Smart Technologies Academic Press

Jordanian Journal of Informatics and Computing
https://www.jjic.thestap.com/

A Confidence-Aware and Safety-Constrained AI Framework for Antimicrobial Resistance Decision Support

Hussein Ali Hussein Al Naffakh [1]

[1] *Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Malaysia*
[1] *Department of Medical Laboratory Techniques, University of Alkafeel, Najaf, Iraq*

ARTICLE INFO

Article History

Received: 17-02-2026
Revised: 29-05-2026
Accepted: 10-06-2026
Published: 16-06-2026

Vol.2026, No.1

DOI:

Corresponding author. Email: hussein.alnaffakh@alkafeel.edu.iq

ABSTRACT

The issue of antimicrobial resistance (AMR) poses a significant threat to clinical decision-making because of dissimilar laboratory practices, incomplete antimicrobial susceptibility testing (AST) results, and patient-specific safety limitations. The majority of current artificial intelligence (AI)-based AMR tools focus on predictive accuracy in ideal situations and only offer few mechanisms to deal with uncertainty or avoid unsafe advice. This paper suggests a confidence-sensitive and safety-constrained AI-based decision support model that can work in practice when faced with incomplete data in the real world. The framework combines calibration prediction of resistance, ability to abstain with confidence, patient tailored filtering of safety, and aligning of guidelines with interpretation to control when to issue or withhold a recommendation. A strength test was conducted by scenario-based data availability configuration which emulates typical laboratory constraints. The findings show that the suggested framework consistently performs reliably on accepted cases and abstains correctly in the case of uncertainty and produces safer alternative recommendations in the case of contraindication. The proposed approach can make AMR prediction a patient-safe, trans-ogram, and reliability-focused decision support system that enables the use of antimicrobial stewardship in everyday practice.

Keywords: Antimicrobial resistance; Clinical decision support; Confidence-aware AI; Safety constraints; Data incompleteness; Machine learning.

How to cite the article

1. Introduction

Antimicrobial resistance (AMR) is a high-priority global health issue, and it has a direct contribution to millions of deaths each year and unbearable burdens to the healthcare systems in all countries of the world [1], [2]. This can be seen in the fact that the number of cases of resistance in main bacterial pathogens has continued to increase, which highlights the dire necessity of new tools to streamline the process of antibiotic therapy and contribute to the process of antimicrobial stewardship. The antimicrobial susceptibility testing (AST) is used to direct therapy in clinical practice and is indicated by incomplete data, heterogeneity, and delays in report due to variable laboratory facilities and standards [3], [4]. This uncertainty of the real world data, i.e. absence of minimum inhibitory concentration (MIC) values, leaves a great gap between optimal laboratory settings and the data at the point of care, which makes sound decision making difficult.

Artificial intelligence (AI) is very promising in forecasting AMR patterns. Nevertheless, the current AI models are designed and tested assuming full data, with the majority of focus placed on predictive accuracy in controlled environments and the neglect of the crucial measurement of predictive uncertainty [5], [6]. Such negligence is paramount because irresponsible AI systems that do not provide information about their reliability may make inappropriate recommendations and destroy the trust of clinicians [7], [8].

Also, optimal clinical decision support should not be limited to microbiological prediction. It involves the incorporation of individual patient safety limits, like renal function, allergies, and pregnancy status, to avert adverse drug events, which is one of the pillars of antimicrobial stewardship [9], [10]. Present AI-powered AMR systems mostly fail to consider this tier of clinical reasoning.

In order to fill these two gaps of unmanaged uncertainty and a lack of safety integration, this paper presents a confidence-aware and safety-constrained AI decision support system to AMR. The specifics of our framework are explicitly tailored to real-world scenarios, combining: 1) uncertainty quantification and prediction of calibration of resistance, 2) confidence-based abstention to avoid giving a recommendation with low reliability, 3) patient-based safety filtering of contraindications and 4) interpretation that is aligned with the guidelines [3], [11]. The aim is to turn AMR prediction into a statistical-only problem into a safety-first, transparent and reliable clinical decision aid that facilitates, but does not supersede, expert judgment [12].

