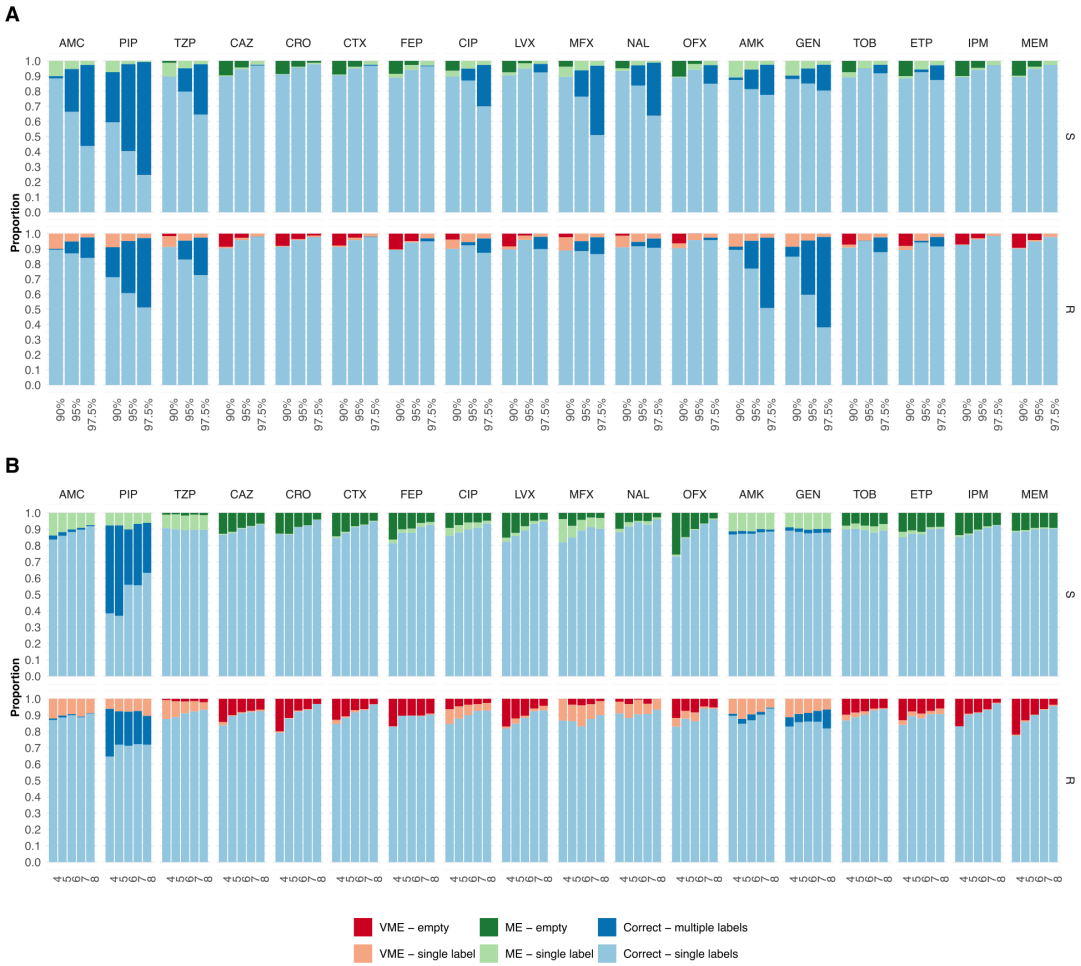

FIG 5 The proportion of correct predictions as a function of the number of labels in the prediction set for *K. pneumoniae* isolates. The proportion of correct predictions with a single label (light) and multiple labels (dark), MEs, and VMEs with a single label (light) and empty set (dark) predictions for resistant (R) and susceptible (S) isolates, shown for each antibiotic and pathogen using three different confidence levels: 90%, 95%, and 97.5%. (A) The proportions are shown for each antibiotic at three confidence levels (90%, 95%, and 97.5%), considering all possible numbers of input AST results. (B) The proportions are shown as a function of the number of input AST results (90% confidence level).

ability to set confidence scores for susceptible and resistant predictions individually enables the model to be adjusted for different clinical scenarios.

