IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155298

AMR INSIGHT: MACHINE LEARNING-BASE MICROBE RESISTANCE PREDICTION USING ARFA

Dr. Swetha Singh[1], Subash M, Akash V[2]

Assistant Professor , Artificial Intelligence and Data Science, Dhanalakshmi Srinivasan College of Engineering and Technology, Chengalpattu, India[1]

Artificial Intelligence and Data Science, Dhanalakshmi Srinivasan College of Engineering and Technology Chengalpattu,India[2]

Abstract: Antimicrobial Resistance (AMR) has emerged as one of the most significant global healthcare challenges, leading to treatment failures, prolonged hospitalization, increased mortality rates, and rising healthcare costs. Conventional antimicrobial susceptibility testing methods require significant laboratory time and resources, making them unsuitable for rapid clinical decision-making. This paper proposes a Machine Learning-Based Antimicrobial Resistance Prediction System integrated with an Adaptive Resistance Fusion Algorithm (ARFA) to improve resistance prediction accuracy. The proposed framework utilizes clinical and microbiological data including patient demographics, microbial species information, infection characteristics, and antibiotic usage history. ARFA performs weighted feature fusion to generate a structured resistance risk representation before classification. Multiple supervised machine learning algorithms including Random Forest, Support Vector Machine (SVM), Logistic Regression, and Decision Tree are employed for resistance prediction. The system provides rapid resistance assessment, improves clinical decision support, and enhances antimicrobial stewardship practices. Experimental analysis demonstrates improved prediction performance and interpretability compared to traditional approaches. The proposed solution offers a scalable and intelligent framework for early antimicrobial resistance detection in modern healthcare environments.

Keywords: Antimicrobial Resistance, Machine Learning, ARFA, Random Forest, Support Vector Machine, Clinical Prediction, Healthcare Analytics, Antibiotic Resistance.

1 INTRODUCTION

Antimicrobial Resistance (AMR) has become one of the most critical threats to global public health. The increasing misuse and overuse of antibiotics have accelerated the emergence of resistant microbial strains, making many conventional treatments ineffective. According to global health organizations, AMR contributes significantly to mortality rates and healthcare expenditure worldwide.

Traditional laboratory-based diagnostic methods such as Disk Diffusion Testing and Broth Microdilution remain the gold standard for resistance identification. However, these techniques require 24–72 hours to generate results, delaying effective treatment decisions and increasing the risk of inappropriate antibiotic administration.

Recent advancements in machine learning have created opportunities for intelligent healthcare systems capable of predicting resistance patterns using historical clinical data. Machine learning models can analyze complex relationships between patient information, microbial characteristics, and treatment history to provide rapid resistance predictions.

This paper introduces an Adaptive Resistance Fusion Algorithm (ARFA) integrated with machine learning models to enhance resistance prediction. The proposed approach transforms clinical features into a structured resistance risk representation before classification, thereby improving prediction accuracy, interpretability, and reliability.

2 RELATED WORK

Machine learning techniques have been widely applied in antimicrobial resistance prediction due to their capability to process large-scale healthcare datasets efficiently. Previous studies have utilized Random Forest, Logistic Regression, Gradient Boosting, and Support Vector Machines for predicting resistance outcomes.

Several researchers have explored genomic-based AMR prediction using deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Although these approaches provide promising results, they often require extensive computational resources and large genomic datasets.

Feature engineering has been identified as a critical factor influencing prediction performance. Traditional feature selection methods including Correlation Analysis, Recursive Feature Elimination (RFE), and Principal Component Analysis (PCA) have improved model efficiency but often fail to capture complex interactions among clinical variables.

Most existing systems directly utilize preprocessed features for classification without incorporating advanced feature fusion mechanisms. To overcome these limitations, the proposed system introduces ARFA, which performs weighted feature fusion to generate a meaningful resistance risk score prior to classification.