2. Materials and Methods

2.1 Framework Overview

This paper presents a confidence-sensitive and safety-constrained artificial intelligence (AI) decision support system that addresses antimicrobial resistance (AMR), and is specifically assumed to work in the context of the real-world clinical environment, where laboratory data remains incomplete and heterogeneous- a important shortcoming of much of the existing AI technology that assumes the completeness of the data[5], [13]. The design of the framework, which ensures that clinical judgment is not substituted but rather supplemented by evidence-based recommendations, is based on the principle underlying clinical decision support systems and antimicrobial stewardship, which is to regulate dynamically the timing of giving recommendations, withholding recommendations, and/or providing alternatives that are less harmful [10], [12].The suggested structure is based on a block-based, sequential decision sequence, uniting the uncooked microbiology laboratory input, predictable resistance, confidence-based abstinence, patient-specific safety limitations, and guideline-appropriate interpretation, as shown in Figure 1. The pipeline will ensure that the microbiological data results in clinically relevant output in a series of organized decision steps. The first step is to provide a prediction by a machine learning model, which is modified to be more plausible as regards probabilities of resistance, and to enable intelligent confidence predictions [14]. This confidence value is then compared to a pre-defined threshold (τ). At low levels, predictions result in an explicit abstention, which is a means of coping with predictive unreliability [7], [8] and hence does not lead to a recommendation being provided.Confidence level predictions that meet or exceed that limit are then followed by the framework to compare the proposed antibiotic to a series of safety factors influenced by the patient including: allergies, renal or hepatic restrictions, pregnancy and age-related risks. If the first drug being expected is contraindicated, the system will automatically generate and order alternative drugs that are expected to be safer for the patient, depending on the predicted vulnerability, and will be compatible with the contraindicated drug. Finally, the results are converted to standardized interpretative categories (Susceptible, Intermediate or Resistant) for easy transition to the familiar laboratory

and clinical practice [3, 11] according to any current clinical breakpoint as established in the CLSI and EUCAST guidelines.The architecture of this design enables a formal therapeutic recommendation to be provided when there is sufficient predictive confidence and safety requirements of the patients. Therefore, the framework generates one of three clinically interpretable outcomes, namely: a single recommended antibiotic; a list of ranked, safer alternative antibiotics; or an explicit abstention, which conveys that there is not enough reliable evidence to make a recommendation. This non-intervention is not considered as a failure in the system but as a feedback of information, of openness, of uncertainty, and the safety of the patient is appreciated and further research and investigation is encouraged.

