Journal of Communicable Diseases
Volume 58, Issue 1 - 2026, Pg. No. 132-144
Peer Reviewed & Open Access Journal

Research Article

Supervised Machine Learning for MDR Detection in Gram-Positive Pathogens: A Cross-Sectional Hospital-Based Dataset Analysis

Umaarasu Thirunavukkarasu[1], Thirunavukkarasu Dharmalingam[2], Suprraja Subramani[3], Senthil Kumar S[4], Loganathan Thangavel[5], Arbind Kumar Choudhary[6]

[1]Assistant Professor, Department of Microbiology, Swamy Vivekanandha Medical College Hospital and Research Institute, Tiruchengode, India
[2]Assistant Professor, Department of Microbiology, Government Mohan Kumaramangalam Medical College, Salem, India
[3]Assistant Professor, Faculty of Pharmacy, Karpagam Academy of Higher Education, Coimbatore, India
[4,5]Saveetha College of Physiotherapy, Saveetha Institute of Medical and Technical Sciences, SIMATS, Chennai, India
[6]Department of Pharmacology, Government Erode Medical College and Hospital, Erode, India
DOI: https://doi.org/10.24321/0019.5138.202618

I N F O

Corresponding Author:
Arbind Kumar Choudhary, Department of Pharmacology, Government Erode Medical College and Hospital, Erode, India
E-mail Id:
arbindkch@gmail.com
Orcid Id:
https://orcid.org/0000-0001-8910-1745
How to cite this article:
Thirunavukkarasu U, Dharmalingam T, Subramani S, Kumar S S, Thangavel L, Choudhary A K, Supervised Machine Learning for MDR Detection in Gram-Positive Pathogens: A Cross-Sectional Hospital-Based Dataset Analysis. J Commun Dis. 2026;58(1):132-144.

Date of Submission: 2025-09-08
Date of Acceptance: 2026-01-11

A B S T R A C T

Background: Multidrug-resistant (MDR) Gram-positive pathogens are a growing challenge in hospital practice due to limited treatment options and poor clinical outcomes. This study assessed MDR prevalence among major Gram-positive pathogens and evaluated supervised machine learning models for MDR prediction.

Methods: A cross-sectional hospital-based study was conducted on 458 clinical specimens from a tertiary care hospital in Salem, Tamil Nadu, India, including 393 culture-positive Gram-positive isolates. MDR was defined as resistance to three or more antimicrobial classes. Logistic regression, random forest, XGBoost, and multilayer perceptron models were developed using demographic, clinical, and microbiological variables and evaluated using accuracy, precision, recall, F1-score, and AUC-ROC.

Results: Enterococcus was the predominant isolate (67.0%), followed by Streptococcus (15.5%) and MRSA (3.3%). MDR was identified in 82.9% of Enterococcus isolates, 76.9% of Streptococcus isolates, and all MRSA isolates. MDR distribution did not differ meaningfully by age or gender. Random forest and XGBoost showed the most balanced performance, with accuracy of about 0.68-0.70, F1-scores of 0.77-0.80, and AUC-ROC values of 0.58-0.59. The multilayer perceptron achieved the highest recall (1.00) and F1-score (0.91), although with lower discrimination (AUC-ROC 0.46). Older age, culture negativity, and Enterococcus isolation were the strongest predictors of MDR.

Conclusions: Gram-positive pathogens in this setting showed a high burden of multidrug resistance. Machine learning models demonstrated moderate predictive value and may complement routine microbiological surveillance and antimicrobial stewardship efforts.

Keywords: Multidrug Resistance, Enterococcus, Streptococcus, MRSA, Machine Learning, Antimicrobial Stewardship, Tertiary Care, Predictive Modelling

Introduction

Antimicrobial resistance has become a major global public health concern, affecting the treatment of both community-acquired and hospital-associated infections. Among resistant organisms, Gram-positive pathogens remain especially important because they are frequently associated with bloodstream infections, wound infections, respiratory tract infections, urinary tract infections, and device-related infections.[1,2] Common Gram-positive organisms such as Enterococcus spp., Streptococcus spp., and Staphylococcus aureus continue to cause substantial morbidity, prolonged hospital stay, and increased healthcare costs, particularly when they exhibit resistance to multiple antimicrobial agents.[3,4]