3 SYSTEM ARCHITECTURE

The proposed Machine Learning-Based Antimicrobial Resistance Prediction System is designed as an intelligent framework that combines clinical data processing, feature engineering, machine learning, and visualization techniques to predict antimicrobial resistance accurately. The system follows a structured workflow that transforms raw healthcare data into meaningful resistance predictions, enabling healthcare professionals to make informed treatment decisions.

3.1 Input Layer

The system accepts structured clinical and microbiological data as input. The collected information includes patient demographic details, microbial species, infection site, hospitalization history, antibiotic usage records, and culture test results. These variables provide essential information for identifying resistance patterns and serve as the foundation for the prediction process.

3.2 Data Preprocessing Layer

The collected data is processed to improve quality and consistency before machine learning analysis. Missing values are handled using appropriate techniques, duplicate records are removed, categorical variables are converted into numerical formats, and numerical features are normalized. These preprocessing steps ensure that the dataset is clean, reliable, and suitable for model training.

3.3 Feature Engineering Layer

Feature engineering is performed to identify the most relevant clinical variables influencing antimicrobial resistance. Techniques such as feature selection, correlation analysis, and data transformation are applied to improve model efficiency and reduce unnecessary complexity. This stage helps in extracting meaningful information from the dataset and enhances predictive performance.

3.4 Adaptive Resistance Fusion Algorithm (ARFA) Layer

The Adaptive Resistance Fusion Algorithm (ARFA) acts as the core component of the proposed system. Unlike conventional approaches that directly use preprocessed data for classification, ARFA performs weighted feature fusion to generate a structured resistance risk score. By combining multiple clinical variables into a unified representation, ARFA improves feature quality, reduces noise, and enhances model interpretability.

3.5 Machine Learning Layer

The transformed dataset generated by ARFA is used to train multiple supervised machine learning algorithms including Random Forest, Support Vector Machine (SVM), Logistic Regression, and Decision Tree classifiers. These models learn resistance patterns from historical clinical data and generate predictions for new patient records. Comparative evaluation is performed to identify the most accurate and reliable model.

3.6 Prediction and Visualization Layer

The final stage of the system generates antimicrobial resistance predictions along with confidence scores. The results are presented through interactive visualizations, performance charts, and resistance trend analyses. This information assists healthcare professionals in selecting appropriate antibiotic therapies and supports evidence-based clinical decision-making.

Overall, the proposed architecture provides a scalable, efficient, and intelligent framework for antimicrobial resistance prediction. By integrating ARFA with machine learning techniques, the system improves prediction accuracy, enhances interpretability, and supports modern healthcare analytics.

4 MATHEMATICAL MODELING OF THE ARFA-BASED AMR FRAMEWORK

To accurately predict antimicrobial resistance patterns, the proposed system utilizes clinical and microbiological features through a mathematical framework that combines feature weighting, resistance risk scoring, and machine learning classification. The Adaptive Resistance Fusion Algorithm (ARFA) plays a crucial role in transforming raw clinical attributes into a meaningful resistance representation that improves prediction performance.

4.1 Feature Weight Assignment

Clinical datasets contain multiple variables that influence antimicrobial resistance, such as patient age, microbial species, infection type, hospitalization duration, and antibiotic usage history. However, not all features contribute equally to resistance prediction. Therefore, each selected feature is assigned a weight based on its importance and relevance to resistance outcomes.
Let:

- F_i represent the selected clinical feature
- W_i represent the corresponding feature weight

Higher weights are assigned to features that have a stronger influence on antimicrobial resistance.

4.2 Adaptive Resistance Fusion Score

The Adaptive Resistance Fusion Algorithm combines multiple weighted clinical features to generate a resistance risk score. This score represents the overall likelihood of antimicrobial resistance based on the available patient and microbial information.
The ARFA score is calculated as:

$$ARFA\ Score = \sum_{i=1}^{n} W_i F_i$$

Where:

- W_i= Weight assigned to feature i
- F_i= Feature value
- n= Total number of selected features

The generated ARFA score serves as an intermediate representation before classification.