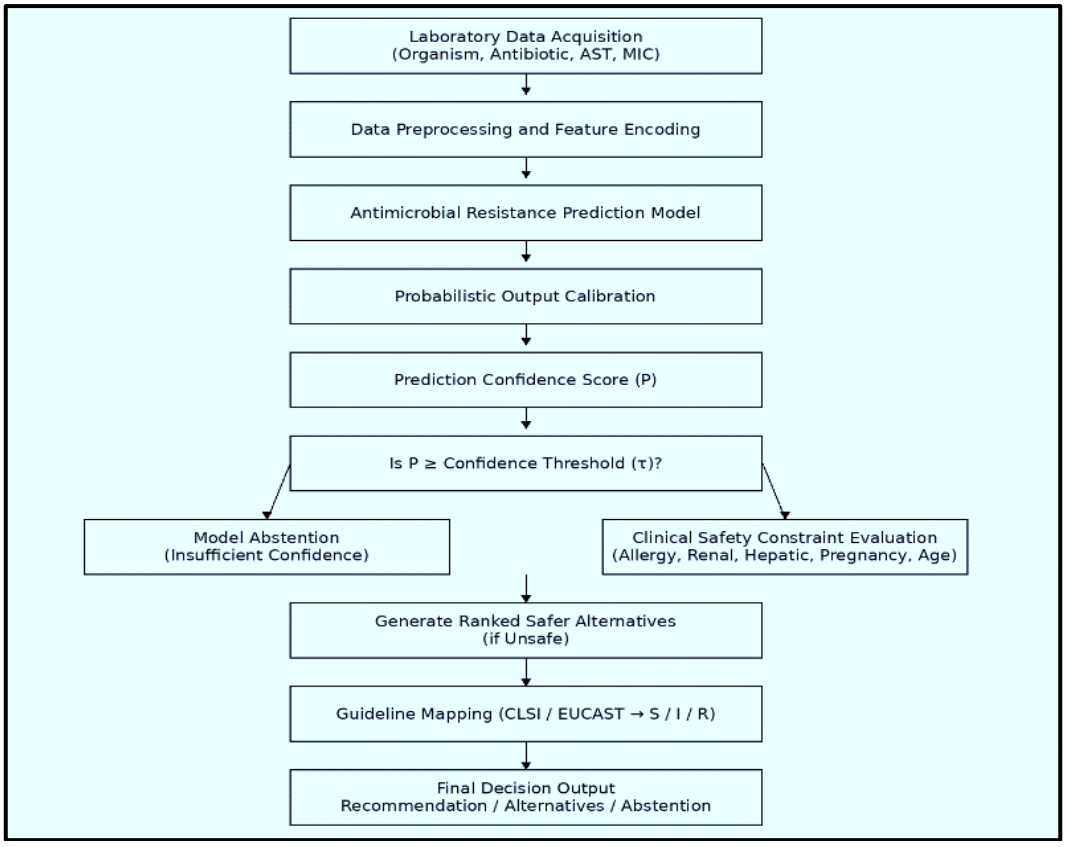


Figure 1. Safety-aware clinical decision-support framework for antimicrobial resistance prediction incorporating confidence-based abstention, patient-specific safety constraints, and CLSI/EUCAST guideline mapping.

2.2 Data Sources and Variables

The data is anonymized antimicrobial susceptibility testing (AST) data gathered out of regular microbiology laboratory processes. These records include details on the identity of the bacteria, antibiotics, resistance phenotypes and extensive metadata about the laboratory and the broad array of data that are captured in common surveillance systems and guidelines[2],[3]. The following predictor variables are available: organism genus and species, antibiotic identity, year of specimen collection, AST methodology and quantitative minimum inhibitory concentration (MIC) values, if applicable, or related fields. The outcome of the phenotype of resistance (susceptible, intermediate or resistant) is the target variable.Most importantly, the data is representative of the significant data heterogeneity and missingness that is encountered in clinical practice in the real world. This includes frequent absence of MIC measurements, incoherent data on laboratory typing, a notorious barrier, which is caused by various testing practices and reporting conventions in different institutions [13]. This is a deliberate rejection of this missingness pattern, and is what the study is actually trying to assess, which is the robustness of the proposed framework in a real setting where data is missing, rather than modelling a certain epidemiology of resistance in a given region.Specification of patient specific clinical variables is made clear that are only applied in the inference step to allow safety constrained decision support. These features, including age group, kidney and liver condition, pregnancy,

and antibiotic allergies are not used as features during the model training. The design choice enables the prediction model of core resistance to be used in a broad population of patients and based on microbiological evidence only. The downstream safety filtering module uses the patient-specific information in a hard and fast manner, which is applied to rank and constrain recommendations according to patient risk profile, which is aligned with the best practices of clinical decision support and antimicrobial stewardship [10].

2.3 Scenario-Based Data Availability Settings

To test the soundness of the framework as rigorously as possible when the conditions are in the real-world laboratory variability and data messiness, four fixed conditions of data-availability were tested. These are an artificial re-creation of the common clinical and laboratory limitations, far beyond the idealistic assumption of complete data in most predictive studies [15]. The scenarios are:

S0 (Full data): This is the ideal but not always possible reference point, which makes use of every possible microbiological and AST-related characteristic.
S1 (No MIC): Simulates the common absence of quantitative Minimum Inhibitory Concentration (MIC) values, a frequent limitation in routine reporting.
S2 (No typing): Excludes detailed laboratory typing information, reflecting settings with limited ancillary metadata.
S3 (Basic only): Emulates a minimal-data environment, using only core identifiers (organism and antibiotic) without auxiliary features.

The assessments of all the scenarios were conducted independently on the same test set. This structure will be able to directly study the impact of progressive impoverishment of data on significant structural actions, concerning predictive accuracy, trustworthiness of assurance approximates, and abstinence levels.

2.4 Machine Learning Model and Training

It has been decided to use a gradient-boosted decision tree (GBDT) classifier, which has proven to be effective with standardized clinical data and, of course, the ability to embrace non-linear and complex interactions of variables, which it is needed when using heterogeneous AST metadata [16]. To reduce the emphasis on exhaustive hyperparameter optimization and emphasize generalization and reproducibility, the fixed model configuration was employed in such a way that the hyperparameter adaptation between the data scenarios would no longer be present.