Multidrug resistance (MDR) in Gram-positive pathogens presents a serious challenge to clinical management. Resistance to commonly used antibiotics narrows therapeutic options, complicates empirical treatment, and may contribute to treatment failure and adverse patient outcomes.[5] In hospital settings, the burden of MDR is further influenced by factors such as repeated antibiotic exposure, invasive procedures, prolonged admission, and local prescribing practices. Because resistance patterns vary across regions and institutions, local epidemiological data are essential for guiding antimicrobial stewardship and infection control strategies.[6,7]

Conventional microbiological methods remain the foundation for organism identification and antimicrobial susceptibility testing; however, these approaches may not always support rapid risk stratification at the point of care.[8,9] In recent years, machine learning techniques have gained attention in clinical research for their ability to analyse complex datasets, identify hidden patterns, and support predictive decision-making. In the context of infectious diseases, supervised machine learning models may help estimate the likelihood of MDR using demographic, clinical, and laboratory variables, thereby complementing routine microbiological surveillance.[11]

Despite increasing interest in predictive modelling, there remains limited evidence on the application of machine learning for MDR prediction among Gram-positive pathogens in hospital-based datasets from India. In addition, the distribution of Gram-positive isolates and their resistance patterns may differ across institutions, making context-specific analysis necessary.[12]

Therefore, the present study was undertaken to determine the prevalence of multidrug resistance among Gram-positive pathogens isolated from clinical specimens in a tertiary care hospital and to evaluate the performance of supervised machine learning models in predicting MDR using demographic, clinical, and microbiological variables.

Materials and Methods

Study Design and Setting

This is a single-centre, cross-sectional, hospital-based study conducted at a tertiary care hospital in Salem, Tamil Nadu, India, with the primary objective of characterising MDR among GP pathogens—Enterococcus, Streptococcus, and MRSA. The study was approved by the institutional review board at the host facility, and informed consent was waived due to the use of de-identified, routinely collected clinical and microbiological data. The overall methodological workflow, from specimen collection to ML model evaluation, is summarised in Figure 1.

Study Population and Data Collection

All consecutive clinical specimens, including blood, pus, sputum, urine, endotracheal tube aspirates, pleural fluid, and stool, received in the hospital microbiology laboratory from January 2023 to December 2023 were prospectively included.Demographic data (age and gender), clinical specimen source, and bacterial culture results (organism identification and antimicrobial susceptibility testing) were extracted from the laboratory information system. Specimens yielding "No growth" or non-GP pathogens were retained in the demographic analysis but excluded from organism-specific resistance calculations. MDR was defined as resistance to three or more classes of antibiotics based on local laboratory protocols and international guidelines.

A total of 458 clinical specimens were included for machine learning analysis. Culture-positive Gram-positive isolates (n = 393) were used for organism-specific and resistance analyses. Minor variations in isolate counts across tables reflect subgroup-specific exclusions and specimen-level stratification applied during analysis.

All specimens were processed using standard microbiological techniques. Identification of bacterial isolates was performed using conventional biochemical methods and automated systems as per clinical laboratory protocols. Antimicrobial susceptibility profiles were determined for each GP isolate against a panel of clinically relevant antibiotics, with categorisation as susceptible, intermediate, or resistant according to the Clinical and Laboratory Standards Institute (CLSI) or European Committee on Antimicrobial Susceptibility Testing (EUCAST) breakpoints. MDR status was recorded as a binary outcome (Yes/ No) based on the susceptibility results for each isolate. The study proposal was approved by the Government Mohan Kumaramangalam Medical College & Hospital Ethical Committee prior to the initiation of the study (Ref: 7713/ME1/2023). The overall methodological workflow, from specimen collection to ML model evaluation, is summarised in Figure 1.

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202618

CONSORT-style flow diagram depicting specimen collection (N = 458), exclusion of non-Gram-positive isolates (N = 65), inclusion of Gram-positive isolates in the final analysis (N = 393), MDR classification, data preprocessing, development of supervised machine learning models, and model evaluation using accuracy, precision, recall, F1-score, and AUC-ROC.