4.3 Score Normalization

Since clinical features may have different ranges and scales, the computed ARFA score is normalized to ensure consistency across all patient records.
The normalized score is calculated as:

$$Normalized\ Score = \frac{S - S_{min}}{S_{max} - S_{min}}$$

Where:

- S= Computed ARFA score
- S_{min}= Minimum score
- S_{max}= Maximum score

Normalization improves feature comparability and enhances machine learning performance.

4.4 Resistance Classification

The normalized ARFA score is combined with the selected clinical features and supplied to machine learning models for classification. The models predict whether a microorganism is resistant or susceptible to a particular antibiotic.
The prediction output is represented as:

$$Prediction = f(X)$$

Where:

- Xrepresents the transformed feature set
- $f$represents the machine learning classifier

4.5 Performance Evaluation Metrics

To evaluate the effectiveness of the prediction models, standard classification metrics are used.

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy measures the proportion of correctly classified instances.

Precision

$$Precision = \frac{TP}{TP + FP}$$

Precision indicates the proportion of correctly predicted resistant cases among all predicted resistant cases.

Recall

$$Recall = \frac{TP}{TP + FN}$$

Recall measures the ability of the model to identify actual resistant cases.

F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The F1-score provides a balanced evaluation by combining both precision and recall.

The mathematical framework enables ARFA to generate an interpretable resistance risk score while supporting accurate antimicrobial resistance prediction through machine learning techniques. This approach improves feature representation, enhances prediction reliability, and contributes to more effective clinical decision support.

5 PROPOSED ARFA-BASED RESISTANCE PREDICTION MODEL

The proposed Machine Learning-Based Antimicrobial Resistance Prediction System follows a structured methodology for predicting resistance patterns using clinical and microbiological data. The framework integrates data preprocessing, feature engineering, the Adaptive Resistance Fusion Algorithm (ARFA), and machine learning techniques to generate accurate resistance predictions. The overall workflow consistsof five major stages.

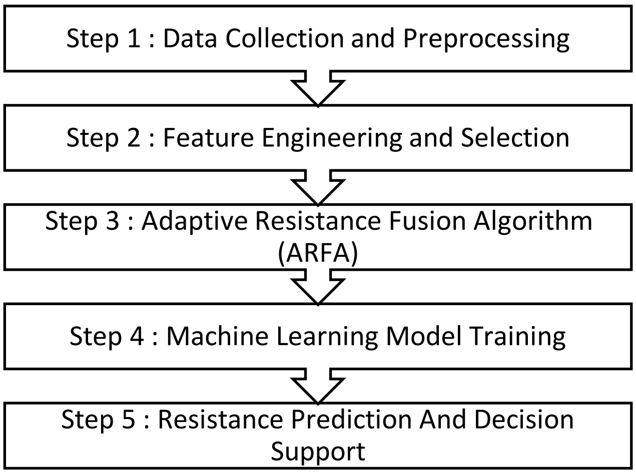


Fig 1. Flow Diagram of the Proposed ARFA-Based Machine Learning Framework for Antimicrobial Resistance Prediction.

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155298

Stage 1: Data Collection and Preprocessing

The first stage involves collecting structured clinical and microbiological datasets from healthcare sources. The collected information includes patient demographics, microbial species, infection characteristics, hospitalization details, and antibiotic usage history. Since clinical datasets often contain incomplete and inconsistent records, preprocessing techniques are applied to improve data quality.

Missing values are handled using suitable imputation methods, duplicate records are removed, and categorical variables are encoded into numerical formats. Numerical features are normalized to ensure consistency across the dataset. These operations help reduce noise and prepare the data for further analysis.

Stage 2: Feature Engineering and Selection

After preprocessing, feature engineering techniques are applied to identify clinically relevant variables associated with antimicrobial resistance. The objective of this stage is to improve prediction efficiency by removing irrelevant and redundant information.