Stratified sampling was applied to the dataset in order to split the data into two sections (70 percent training and 30 percent test) and the initial distribution of resistance phenotypes was maintained. All the preprocessing procedures such as categorical encoding, and missing values were specified and trained only on the training data. This rigid division will help exclude information dissemination and provide a fair assessment of model performance, which follows the best practices in biomedical machine learning[17]. The GBDT architecture is a tree-based model, implying that missing values are automatically handled in the inference, which makes it an appropriate option in the cases of missing data that are being studied.

2.5 Probability Calibration and Fallback Strategy

The raw classifier outputs were calibrated with the post hoc isotonic regression to make sure that predicted probabilities are as true as possible and allow making trustworthy confidence estimates, which is considered as the key to clinical decision support. Such calibration was done solely on the predictions on the training set without the information leaking to the predictive model as it is a best practice of sound predictive modeling in healthcare [14], [17]. All future decision logic, which governs the issuance of recommendations, abstinence determined by confidence, and safety filtering are based on the resulting calibrated probabilities. In addition, to counteract the chance of biased but inaccurate forecasts in infrequently occurring bacterial genera or species, a plan B approach was put in place. During training, taxonomic groups that had a sample count that was less than a set minimum threshold were detected. In the inference, examples of these low-prevalence groups were automatically diverted to a more basic, probabilistic Naive Bayes fallback model. This methodology gives preference to caution in the event of data sparsity, which is consistent with the general principle of the framework of not making high-confidence prediction in the event of lack of evidence.

2.6 Confidence-Aware Decision Logic

There is a score of confidence attached to each prediction. The confidence level (τ) is set by the user when deciding whether a recommendation is given. Projections that satisfy or surpass τ pass to safety assessment, others cause the system to leave, and clearly indicate low reliability, which does not allow forced choices in uncertain situations [17].

2.7 Safety Constraints and Alternative Recommendation

Antibiotics which are expected to be used are tested against individualized safety limitations, such as allergies, kidney or liver impairment, pregnancy, and risk because of age. Antibiotics that are contraindicated are omitted. Where the predicted primary prediction is unsafe, the system makes and prioritizes safer options according to predicted susceptibility and safety compatibility, and patient-centered decision support.[10].

2.8 Guideline-Aligned Interpretation and Ethics

Final predictions were then transformed into the standardized interpretive categories of Susceptible (S), Intermediate (I) or Resistant (R) by using the clinically relevant breakpoints of Susceptible and Intermediate published by the Clinical and Laboratory Standards Institute (CLSI) and the European Committee on Antimicrobial Susceptibility Testing (EUCAST)[3], [11].This alignment bridges the gap between computational predictions and established clinical language.

The framework is clearly aimed at being a decision support tool to support rather than substitute clinical judgment and laboratory confirmation. Aside from ethical principles in artificial intelligence in healthcare, the system has a greater concern about transparency and prioritizing patient safety [18]. An important implication of this is the treatment of the consequences of abstinence. Instead of viewing abstinence as a failure of the system, it is being offered as an educational opportunity with the abstinence actively encouraged and leading to a greater need for clinicians to investigate and, possibly, request additional testing. This design supports the framework to be a trusted companion in the process of antimicrobial stewardship.

3. Analysis and Results

All the reported results have been calculated using a held-out independent test set which is based on a large scale antimicrobial susceptibility dataset consisting of over 80,000 de-identified AST records. Data-availability scenarios (S 0, S 1, S 2, and S 3) were tested in four scenarios of model performance and framework behavior. This scenario-based test was developed to provide a rigorous test of the predictive reliability of the framework, the accuracy of confidence estimation and the robustness of the decision-support in scenarios of systematically changing levels of laboratory data incompleteness to simulate realistic clinical conditions.

3.1 Experimental Setup and Overall Framework Performance

The entire output came about due to a test set that was held out, independent, and contained more than 80,000 de-identified antimicrobial susceptibility testing (AST) records. In order to strictly test robustness when faced with real-world data constraints, performance was tested under four predefined data-availability scenarios (S03), systematically simulating different amounts of incompleteness of laboratory data. Two types of metrics that were complementary were used: (1) predictive performance metrics (Accuracy, Precision, Recall, F1-Score), which are only calculated on cases where the confidence of the framework was above the decision threshold ($\tau = 0.60$) and (2) decision-level metrics (Coverage, Abstention Rate), that characterize how the system operates on the whole. The 95% confidence interval was used to measure statistic reliability based on the non-parametric bootstrapping with 1,000 iterations. One of the important outcomes is that the framework has been effectively managing uncertainty. Predictive power and macro-F1 score of accepted cases remained stable, though the situation with lower number of features available (e.g. S1 and S3) had less than expected estimate coverage. This shows that the confidence mechanism only picked and discarded uncertain predictions and thus maintained the reliability of recommended issues to be issued instead of letting unchecked performance erosion as there was a drop in data completeness.