The figure 1 diagram shows the inclusion of clinical specimens (N = 458), exclusion of non–Gram-positive isolates (N = 65), and final inclusion of Gram-positive isolates (N = 393). Organism distribution included Enterococcus (n = 307), Streptococcus(n = 71), and methicillin-resistant Staphylococcus aureus (MRSA) (n = 15). Multidrug resistance (MDR) was defined as resistance to three or more antibiotic classes. Data preprocessing steps included encoding of demographic and clinical variables, one-hot encoding of organism categories, and handling of missing data. Supervised machine learning models—logistic regression, random forest, extreme gradient boosting (XGBoost), and multilayer perceptron (MLP)—were developed to predict MDR status. Model performance was evaluated using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC).

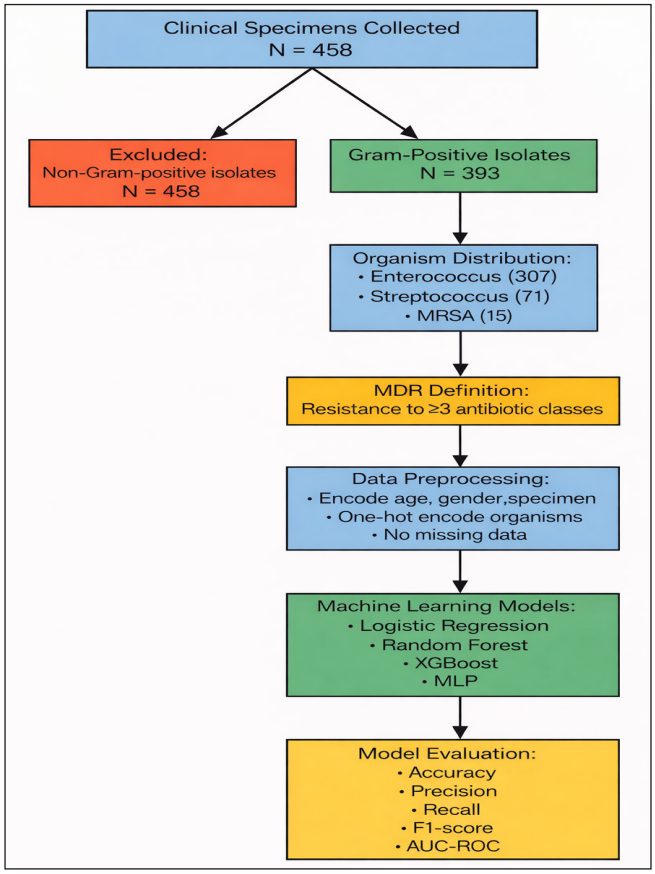

Figure 1.CONSORT-style flow diagram illustrating specimen selection and machine learning

Data Preprocessing and Feature Engineering

For ML analysis, the dataset was structurally cleaned and standardised. Age was included as a continuous variable. Gender was encoded as a binary variable (female = 0, male = 1), and minor inconsistencies in data entry were harmonised. Specimen types (blood, pus, sputum, urine, and others) were each represented as binary indicator variables (0 = absent, 1 = present); any non-zero value in the "Other" specimen field (endotracheal tube, pleural fluid, and stool) was recoded to 1. Organism categories (Enterococcus, Streptococcus, MRSA, and "No growth") were one-hot encoded to avoid implicit ordinal relationships. No missing data were present in the selected variables, obviating the need for imputation. MDR status served as the binary target variable for supervised learning.

Supervised Machine Learning Modelling

Four supervised ML algorithms were evaluated for MDR prediction: logistic regression (as a baseline interpretable model), random forest, XGBoost, and a multi-layer perceptron (MLP) neural network. The dataset was randomly split into training (70%) and test (30%) sets, preserving the distribution of MDR status in both subsets. Model performance was assessed using standard metrics: accuracy, precision, recall, F1-score, and area under the receiver-operating characteristic curve (AUC-ROC). Feature importance was extracted from the random forest model to identify key clinical and microbiological predictors of MDR.

Statistical Analysis

Descriptive statistics (counts, percentages, means, standard deviations, medians, and ranges) were used to summarise demographic, clinical, and resistance data. Chi-square tests (or Fisher's exact test for small cell counts) were employed to assess differences in MDR prevalence across age groups, gender, and specimen types. All analyses were performed using Python (pandas, scikit-learn, xgboost, tensorflow/ keras) and R, with statistical significance set at $p < 0.05$.