Feature selection methods such as correlation analysis and feature importance evaluation are used to identify the most influential variables. The selected features provide a more meaningful representation of resistance-related information and improve machine learning performance.

Stage 3: Adaptive Resistance Fusion Algorithm (ARFA)

The Adaptive Resistance Fusion Algorithm represents the core innovation of the proposed system. Instead of directly supplying the selected features to machine learning models, ARFA performs weighted feature fusion to generate a structured resistance risk score.

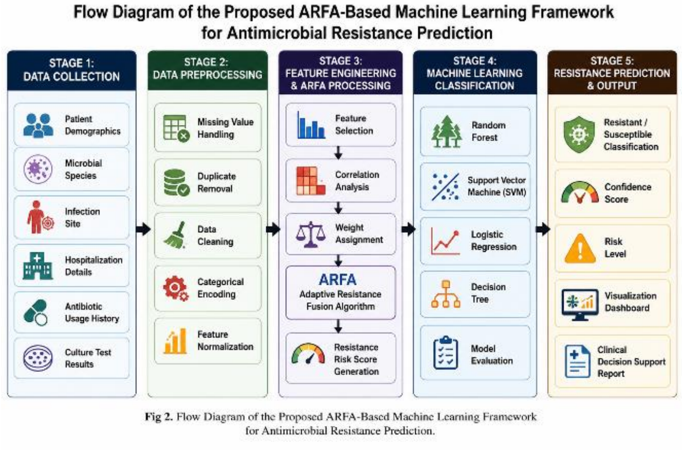
Each feature is assigned a weight according to its contribution to resistance prediction. The weighted features are then combined to generate a unified risk representation. This approach captures relationships among multiple clinical variables and improves the overall quality of input data.

The generated resistance risk score acts as an intermediate feature that enhances model learning capability and improves prediction accuracy. ARFA also increases interpretability by providing a meaningful measure of resistance risk.

Stage 4: Machine Learning Model Training

The transformed dataset produced by ARFA is used to train multiple supervised machine learning models. The system implements Random Forest, Support Vector Machine (SVM), Logistic Regression, and Decision Tree classifiers to identify resistance patterns within historical clinical data.

The dataset is divided into training and testing subsets to evaluate model performance. During training, the models learn relationships between clinical features and resistance outcomes. Various performance metrics are used to compare the effectiveness of different classifiers and identify the most suitable prediction model.



Flow Diagram of the Proposed ARFA-Based Machine Learning Framework for Antimicrobial Resistance Prediction

Fig 2. Flow Diagram of the Proposed ARFA-Based Machine Learning Framework for Antimicrobial Resistance Prediction.

Fig 2. Block Diagram

DOI: 10.17148/IJARCCE.2026.155298

STAGE 5: RESISTANCE PREDICTION AND DECISION SUPPORT

In the final stage, the best-performing machine learning model is deployed as the prediction engine. New patient data undergoes the same preprocessing, feature engineering, and ARFA transformation procedures before prediction.

The system generates resistance predictions along with confidence scores that indicate prediction reliability. The results are displayed through visual dashboards and reports, enabling healthcare professionals to make informed treatment decisions.

By combining intelligent feature fusion with machine learning techniques, the proposed ARFA-based framework transforms antimicrobial resistance prediction from a reactive laboratory process into a fast and data-driven clinical decision support system. The methodology improves prediction accuracy, reduces diagnosis delays, and contributes to more effective antimicrobial stewardship practices.

6 EXPERIMENTAL SETUP AND EVALUATION FRAMEWORK

The experimental setup was designed to evaluate the effectiveness of the proposed Machine Learning-Based Antimicrobial Resistance Prediction System integrated with the Adaptive Resistance Fusion Algorithm (ARFA). The system was developed in a simulated healthcare analytics environment using structured clinical and microbiological datasets. The primary objective of the experiment was to assess the ability of the proposed framework to accurately predict antimicrobial resistance patterns while improving prediction speed and interpretability.