The model showed a clear difference in predictive performance between antibiotics, notoriously lower for penicillins, especially compared to cephalosporins. Historically, penicillins were among the earliest classes of antibiotics and the first beta-lactam antibiotics introduced. Resistance to beta-lactam antibiotics is typically caused by enzymes that hydrolyze the antibiotics (41). Penicillins have been widely used for over 80 years, and bacteria can consequently acquire a wide diversity of resistance genes (42). In contrast, resistance to cephalosporins—a later generation of beta-lactam antibiotics—is, to a larger extent, dependent on additional genetic events, such as mutations in chromosomal genes or the acquisition of broader-spectrum resistance mechanisms (41). These patterns were reflected in the data, where resistance to penicillins was common. Indeed, as many as 87.7% of the *E. coli* and 61% of the *K. pneumoniae* bacterial isolates resistant to a single antibiotic were resistant to a penicillin. Furthermore, it is plausible that the order of evolution of multidrug-resistant bacteria affects the performance of our model. We also consider resistance phenotypes that are initially acquired to be harder to predict than those that commonly appear in later stages of evolution, simply because there are no other correlating susceptibilities. Results from additional diagnostic assays, such as targeted molecular tests that detect antibiotic resistance

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

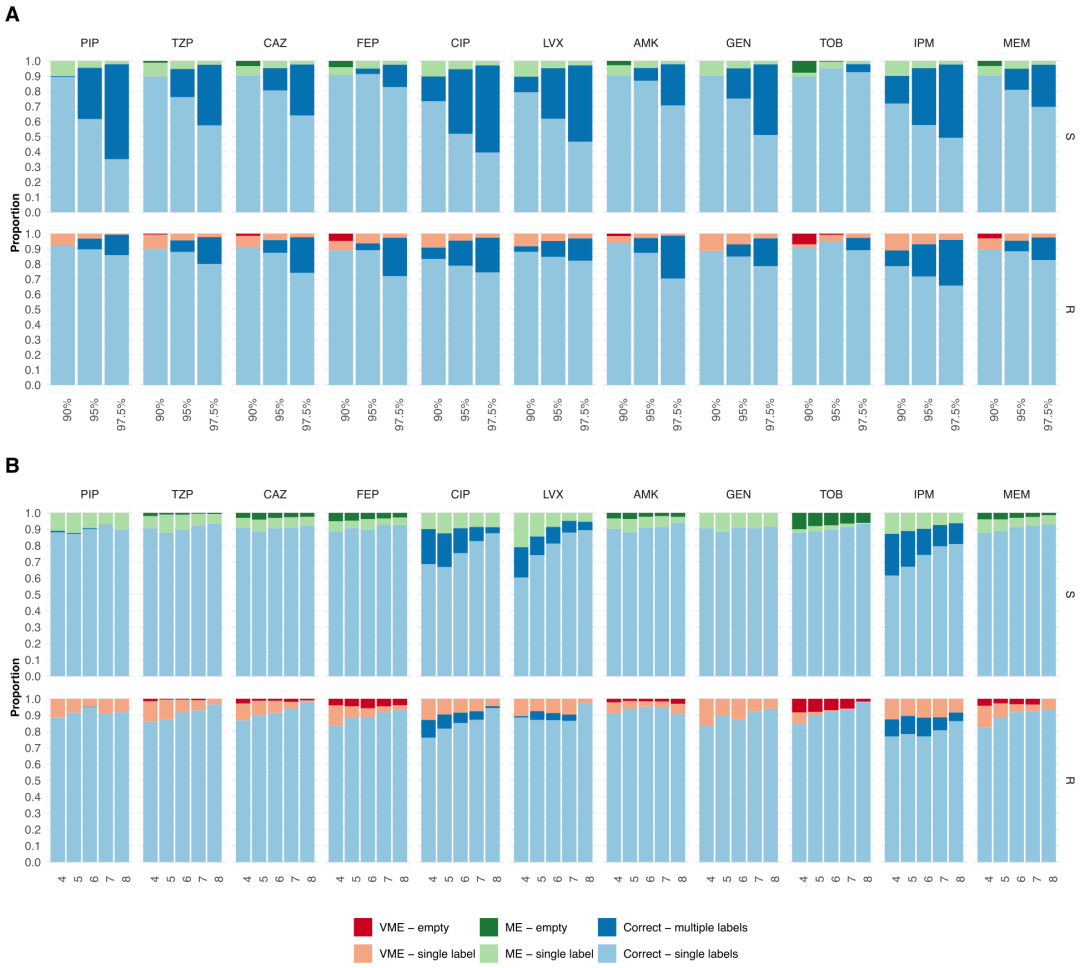

FIG 6 The proportion of correct predictions with respect to the number of labels in the prediction set for *P. aeruginosa* isolates. The proportion of correct predictions with a single label (light) and multiple labels (dark), MEs, and VMEs with a single label (light) and empty set (dark) predictions for resistant (R) and susceptible (S) isolates, shown for each antibiotic and pathogen using three different confidence levels: 90%, 95%, and 97.5%. (A) The proportions are shown for each antibiotic at three confidence levels (90%, 95%, and 97.5%), considering all possible numbers of input AST results. (B) The proportions are shown as a function of the number of input AST results (90% confidence level).

genes and genomic data, could further improve performance. Indeed, the flexibility of the transformers allows incorporating other types of diagnostic information, including genotypic information, as new words in the input sentence.

Ultimately, the AI methodology presented here has the potential to enhance the diagnosis of infections caused by antibiotic-resistant bacteria. We argue that data-driven methods have the potential to replace selected diagnostic assays, thereby providing physicians with more comprehensive decision support at an earlier stage. This has the potential to improve the treatment of bacterial infections, thereby decreasing patient morbidity and mortality, reducing costs, and limiting the overprescription of antibiotics.

MATERIALS AND METHODS

Data description