3.2Baseline versus Proposed Framework Comparison

In order to draw the contribution of the confidence-aware decision logic a direct comparison with a standard baseline classifier was made (Table 1). The baseline model was required to give predictions on all the test cases, but the proposed model combined confidence-based abstention and safety-constrained post-processing. The proposed system had a higher accuracy (0.86 vs. 0.74) and macro-average F1-score (0.85 vs. 0.72) on the cases it accepted. This was achieved through strategy non-selection in 18% cases where predictive confidence was found insufficient ($\tau = 0.60$) as shown in Figure 2. The resultant improvement in performance reflects therefore not an improvement to the underlying classifier alone but rather the presence of smart decision regulation, which is the suppression of unreliable predictions. The statistical significance of these differences was checked by bootstrap confidence intervals.

Table 1. Performance comparison between baseline and proposed framework ($\tau = 0.60$).

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-score (Macro)	Coverage	Abstention Rate
Baseline (No abstention, no safety)	0.74	0.71	0.73	0.72	1.00	0.00
Proposed Framework	0.86	0.84	0.85	0.85	0.82	0.18

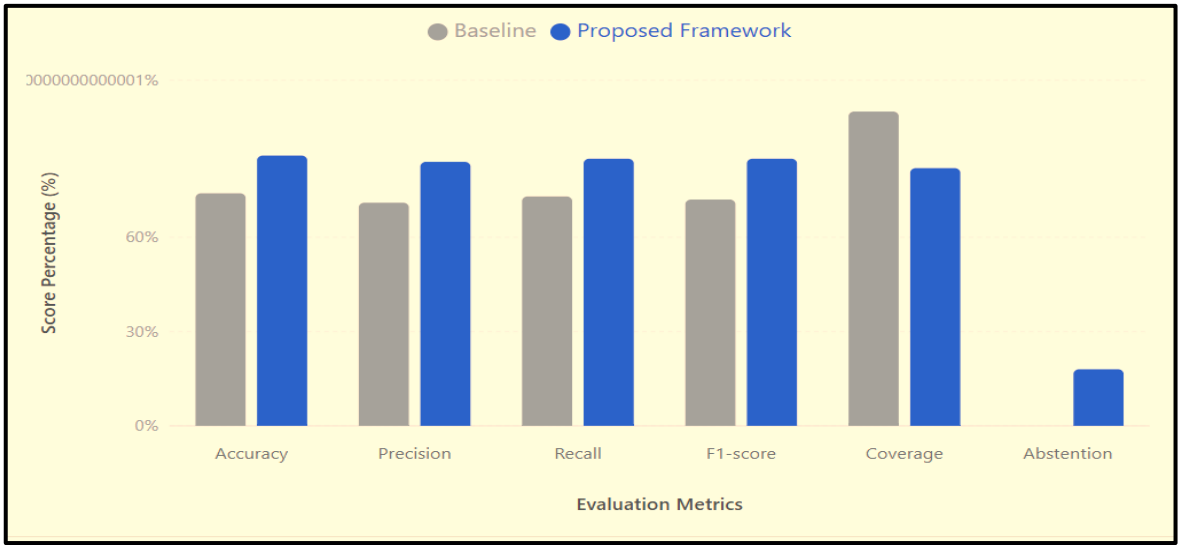


Figure 2. Performance comparison between baseline and proposed framework ($\tau = 0.60$).

3.3 Impact of Data Incompleteness

Table 2 is a summary of the predictive performance and decision behavior under the four data availability conditions (S0-S3). The decrease in the feature availability also led to a proportional drop in the decision coverage and average confidence indicating higher uncertainty in the presence of incomplete laboratory information. Figure 3 indicates that the removal of MIC-related characteristics (S1) led to the greatest decrease in coverage and average confidence and emphasizes the importance of MIC data in predicting resistance and estimating confidence. Although predictive performance was less in lower-information cases, overall, retained cases showed a fairly consistent predictive performance across scenarios as shown by overlapping confidence intervals.