Results

Demographic and Clinical Characteristics of the Study Cohort with a Nested Specimen × Organism × MDR Breakdown:

The study enrolled 458 patients from a tertiary care hospital setting (Table 1), providing a comprehensive overview of demographic, clinical, and microbiological characteristics relevant to MDR GP infections. The cohort reflects a typical hospitalised population with diverse ages, genders, and infection sources, supporting robust ML analysis for MDR prediction.

†GP: Gram-positive isolates only (Enterococcus, Streptococcus, and MRSA); excludes "No growth" and non-Gram-positive cases MDR: Multidrug resistance, defined as

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202618

resistance to ≥ 3 antibiotic classes Percentages are calculated as the proportion of MDR cases within each subgroup. Specimen types are mutually exclusive per patient in this dataset.

"Others" includes ET tube, stool, and pleural fluid specimens. Total specimen counts may not sum up to Gram-positive totals due to some patients contributing multiple specimen types.

The cohort demonstrates a remarkably high prevalence of MDR among GP pathogens (84.4% for Enterococcus, 81.7% for Streptococcus, and 100% for MRSA), with no significant variation by age, gender, or specimen source. Enterococcus predominates across all infection sites, underscoring its role as the leading cause of resistant infections in this setting. The uniformly elevated MDR rates highlight an urgent need for enhanced antimicrobial stewardship and alternative therapeutic strategies. The nested breakdown by specimen and organism provides actionable insights for empirical therapy and infection control, and confirms the dataset's suitability for predictive modelling of MDR in hospitalised patientsDifferences in specimen counts across tables reflect data preprocessing steps applied during machine learning feature engineering, where specimen categories were standardised and recoded.

Table 1.Demographic and Clinical Characteristics of the Study Cohort

Variable	Category/ Subgroup	Count	Percentage (%)	MDR Yes (%)	Notes
All Patients	—	458	100.0	—	—
Age (years)	Mean ± SD, Range	46.6 ± 22.4, 2–87	—	—	Age distribution spanning paediatric to the elderly
	< 20	41	9.0	—	—
	20–39	156	34.1	—	—
	40–59	109	23.8	—	—
	≥ 60	152	33.2	—	—
Gender	Female	231	50.4	—	Nearly equal gender representation
	Male	227	49.6	—	—
Specimen source	Pus	109	23.8	—	—
	Urine	106	23.1	—	—
	Sputum	84	18.3	—	—
	Blood	47	10.3	—	—
	Others (ET, stool, pleural)	90	19.7	—	—
Gram-positive isolates	*Enterococcus*	307	67.0	84.4	Dominant organism; high MDR rate
	Streptococcus	71	15.5	81.7	—
	MRSA	15	3.3	100.0	All MRSA isolates were MDR
	No growth	65	14.2	—	Excluded from MDR modelling
MDR by age (GP†)	< 20	35	85.4	85.7	Consistently high MDR rates across age groups
	20–39	141	90.4	80.9	—
	40–59	90	82.6	85.6	—
	≥ 60	127	83.6	83.5	—

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202618

MDR by gender (GP†)	Female	195	84.4	84.6	No major gender disparity in MDR
	Male	198	87.2	83.8	—
MDR by specimen (GP†)	Blood	38	80.9	86.8	High MDR prevalence seen in bloodstream infections
	Pus	100	91.7	89.0	—
	Sputum	74	88.1	87.8	—
	Urine	88	83.0	86.4	—
	Other	73	81.1	86.3	—

Nested Specimen × Organism × MDR Breakdown

	Organism	Count	MDR Yes	MDR Yes (%)
Specimen Blood	*Enterococcus*	30	26	86.7
	Streptococcus	8	7	87.5
	MRSA	0	—	—
Pus	*Enterococcus*	85	76	89.4
	Streptococcus	14	13	92.9
	MRSA	1	1	100.0
Sputum	*Enterococcus*	61	54	88.5
	Streptococcus	13	11	84.6
	MRSA	0	—	—
Urine	*Enterococcus*	66	57	86.4
	Streptococcus	12	10	83.3
	MRSA	0	—	—
Others	*Enterococcus*	65	56	86.2
	Streptococcus	8	7	87.5
	MRSA	0	—	—

MRSA: Methicillin-resistant *Staphylococcus aureus*

Distribution of Gram-Positive Pathogen Isolates

The distribution of Gram-positive pathogen isolates, including *Enterococcus*, *Streptococcus*, *methicillin-resistant Staphylococcus* aureus (MRSA), and cases with no growth, is presented in Table 2. The table also summarises the overall distribution of culture-positive Gram-positive isolates derived from the study dataset.