This study is based on data from the European Surveillance System (TESSy), collected as part of surveillance conducted by the European Centre for Disease Prevention and Control. The data set contains more than 12 million AST results for bacteria isolated from blood and cerebrospinal fluid of hospitalized patients across 30 European countries. For

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

each bacterial isolate, we retrieved the species name, AST results, the patient's gender and age, the date of bacterial isolation, and the country where the test was performed. This metadata will be referred to collectively as the patient data. The analysis was limited to the susceptibility of *E. coli*, *K. pneumoniae*, and *P. aeruginosa* isolates, which were the most common species, from blood infections collected between 2007 and 2020. Tests that resulted in either a resistant (R) or a susceptible (S) result were included, while tests with an I result were excluded. Only antibiotics with a resistance rate of at least 8% were considered. This resulted in data covering 16 antibiotics from five clinically relevant classes: aminoglycosides, cephalosporins, penicillins, quinolones, and carbapenems (Table 1). Furthermore, isolates tested for fewer than five antibiotics were removed. If an antibiotic was tested multiple times for the same bacterial isolate, only the most recent test was included. The final data set contained 1,629,257 bacterial isolates with 12,433,856 AST results, resulting in an average of 7.6 tests (standard deviation of 2.1) per isolate. After randomization, the bacterial isolates were divided into three datasets: a training data set (70%), a calibration data set (15%), and a test data set (15%). The train data set was further divided into five groups for fivefold cross-validation.

Data expansion

To increase the number of AST result combinations, each data set was expanded. For each bacterial isolate *j*, multiple data points were generated by randomly splitting the susceptibility test results into two groups. The first group, together with the patent information, x_j, was considered known and, thus, used as input to the model. The second group, y_j, was hidden from the model and used to evaluate the predictive performance. The full details of the data expansion are provided in Table S1. A made-up example of the information available for a bacterial isolate is presented below.

Available information for the bacterial isolate *j*: "ESCCOL SV 30 M 2013_01 LVX_R AMC_S AMP_S TZP_R CTX_S GEN_S", represents an *E. coli* bacterium isolated at a Swedish hospital (SV) from a 30-year-old (30) male (M) patient in January 2013 (2013_01) where the isolated bacterium was tested against six antibiotics. The AST results indicated resistance to the antibiotics LVX and PIP/tazobactam and susceptibility to amoxicillin/clavulanic acid, AMP, cefotaxime, and gentamicin. Two example data points, (x_k, y_k) and $(x_{k'}, y_{k'})$, that could potentially be created from this isolate:

$$(x_k, y_k) = (\text{"ESCCOL SV 30 M 2013\_01 LVX\_R AMC\_S AMP\_S TZP\_R"}, \text{"CTX\_S GEN\_S"}),$$
$$(x_{k'}, y_{k'}) = (\text{"ESCCOL SV 30 M 2013\_01 LVX\_R AMC\_S AMP\_S CTX\_S GEN\_S"}, \text{"TZP\_R"}).$$