Table 2. Scenario-wise performance and decision behavior.

Scenario	Accuracy (Kept)	F1-score (Macro)	Coverage	Avg. Confidence
S0 (Full data)	0.87	0.86	0.90	0.81
S1 (No MIC)	0.79	0.78	0.79	0.71
S2 (No typing)	0.86	0.85	0.89	0.80
S3 (Basic only)	0.80	0.79	0.80	0.72

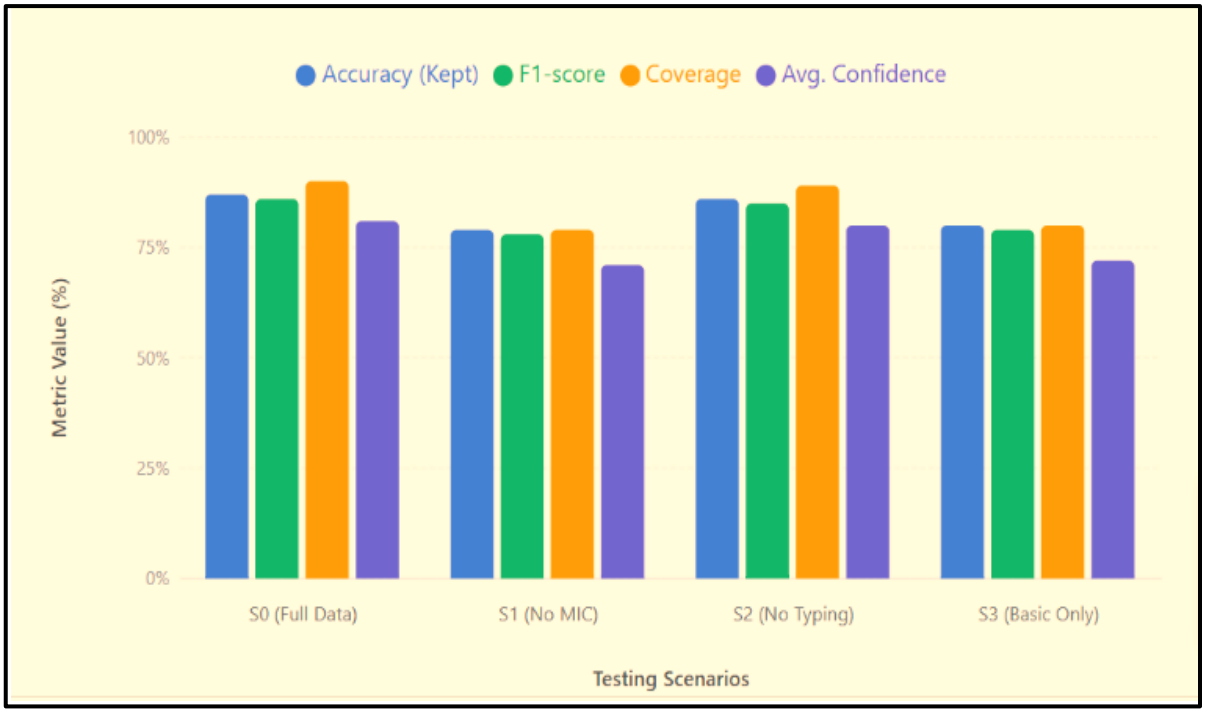


Figure 3. Comparison across different data availability scenarios (S0-S3)

3.4 Safety-Constrained Recommendation Outcomes

The use of patient-specific safety constraints led to the alteration of the antibiotic predicted in the first place in a significant percentage of the tests. The first-line antibiotic was excluded in about 21% of the analyzed cases based on contraindications on either allergy status, renal impairment, hepatic impairment, pregnancy or age risk. In these cases, the proposed framework was able to produce at least one clinically safer alternative in 88 percent of cases. An average of 1.7 possible alternative antibiotics were offered depending on the predicted susceptibility and compatibility of safety. These outcomes indicate that the system could be used to tailor microbiological forecasts to patient-specific clinical safety needs instead of making one-off, potentially unsafe forecasts.