Gram-positive (GP) isolates include *Enterococcus, Streptococcus*, and MRSA only; "No growth" and non-Gram-positive pathogens were excluded from this analysis. Sex: F: female, M: male; based on recorded gender in the dataset (minor whitespace variations in records were standardised for analysis). Age: Median and range in years for each organism group. % of all specimens: Percentage of each isolate type out of all 458 submitted specimens. % of Gram-positive isolates: Percentage of each isolate type out of all confirmed Gram-positive isolates (N = 393). "No growth": Culture-negative specimens; excluded from Gram-positive isolate totals and resistance analyses. Total GP isolates: Sum of *Enterococcus, Streptococcus*, and MRSA specimens (N = 393).

Enterococcus dominates the GP landscape in this cohort (78.1% of GP isolates), with Streptococcus and MRSA representing smaller but clinically significant proportions (18.1% and 3.8%, respectively). The culture-negative rate ("No growth") is 14.2%, which is consistent with tertiary hospital microbiology practice and should be considered

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202618

when interpreting resistance patterns. The sex and age distributions reflect a broad, adult-predominant hospitalised population, with no major sex bias in pathogen distribution. This breakdown establishes the epidemiologic foundation for subsequent resistance and ML analyses, and highlights the predominance of Enterococcus as the key target for MDR surveillance and intervention in this setting.The distribution and resistance patterns of Gram-positive pathogens are presented across Tables 2-5. While the overall number of Gram-positive isolates was 393, slight variations in subgroup totals are observed in stratified analyses due to specimen-level categorisation and exclusion criteria. However, the overall trends in multidrug resistance remained consistent across analyses.

Table 2.Distribution of Gram-Positive Pathogen Isolates

Category	Count	% of All Specimens (N = 458)	% of Gram-Positive Isolates (N = 393)	Sex (F/M)	Age (Median, Range)
Enterococcus	307	67.0	78.1	141/166	48 (2–87)
Streptococcus	71	15.5	18.1	38/33	53 (2–87)
MRSA	15	3.3	3.8	7/8	54 (17–80)
No growth	65	14.2	—	—	—
Total GP isolates	393	85.8	100.0	—	—

MRSA: Methicillin-resistant Staphylococcus aureus, GP: Gram-positive only

Prevalence of Multidrug Resistance (MDR)

The prevalence of multidrug resistance (MDR) among culture-positive Gram-positive (GP) pathogens isolated from clinical specimens in a hospital setting is summarised in Table 3. The stratified assessment presented in Table 3 by organism, specimen type, age, and gender provides a detailed understanding of resistance patterns, which is essential for guiding antimicrobial stewardship strategies and informing the development of predictive models for MDR detection.

(Enterococcus, Streptococcus, MRSA); excludes "No growth" and other isolates. Sex: Females and males as defined in the dataset (minor coding inconsistencies standardised for analysis). Most common specimen: Determined by the highest absolute count per organism. Specimen-specific counts (n): Total isolates of organism from that specimen. MDR Yes, MDR%: Number and percentage of MDR isolates within each subgroup. Age: Median age in years for each organism group. Gaps (—): Not calculated or not applicable.

Table 3.Prevalence of Multidrug Resistance by Organism, Specimen, and Demographics

Organism	Total Isolates	MDR Yes	MDR No	MDR Rate (%)	Median Age (Years)	Female (n)	Male (n)	Most Common Specimen	Blood (n, MDR Yes, MDR%)	Pus (n, MDR Yes, MDR%)	Sputum (n, MDR Yes, MDR%)	Urine (n, MDR Yes, MDR%)	Other (n, MDR Yes, MDR%)
Enterococcus	281	233	48	82.9	43	135	146	Pus	17, 13, 76.5	92, 78, 84.8	53, 44, 83.0	65, 55, 84.6	54, 43, 79.6
Streptococcus	65	50	15	76.9	49	30	35	Blood	7, 2, 28.6	18, 13, 72.2	11, 8, 72.7	9, 6, 66.7	6, 4, 66.7