Model architecture and training

Given a data point, (x_k, y_k), the model takes the input x_k and makes predictions $\hat{y}_k$ of the susceptibilities y_k. The input sentence x_k is split into patient data and AST data, both of which contain the word for the bacterial species, and both are complemented at the start with the classification word, *CLS*. The patient data and the AST data are padded to a length $L = 6$ and $L = 21$ with *PAD* words if needed. Each word is then converted into a linear embedding representation in the form of a *d*-dimensional vector that provides semantic meaning to the model ($d = 128$). These word embeddings are passed through two transformer encoder layers, each with two attention heads, followed by an add-and-normalize layer, a position-wise feed-forward layer (using 256 nodes), and, finally, another add-and-normalize layer. The first vector of the output from each encoder—representing the *CLS* word and containing information at the sentence level—is concatenated and fed to a linear transformation. The output is used as input to $M = 20$ independent antibiotic-specific feed-forward networks, each of depth two. The intermediate layer of the network has 128 nodes with a rectified linear unit activation function, followed by a normalization step, while the final layer outputs a linear transformation to vectors of length two, which are used for binary classification. The isolate was classified as resistant or susceptible based on the largest output value. Additionally, models with the combinations of 1 or 2 encoder heads and layers, with

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

word embedding of size 64 or 128, and a position-wise feed-forward layer of 128 or 256, with patient data and no patient data, were trained (Fig. S13 through S15). The proposed model achieved higher accuracy, lower ME, and lower VME in 23, 14, and 10 out of 45 pathogen-antibiotic combinations, respectively. The combination of one head, one layer, with word embedding of size 128 and a position-wise feed-forward layer of size 256 was the only one to have more pathogen-antibiotic combinations with lower VME (13).

The model was trained as follows. At each epoch, 248,000 bacterial isolates were randomly sampled from the train data set and expanded as described in "Data expansion," above. The model was then trained on 512,000 randomly selected data points from the data expansion step, divided into mini-batches of size 512. The loss was based on the cross-entropy between the known (y_k) and predicted ($\hat{y}_k$) labels. The Adam optimizer was used to minimize the loss over 700 epochs with a fixed learning rate of 1×10^{-6}. The model was implemented and trained using PyTorch version 1.7.1.

Uncertainty control

An algorithm based on conditional inductive conformal prediction was used to quantify the uncertainty of the predictions (34) with respect to the antibiotic and label (i.e., "susceptible" and "resistant") (34). Conformal prediction uses empirical data to estimate uncertainty and has previously been shown to suit complex diagnostic data for which valid distributional assumptions may be hard to make (43). For a data point (x_k, y_k) belonging to pathogen q, the algorithm estimates a prediction set $\Gamma_k^{\varepsilon, q}$, containing the predictions that are deemed sufficiently confident given a predefined confidence level $1 - \varepsilon$. The uncertainty for a prediction was based on its conformity measure, defined as the softmax transformation of the outputs of a neural network. The conformity measure and, thus, the uncertainty were derived individually for each antibiotic and each label (i.e., resistance or susceptible).

We estimated the empirical distributions for each conformity measure from the calibration data set separately for each pathogen, which, for an antibiotic a and pathogen q, were assumed to contain $l_{q,a} = l_{q,a,S} + l_{q,a,R}$ data points, where $l_{q,a,S}$ and $l_{q,a,R}$ are the number of data points for susceptible and resistant bacterial isolates of pathogen q, respectively. For a data point (x_k, y_k) derived from pathogen q, let $\alpha_k^{a,S}$ and $\alpha_k^{a,R}$ denote the softmax score for the prediction of susceptibility and resistance to antibiotic a, respectively. The prediction sets were decided based on the empirical p-values $p_k^{q,a,S}$ and $p_k^{q,a,R}$, which were calculated according to

$$p_k^{q,a,S} = \frac{|i = 1, \ldots, l_{q,a,S} : \tilde{\alpha}_i^{q,a,S} \leq \alpha_k^{a,S}| + 1}{l_{q,a,S} + 1},$$