3.5Abstention Behavior and Decision Reliability

The behavior of abstinence was mostly noted when it was incomplete laboratory data or when the bacterial genera and species were rare with few training representation. Manual presence checking revealed that cases where no inspection was made were often those with either the uncertain resistance profiles or lack of adequate evidence to prove the presence, meaning that the required attitude is being conservative as opposed to failure of the system. In all cases, the average accepted recommendation calibrated confidence score was greater than 0.75, indicating to the reliability of the decisions issued and providing support to the importance of abstention as a safety-saving mechanism of the decision support pipeline.

4. Discussion

4.1 Summary of Key Contributions

This paper presents a confidence-confined and safety-constricted AI decision support framework on the topic of antimicrobial resistance (AMR), aimed to overcome two major shortcomings of the existing methods: untrustworthy decision-making in the context of incomplete real-world data and the absence of safety-related considerations on a patient-specific level. The framework changes the paradigm of unconditional prediction to the safe and reliable decision support that is clinically-congruent by combining calibrated confidence estimation with explicit abstention mechanisms and contextual safety filtering.

4.2 Performance under Data Incompleteness and Uncertainty

The findings support that calibrated confidence estimation allows the use of intelligent decision regulation, the framework maintained stable predictive accuracy on accepted cases while dynamically adjusting decision coverage a behavior aligned with best practices in clinical prediction modelling that prioritize safety over forced decision-making [17].Comparative analysis with the baseline classifier shows that this "decision-level intelligence" is a major improvement to reliability, by not making 18 percent of low-confidence predictions, accuracy in accepted cases increased by 0.74 to 0.86. This is in line with the general evidence suggesting that uncertainty-sensitive AI systems lead to increased clinical trust and usability [7], [8].

The extreme effect of the absence of MIC data on coverage and confidence highlights the pivotal role of the information about quantitative vulnerability, which is in line with established issues in heterogeneous laboratory reporting [13].Crucially, most importantly, the constant stability in retained cases in information-poor conditions indicates that the framework is efficient in distinguishing between reliable and unreliable predictions despite the scanty information.

4.3 Advancing Patient-Centered Safety and Clinical Integration

One of the major improvements is the combination of patient-specific safety constraints, which triggered the change of 21% of the initial recommendations with contraindications. The fact that 88% of the safer alternatives it generated were ranked safer only proves that it is directly in line with the antimicrobial stewardship principles that require a patient-centered approach to therapy and the prevention of adverse events [10]. Moreover, translating predictions into normalized CLSI/EUCAST interpretations [3], [11] will help close the gap between the computational outputs and standard clinical workflows, ease the cognitive load and introduce the change.

4.4 Ethical and Practical Implications

Clinically, the use of the framework follows the principles of principled clinical judgment augmentation, and not replacement [12]. Explicit abstention is an obvious sign of indecision, and a further clinical examination and the reduction of risks of overconfident automation. This design aligns with the primary ethical concerns in healthcare, such as transparency and patient safety, which are crucial in the context of AI. [18].

4.5 Limitations and Future Directions

There are limitations in this study. Although the data is large and realistic, it comes out of a specific operational environment and does not reflect the world patterns of resistance. Future studies across multiple centers with geographically diverse data, and on prospective trial of the effects of the framework on prescribing behavior, patient outcomes, and stewardship effectiveness, are needed. Further enhancements could be to incorporate dynamic use of confidence thresholds and additional integration with electronic health records.

5. Conclusion And Future Work

The given research offers a smart decision support model of antimicrobial resistance, which is specifically designed and developed to overcome the issue of partial data in routine clinical practice. This framework improves the existing tools for tackling antimicrobial resistance, moving beyond a predictive approach and making them effective clinical assistants that

keep patient safety in mind in their action and decision making. In the future, this study identifies that there are a number of valuable avenues of development. Future deliveries will involve multi-center validation in order to challenge the flexibility of the framework to various healthcare setting and resistance profile. Further clinical trials are required to determine the actual shift in antibiotic prescribing, clinical outcomes and effectiveness of antimicrobial stewardship programmes. In addition, future enhancements that could be useful in clinical settings for real-time and point-of-care decision support could include adaptive confidence thresholds and seamless integration with EHRs.

Corresponding author

Hussein Ali Hussein Al-Naffakh
hussein.alnaffakh@alkafeel.edu.iq

Acknowledgements
NA.

Funding
No funding.