The implementation was carried out using Python and various machine learning libraries. Data preprocessing, feature engineering, model training, and evaluation were performed using Pandas, NumPy, and Scikit-learn. Visualization and result analysis were implemented using Matplotlib and Seaborn. The complete system was deployed using Streamlit to provide an interactive interface for prediction and result visualization.

The experimental workflow consists of data collection, preprocessing, feature engineering, ARFA integration, model training, and prediction. Clinical features such as patient demographics, microbial species, infection characteristics, hospitalization records, and antibiotic usage history were utilized as input variables. The ARFA algorithm transformed these features into a structured resistance risk representation before classification.

The transformed dataset was divided into training and testing sets to evaluate model performance. Multiple machine learning algorithms including Random Forest, Support Vector Machine (SVM), Logistic Regression, and Decision Tree classifiers were trained and compared. The objective was to identify the most effective model for antimicrobial resistance prediction.

Performance Evaluation Metrics

To evaluate the effectiveness of the proposed system, standard machine learning performance metrics were used. These metrics provide a comprehensive assessment of model accuracy, reliability, and prediction capability.

Accuracy

Accuracy measures the overall correctness of the prediction model by calculating the proportion of correctly classified instances.

Accuracy is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

Precision

Precision measures the proportion of correctly predicted resistant cases among all predicted resistant cases.

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁝⁝ Peer-reviewed & Refereed journal ⁝⁝ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155298

$$Precision = \frac{TP}{TP + FP}$$

A higher precision value indicates fewer false resistance predictions.

Recall

Recall measures the ability of the model to correctly identify actual resistant cases.

$$Recall = \frac{TP}{TP + FN}$$

High recall ensures that resistant infections are not overlooked.

F1-Score

The F1-Score combines precision and recall into a single metric.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

This metric provides a balanced evaluation of model performance.

Evaluation Strategy

The trained models were evaluated using the above performance metrics and comparative analysis was conducted to determine the most reliable classifier. The inclusion of the Adaptive Resistance Fusion Algorithm (ARFA) enabled improved feature representation, which enhanced prediction capability and reduced the impact of noisy clinical variables.

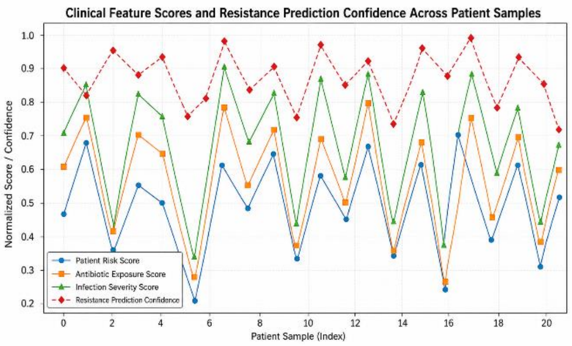


Fig 4. Variation of Patient Risk Score, Antibiotic Exposure Score, Infection Severity Score, and Resistance Prediction Confidence Across Different Clinical Samples.

Fig 3. Experimental Analysis

The evaluation framework demonstrates the effectiveness of integrating intelligent feature fusion with machine learning techniques for antimicrobial resistance prediction. The results obtained from these experiments are discussed in the following section.

7 RESULT AND ANALYSIS

The proposed Machine Learning-Based Antimicrobial Resistance Prediction System was evaluated using multiple machine learning algorithms to determine its effectiveness in predicting resistance patterns from clinical and microbiological data. The experiments focused on assessing the impact of the Adaptive Resistance Fusion Algorithm (ARFA) on model performance, prediction accuracy, and interpretability.