$$p_k^{q,a,R} = \frac{|i = 1, \ldots, l_{q,a,R} : \tilde{\alpha}_i^{q,a,R} \leq \alpha_k^{a,R}| + 1}{l_{q,a,R} + 1},$$

where $\tilde{\alpha}_i^{q,a,S}$ and $\tilde{\alpha}_i^{q,a,R}$ denotes the softmax scores calculated using the data points in the calibration data set for isolates from pathogen q. At a confidence $1 - \varepsilon$, the prediction set was then formed by

$$\Gamma_k^{\varepsilon,a} = \left\{S \text{ if } p_k^{q,a,S} > \varepsilon\right\} \bigcup \left\{R \text{ if } p_k^{q,a,R} > \varepsilon\right\}.$$

Performance

The model's predictive performance was computed based on 257,784, 242,121, and 114,410 train; 285,897, 223,661, and 97,542 calibration; and 280,428, 223,665, and 100,927 test individual AST results for *E. coli*, *K. pneumoniae*, and *P. aeruginosa* isolates, respectively. To evaluate the overall model performance, the F1 score was calculated.

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

In addition, the ME rate, defined as the proportion of true susceptible bacterial isolates erroneously predicted as resistant, and the VME rate, defined as the proportion of true resistant isolates erroneously predicted as susceptible, were also calculated. To measure the performance of the uncertainty control, true predictions were defined as prediction regions containing the true label for each antibiotic, and false predictions were those containing either no labels or only the wrong one.

For comparison, the following naive classifier was included (Fig. S7). For each of the bacterial species, the frequencies between all pairwise antibiotics and their susceptibility profile were calculated in the training data set. Then, the ratios $r_{a_R,b} = \frac{|a_R b_R|}{|a_R b_R| + |a_R b_S|}$ and $r_{a_S,b} = \frac{|a_S b_R|}{|a_S b_R| + |a_S b_S|}$, between antibiotic a, with resistance (R) and susceptible (S) profiles, and the antibiotic b, with resistance profile, were calculated for all pairwise combinations of antibiotics. Finally, for a new prediction of antibiotic b in y_k, resistance was deemed if the average $A = \sum_{a_j \in x_k} \frac{r_{a_j b}}{n_a}$ was larger than 0.5, where a_j represent the antibiotics in x_k with resistance $j \in \{R, S\}$, and n_a is the number of antibiotics in x_k; the isolate was deemed susceptible otherwise. To investigate the effect of patient data, we compared our results with a model based solely on AST results (Fig. S6). The VMEs for the three bacterial species were found to be significantly smaller ($P < 0.01$) when the original model was compared to the model without patient data and to the naive classifier, as determined by Wilcoxon signed-rank tests.

ACKNOWLEDGMENTS

Data from The European Surveillance System – TESSy between 2007 and 2020 were provided by Andorra, Albania, Armenia, Austria, Azerbaijan, Bosnia and Herzegovina, Belgium, Bulgaria, Belarus, Switzerland, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Georgia, Croatia, Hungary, Ireland, Israel, Iceland, Italy, Kyrgyzstan, Kazakhstan, Liechtenstein, Lithuania, Luxembourg, Latvia, Monaco, Republic of Moldova, Montenegro, Republic of North Macedonia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Russian Federation, Sweden, Slovenia, Slovakia, San Marino, Tajikistan, Turkmenistan, Turkey, Ukraine, United Kingdom, Uzbekistan, and Kosovo and released by The European Centre for Disease Prevention and Control (ECDC). The views and opinions of the authors expressed here do not necessarily state or reflect those of ECDC. The accuracy of the authors' statistical analysis and the findings they report are not the responsibility of ECDC. ECDC is not responsible for conclusions or opinions drawn from the data provided. ECDC is not responsible for the correctness of the data and data management, data merging, and data collation after the provision of the data. ECDC shall not be held liable for improper or incorrect use of the data.

We acknowledge support from Swedish Research Council (VR) grant 2018-02835 (E.K.), Swedish Research Council (VR) grant 2019-03482 (E.K.), Chalmers AI Research Centre (CHAIR; E.K.), and National Health and Medical Research Council of Australia under the framework of JPI AMR, grant 2031902, SEARCHER.