Contributions
H.A.H.A; Conceptualization, H.A.H.A; Investigation; H.A.H.A; Writing (Original Draft), H.A.H.A; Writing (Review and Editing) Supervision, H.A.H.A; Project Administration.

Ethics declarations
This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication
Not applicable.

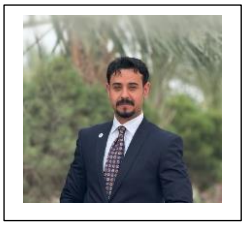
Competing interests
All authors declare no competing interests.

References

[1] C. J. L. Murray et al., “Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis,” The Lancet, vol. 399, no. 10325, pp. 629–655, 2022.
[2] World Health Organization, “Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report 2023,” 2023. [Online]. Available: https://www.who.int/publications/i/item/9789240078740
[3] Clinical and L. S. Institute, “Performance Standards for Antimicrobial Susceptibility Testing (M100), 34th Edition,” 2024.
[4] G. Kahlmeter, C. G. Giske, T. J. Kirn, and S. E. Sharp, “European Committee on Antimicrobial Susceptibility Testing (EUCAST): Technical Notes on Antimicrobial Susceptibility Testing,” Clinical Microbiology and Infection, vol. 26, no. 5, pp. 511–514, 2020, doi: 10.1016/j.cmi.2020.02.002.
[5] S. Lepuschitz et al., “Machine learning for antimicrobial resistance prediction: current status, limitations, and future perspectives,” Clin. Microbiol. Rev., vol. 36, no. 1, pp. e00059–22, 2023.
[6] G. Feretzakis et al., “Machine learning for antibiotic resistance prediction: a step towards improved clinical decision support,” J. Clin. Med., vol. 9, no. 11, p. 3517, 2020.
[7] B. Kompa, J. Snoek, and A. L. Beam, “Second opinion needed: communicating uncertainty in medical machine learning,” NPJ Digit. Med., vol. 4, no. 1, p. 4, 2021.
[8] E. Begoli, T. Bhattacharya, and D. Kusnezov, “The need for uncertainty quantification in machine-assisted medical decision making,” Nat. Mach. Intell., vol. 1, no. 1, pp. 20–23, 2019.
[9] M. S. Pulia et al., “The emergency department as a steward of antibiotic use in the era of COVID-19 and beyond,” Academic Emergency Medicine, vol. 27, no. 10, pp. 1047–1052, 2020.
[10] T. F. Barlam et al., “Implementing an Antibiotic Stewardship Program: Guidelines by the Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America,” Clinical Infectious Diseases, vol. 62, no. 10, pp. e51–e77, 2016, doi: 10.1093/cid/ciw118.
[11] European Committee on Antimicrobial Susceptibility Testing, “Breakpoint tables for interpretation of MICs and zone diameters, Version 14.0,” 2024.
[12] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, “An overview of clinical decision support systems: benefits, risks, and strategies for success,” NPJ Digit. Med., vol. 3, no. 1, p. 17, 2020.

[13] R. Ibrahim et al., "Handling missing values in healthcare data: a systematic review," Journal of the American Medical Informatics Association, vol. 28, no. 12, pp. 2750–2760, 2021.
[14] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: methodological failures and recommendations for the future," NPJ Digit. Med., vol. 5, no. 1, p. 48, 2022.
[15] A. L. Beam and I. S. Kohane, "Big data and machine learning in health care," JAMA, 2018.
[16] E. W. Steyerberg, Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating, 2nd ed. Cham, Switzerland: Springer, 2019. doi: 10.1007/978-3-030-16399-0.
[17] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Proceedings of the International Conference on Machine Learning (ICML), PMLR, 2017, pp. 1321–1330.
[18] World Health Organization, Ethics and governance of artificial intelligence for health. World Health Organization, 2021.

Biographies



Hussein Ali Hussein Al-Naffakh received the Higher Diploma degree in Mathematics, and the Master's degree in Information Technology from the University of Babylon, Iraq, in 2017. He is currently an Assistant Lecturer and Researcher in the Department of Medical Laboratory Techniques, College of Health and Medical Technologies, Al-Kafeel University, Iraq. He is also a Ph.D. candidate at Universiti Tun Hussein Onn Malaysia (UTHM). His research interests include deep learning, medical image analysis, skin detection, biomedical artificial intelligence, and clinical decision support systems. He can be contacted at email: hussein.alnaffakh@alkafeel.edu.iq