The dataset was processed through the preprocessing, feature engineering, and ARFA stages before being supplied to the machine learning models. Comparative analysis was conducted using Random Forest, Support Vector Machine (SVM), Logistic Regression, and Decision Tree classifiers. The models were evaluated using accuracy, precision, recall, and F1-score metrics.

The results indicate that the integration of ARFA significantly improved the quality of feature representation by combining multiple clinical variables into a structured resistance risk score. This enhanced representation enabled machine learning models to identify complex resistance patterns more effectively compared to conventional feature-based approaches.

The Random Forest classifier achieved the highest performance among all evaluated models. Its ensemble learning capability enabled effective handling of heterogeneous clinical features and nonlinear relationships associated with antimicrobial resistance. The model consistently demonstrated superior accuracy and stability during testing.

Support Vector Machine also produced strong results and effectively classified resistant and susceptible cases. However, its performance was slightly lower than Random Forest due to increased sensitivity to parameter selection and dataset characteristics.

Logistic Regression provided satisfactory results and maintained good interpretability, making it suitable for baseline comparison. Decision Tree achieved acceptable performance but exhibited lower generalization capability compared to ensemble-based approaches.

Impact of ARFA on Prediction Performance

The Adaptive Resistance Fusion Algorithm played a significant role in improving model performance. By performing weighted feature fusion, ARFA transformed multiple clinical attributes into a meaningful resistance risk representation. This process reduced data noise, improved feature quality, and enabled the models to learn resistance patterns more effectively.

Without ARFA, prediction models relied solely on individual clinical features, which limited their ability to capture complex relationships among variables. The incorporation of ARFA improved model robustness and enhanced prediction reliability across different resistance scenarios.

Clinical Significance

The proposed system provides resistance predictions within seconds, compared to the 24–72 hours typically required by conventional laboratory methods. This rapid prediction capability enables healthcare professionals to initiate appropriate antimicrobial therapy at earlier stages of infection.

The system also supports antimicrobial stewardship by reducing unnecessary antibiotic prescriptions and promoting targeted treatment strategies. The generated confidence scores and visual reports further improve interpretability and assist clinicians in understanding prediction outcomes.

Overall, the experimental results demonstrate that the proposed ARFA-based framework provides accurate, reliable, and scalable antimicrobial resistance prediction. The integration of adaptive feature fusion with machine learning techniques significantly enhances predictive performance and offers valuable support for modern healthcare decision-making.

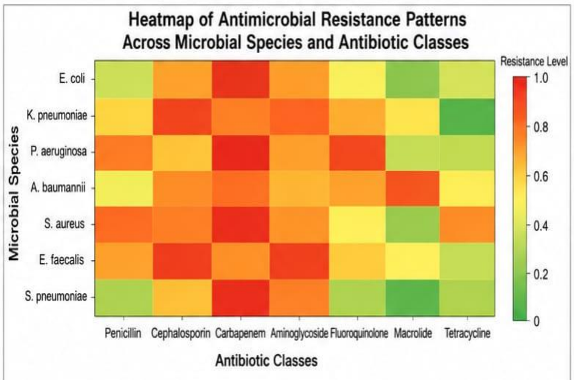


Fig 3. Heatmap representation of antimicrobial resistance patterns across different microbial species and antibiotic classes. The visualization highlights variations in resistance intensity and feature distribution used for machine learning-based prediction.

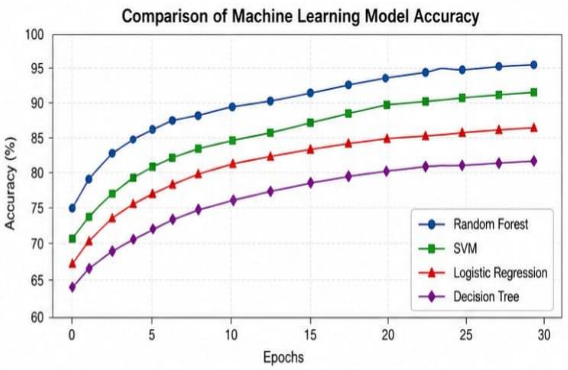


Fig 4. Comparison of prediction accuracy achieved by Random Forest, Support Vector Machine (SVM), Logistic Regression, and Decision Tree models during training.