AUTHOR AFFILIATIONS

[1]Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden
[2]Center for Antibiotic Resistance Research (CARe), University of Gothenburg, Gothenburg, Sweden
[3]Centre for Microbiome Research, School of Biomedical Sciences, Queensland University of Technology, Translational Research Institute, Woolloongabba, Australia
[4]Department of Systems and Data Analysis, Fraunhofer-Chalmers Center, Gothenburg, Sweden
[5]Department of Clinical Microbiology, Sahlgrenska University Hospital, Gothenburg, Sweden

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

[6]Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden

AUTHOR ORCIDs

Juan S. Inda-Díaz http://orcid.org/0000-0002-3735-8300
Anna Johnning http://orcid.org/0000-0003-2185-2432
Erik Kristiansson http://orcid.org/0000-0002-8609-2414

FUNDING

Funder	Grant(s)	Author(s)
Vetenskapsrådet	2018-02835	Erik Kristiansson
Vetenskapsrådet	2019-03482	Erik Kristiansson
Chalmers Tekniska Högskola	Chalmers Artificial intelligence Research Centre (CHAIR)	Erik Kristiansson
National Health and Medical Research Council	2031902	Juan S. Inda-Díaz
Vetenskapsrådet	2024-06177	Erik Kristiansson

AUTHOR CONTRIBUTIONS

Juan S. Inda-Díaz, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review and editing | Anna Johnning, Formal analysis, Supervision, Writing – original draft, Writing – review and editing | Magnus Hessel, Formal analysis, Writing – original draft, Writing – review and editing | Anders Sjöberg, Formal analysis, Writing – original draft, Writing – review and editing | Anna Lokrantz, Formal analysis, Writing – original draft, Writing – review and editing | Lisa Helldal, Formal analysis, Writing – original draft, Writing – review and editing | Mats Jirstrand, Formal analysis, Writing – original draft, Writing – review and editing | Lennart Svensson, Conceptualization, Formal analysis, Writing – original draft, Writing – review and editing | Erik Kristiansson, Conceptualization, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The data used in this study come from the European Surveillance System (TESSy). We acknowledge TESSy for data availability and refer the reader to the European Surveillance System for data access. The model is available in the GitHub repository at https://github.com/indajuan/Confidence-based-Prediction-of-Antibiotic-Resistance.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental Material (mBio03431-25-s0001.docx). Supplemental figures and tables.

REFERENCES

1. World Health Organization. 2015. Global action plan on antimicrobial resistance
2. Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, Han C, Bisignano C, Rao P, Wool E, et al. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet 399:629–655. https://doi.org/10.1016/S0140-6736(21)02724-0
3. O'Neill J. 2016. Tackling drug-resistant infections globally: final report and recommendations. Wellcome Trust and Government of the United Kingdom
4. Mercer DK, Torres MDT, Duay SS, Lovie E, Simpson L, von Köckritz-Blickwede M, de la Fuente-Nunez C, O'Neil DA, Angeles-Boza AM. 2020. Antimicrobial susceptibility testing of antimicrobial peptides to better

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

predict efficacy. Front Cell Infect Microbiol 10:326. https://doi.org/10.3389/fcimb.2020.00326
5. Jorgensen JH, Ferraro MJ. 1998. Antimicrobial susceptibility testing: general principles and contemporary practices. Clin Infect Dis 26:973–980. https://doi.org/10.1086/513938
6. Friedman ND, Temkin E, Carmeli Y. 2016. The negative impact of antibiotic resistance. Clin Microbiol Infect 22:416–422. https://doi.org/10.1016/j.cmi.2015.12.002
7. Bassetti M, Rello J, Blasi F, Goossens H, Sotgiu G, Tavoschi L, Zasowski EJ, Arber MR, McCool R, Patterson JV, Longshaw CM, Lopes S, Manissero D, Nguyen ST, Tone K, Aliberti S. 2020. Systematic review of the impact of appropriate versus inappropriate initial antibiotic therapy on outcomes of patients with severe bacterial infections. Int J Antimicrob Agents 56:106–184. https://doi.org/10.1016/j.ijantimicag.2020.106184
8. Battle SE, Bookstaver PB, Justo JA, Kohn J, Albrecht H, Al-Hasan MN. 2017. Association between inappropriate empirical antimicrobial therapy and hospital length of stay in Gram-negative bloodstream infections: stratification by prognosis. J Antimicrob Chemother 72:299–304. https://doi.org/10.1093/jac/dkw402
9. Kumar A, Ellis P, Arabi Y, Roberts D, Light B, Parrillo JE, Dodek P, Wood G, Kumar A, Simon D, Peters C, Ahsan M, Chateau D, Cooperative Antimicrobial Therapy of Septic Shock Database Research Group. 2009. Initiation of inappropriate antimicrobial therapy results in a fivefold reduction of survival in human septic shock. Chest 136:1237–1248. https://doi.org/10.1378/chest.09-0087
10. van den Bosch CMA, Hulscher MEJL, Akkermans RP, Wille J, Geerlings SE, Prins JM. 2017. Appropriate antibiotic use reduces length of hospital stay. J Antimicrob Chemother 72:923–932. https://doi.org/10.1093/jac/dkw469
11. Dolejska M, Papagiannitsis CC. 2018. Plasmid-mediated resistance is going wild. Plasmid 99:99–111. https://doi.org/10.1016/j.plasmid.2018.09.010
12. Johnning A, Karami N, Tång Hallbäck E, Müller V, Nyberg L, Buongermino Pereira M, Stewart C, Ambjörnsson T, Westerlund F, Adlerberth I, Kristiansson E. 2018. The resistomes of six carbapenem-resistant pathogens - a critical genotype-phenotype analysis. Microb Genom 4:e000233. https://doi.org/10.1099/mgen.0.000233
13. Dias SP, Brouwer MC, van de Beek D. 2022. Sex and gender differences in bacterial infections. Infect Immun 90:e0028322. https://doi.org/10.1128/iai.00283-22
14. Murray KA, Preston N, Allen T, Zambrana-Torrelio C, Hosseini PR, Daszak P. 2015. Global biogeography of human infectious diseases. Proc Natl Acad Sci USA 112:12746–12751. https://doi.org/10.1073/pnas.1507442112
15. Yelin I, Snitser O, Novich G, Katz R, Tal O, Parizade M, Chodick G, Koren G, Shalev V, Kishony R. 2019. Personal clinical history predicts antibiotic resistance of urinary tract infections. Nat Med 25:1143–1152. https://doi.org/10.1038/s41591-019-0503-6
16. Kanjilal S, Oberst M, Boominathan S, Zhou H, Hooper DC, Sontag D. 2020. A decision algorithm to promote outpatient antimicrobial stewardship for uncomplicated urinary tract infection. Sci Transl Med 12:eaay5067. https://doi.org/10.1126/scitranslmed.aay5067
17. Stracy M, Snitser O, Yelin I, Amer Y, Parizade M, Katz R, Rimler G, Wolf T, Herzel E, Koren G, Kuint J, Foxman B, Chodick G, Shalev V, Kishony R. 2022. Minimizing treatment-induced emergence of antibiotic resistance in bacterial infections. Science 375:889–894. https://doi.org/10.1126/science.abg9868
18. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. 2021. The role of artificial intelligence in healthcare: a structured literature review. BMC Med Inform Decis Mak 21:1–23. https://doi.org/10.1186/s12911-021-01488-9
19. Yu K-H, Beam AL, Kohane IS. 2018. Artificial intelligence in healthcare. Nat Biomed Eng 2:719–731. https://doi.org/10.1038/s41551-018-0305-z
20. Benjamens S, Dhunnoo P, Meskó B. 2020. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 3:118. https://doi.org/10.1038/s41746-020-00324-0
21. Rajpurkar P, Chen E, Banerjee O, Topol EJ. 2022. AI in health and medicine. Nat Med 28:31–38. https://doi.org/10.1038/s41591-021-01614-0
22. Ali T, Ahmed S, Aslam M. 2023. Artificial intelligence for antimicrobial resistance prediction: challenges and opportunities towards practical implementation. Antibiotics (Basel) 12:523. https://doi.org/10.3390/antibiotics12030523