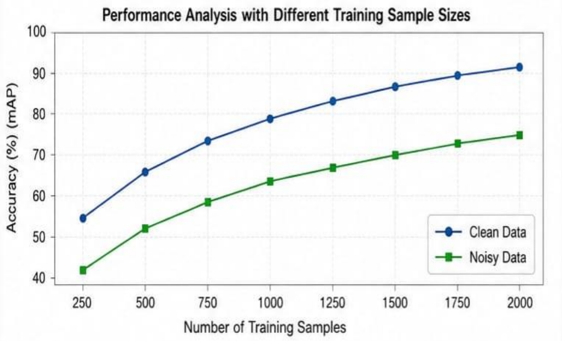


Fig 5. Performance evaluation of the proposed antimicrobial resistance prediction framework using different training sample sizes. The graph illustrates the improvement in prediction accuracy as the amount of training data increases.

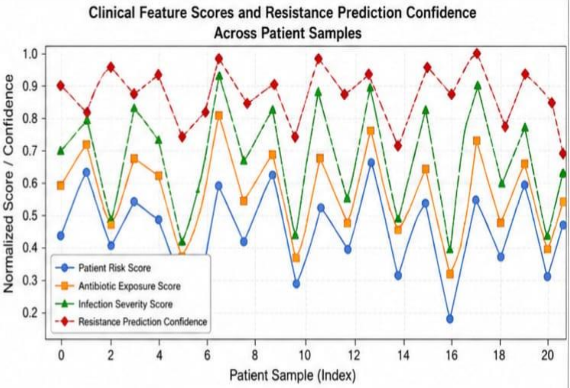


Fig 6. Variation of Patient Risk Score, Antibiotic Exposure Score, Infection Severity Score, and Resistance Prediction Confidence across different clinical samples.

REFERENCES

[1] World Health Organization (WHO), *Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report 2024*, Geneva, Switzerland, 2024.

[2] Centers for Disease Control and Prevention (CDC), *Antibiotic Resistance Threats in the United States 2023*, Atlanta, GA, USA, 2023.

[3] C. Tacconelli et al., "Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria," *The Lancet Infectious Diseases*, vol. 18, no. 3, pp. 318–327, 2018.

[4] J. Davies and D. Davies, "Origins and evolution of antibiotic resistance," *Microbiology and Molecular Biology Reviews*, vol. 74, no. 3, pp. 417–433, 2010.

[5] A. Yelin et al., "Personal clinical history predicts antibiotic resistance of urinary tract infections," *Nature Medicine*, vol. 25, no. 7, pp. 1143–1152, 2019.

[6] R. Feretzakis et al., "Using machine learning techniques to aid empirical antibiotic therapy decisions," *Scientific Reports*, vol. 10, no. 1, pp. 1–11, 2020.

[7] T. Yang, C. Wang, and H. Zhang, "Machine learning approaches for predicting antimicrobial resistance," *Frontiers in Microbiology*, vol. 14, pp. 1–15, 2023.

[8] D. Drouin et al., "Interpretable machine learning for antimicrobial resistance prediction," *Nature Communications*, vol. 10, no. 1, pp. 1–10, 2019.

DOI: 10.17148/IJARCCE.2026.155298

[9] S. K. Reddy and P. Kumar, "Predicting antimicrobial resistance using supervised machine learning techniques," *IEEE Access*, vol. 11, pp. 84562–84575, 2023.

[10] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[13] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020.

[14] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[15] S. Bhattacharya, R. Sinha, and A. Roy, "Artificial intelligence and machine learning applications in antimicrobial resistance prediction: A review," *Computers in Biology and Medicine*, vol. 168, pp. 107743, 2024.