23. Sakagianni A, Koufopoulou C, Feretzakis G, Kalles D, Verykios VS, Myrianthefs P, Fildisis G. 2023. Using machine learning to predict antimicrobial resistance—a literature review. Antibiotics (Basel) 12:452. https://doi.org/10.3390/antibiotics12030452
24. The Medical Futurist. 2022. FDA-approved A.I.-based algorithms. Available from: https://medicalfuturist.com/fda-approved-ai-based-algorithms/
25. Begoli E, Bhattacharya T, Kusnezov D. 2019. The need for uncertainty quantification in machine-assisted medical decision making. Nat Mach Intell 1:20–23. https://doi.org/10.1038/s42256-018-0004-1
26. Kompa B, Snoek J, Beam AL. 2021. Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digit Med 4:4. https://doi.org/10.1038/s41746-020-00367-3
27. Feretzakis G, Loupelis E, Sakagianni A, Kalles D, Martsoukou M, Lada M, Skarmoutsou N, Christopoulos C, Valakis K, Velentza A, Petropoulou S, Michelidou S, Alexiou K. 2020. Using machine learning techniques to aid empirical antibiotic therapy decisions in the intensive care unit of a general hospital in Greece. Antibiotics (Basel) 9:50. https://doi.org/10.3390/antibiotics9020050
28. Feretzakis G, Sakagianni A, Loupelis E, Kalles D, Skarmoutsou N, Martsoukou M, Christopoulos C, Lada M, Petropoulou S, Velentza A, Michelidou S, Chatzikyriakou R, Dimitrellos E. 2021. Machine learning for antibiotic resistance prediction: a prototype using off-the-shelf techniques and entry-level data to guide empiric antimicrobial therapy. Healthc Inform Res 27:214–221. https://doi.org/10.4258/hir.2021.27.3.214
29. Devlin J, Chang M-W, Lee K, Toutanova K. 2018. Bert: pre-training of deep bidirectional transformers for language understanding. arXiv. https://doi.org/10.48550/arXiv.1810.04805
30. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A. 2020. Language models are few-shot learners. Adv Neural Inf Process Syst 33:1877–1901.
31. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. 2017. Attention is all you need. 31st Conference on Neural Information Processing Systems (NIPS 2017); Long Beach, CA, USA. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
32. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2
33. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 40:1023–1025. https://doi.org/10.1038/s41587-021-01156-3
34. Vovk V. 2012. Conditional validity of inductive conformal predictors. Asian Conference on Machine Learning. Proceedings of Machine Learning Research.p 475–490
35. Vazquez J, Facelli JC. 2022. Conformal prediction in clinical medical sciences. J Healthc Inform Res 6:241–252. https://doi.org/10.1007/s41666-021-00113-8
36. Reller LB, Weinstein M, Jorgensen JH, Ferraro MJ. 2009. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. Clin Infect Dis 49:1749–1755. https://doi.org/10.1086/647952
37. European Centre for Disease Prevention and Control, European Medicines Agency. 2009. The bacterial challenge: time to react: a call to narrow the gap between multidrug-resistant bacteria in the EU and the development of new antibacterial agents. Publications Office.
38. Ahmad M, Khan AU. 2019. Global economic impact of antibiotic resistance: a review. J Glob Antimicrob Resist 19:313–316. https://doi.org/10.1016/j.jgar.2019.05.024
39. Vovk V, Gammerman A, Shafer G. 2005. Algorithmic learning in a random world. Springer-Verlag, New York, USA.
40. Papadopoulos H. 2008. Inductive conformal prediction: theory and application to neural networkstools in artificial intelligence. IntechOpen, Rijeka.
41. Bush K, Bradford PA. 2020. Epidemiology of β-lactamase-producing pathogens. Clin Microbiol Rev 33:e00047-19. https://doi.org/10.1128/CMR.00047-19
42. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, et al. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother 75:3491–3500. https://doi.org/10.1093/jac/dkaa345

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.

43. Olsson H, Kartasalo K, Mulliqi N, Capuccini M, Ruusuvuori P, Samaratunga H, Delahunt B, Lindskog C, Janssen EAM, Blilie A, Egevad L, Spjuth O, Eklund M, ISUP Prostate Imagebase Expert Panel. 2022. Estimating diagnostic uncertainty in artificial intelligence assisted pathology using conformal prediction. Nat Commun 13:7761. https://doi.org/10.1038/s41467-022-34945-8

Downloaded from https://journals.asm.org/journal/mbio on 24 February 2026 by 129.16.44.179.