ISSN: 0019-5138
DOI: https://doi.org/

MRSA	14	14	0	100.0	60	6	8	Blood	0, 0, —	6, 6, 100.0	0, 0, —	2, 2, 100.0	1, 1, 100.0
All GP isolates	360	297	63	82.5	—	—	—	—	24, 15, 62.5	116, 97, 83.6	64, 52, 81.3	76, 63, 82.9	61, 48, 78.7

MRSA: Methicillin-resistant *Staphylococcus aureus*

Specimen-Specific Multidrug Resistance Patterns

Specimen-specific multidrug resistance (MDR) patterns among Gram-positive (GP) bacterial isolates from clinical specimens are presented in Table 4. Table 4 summarises both the overall MDR prevalence by specimen type and the detailed MDR rates for Enterococcus, Streptococcus, and methicillin-resistant Staphylococcus aureus (MRSA) within each specimen category. This granular analysis provides clinically relevant insights for guiding empiric therapy and antimicrobial stewardship.

GP isolates: Gram-positive pathogens, including *Enterococcus, Streptococcus*, and methicillin-resistant *Staphylococcus aureus* (MRSA); excludes "no growth," non–Gram-positive isolates, and missing entries.MDR: Multidrug resistance, defined as resistance to ≥ 3 antimicrobial classes. Others: Includes endotracheal tube aspirates, stool, and pleural fluid specimens. n: Number of isolates within each organism–specimen category. MDR Yes: Number of MDR cases within each subgroup. MDR (%): Percentage of MDR cases within each subgroup, calculated as (MDR Yes / n) × 100; values are rounded to one decimal place. MRSA: No isolates from blood, sputum, or urine were classified as non-MDR. The apparent variation in MRSA multidrug resistance across specimen-specific subgroups is attributable to small sample sizes in certain categories, particularly blood specimens.

Table 4.Specimen-Specific Multidrug Resistance Patterns: Overall and by Organism

Specimen Type	All GP Isolates (n, MDR Yes, MDR%)	Enterococcus (n, MDR Yes, MDR%)	Streptococcus (n, MDR Yes, MDR%)	MRSA (n, MDR Yes, MDR%)
Blood	34, 24, 70.6	19, 14, 73.7	13, 10, 76.9	2, 0, 0.0
Pus	125, 109, 87.2	82, 71, 86.6	32, 28, 87.5	11, 10, 90.9
Sputum	103, 91, 88.3	73, 64, 87.7	24, 21, 87.5	6, 6, 100.0
Urine	95, 80, 84.2	77, 62, 80.5	12, 12, 100.0	6, 6, 100.0
Others	60, 51, 85.0	46, 39, 84.8	8, 7, 87.5	6, 5, 83.3

Age- and Gender-Stratified MDR Rates

Multidrug resistance (MDR) prevalence across age and gender subgroups is summarised in Table 5, highlighting the distribution of MDR among culture-positive Gram-positive (GP) pathogens within the study cohort.

GP Isolates: Culture-positive Gram-positive pathogens only (*Enterococcus, Streptococcus*, MRSA; excludes "No growth"). MDR: Multidrug resistance, defined as resistance to ≥ 3 antibiotic classes. p values: Calculated by chi-square or Fisher's exact test as appropriate, comparing each subgroup to the reference (20–39 for age, male for gender). Age groups: < 20 years (paediatric/ non-adult), 20–39 (young adults), 40–59 (middle-aged), and ≥ 60 (elderly). Gender: Minor data anomalies (e.g., "M") were standardised for analysis. No significant differences were detected by age or gender in MDR prevalence.

MDR is highly prevalent across all age groups and both genders, with rates consistently above 80%. None of the age categories or gender groups showed statistically significant differences in MDR prevalence, suggesting that demographic factors do not substantially modulate the risk of MDR in this hospitalised population. This uniform risk profile underscores the need for universal vigilance—all patients with suspected GP infections

ISSN: 0019-5138
DOI: https://doi.org/

in this setting should be considered at high risk for MDR, regardless of age or gender. These findings support broadly targeted antimicrobial stewardship and infection control interventions rather than demographic subgroup-specific strategies.

Table 5.Age- and Gender-Stratified MDR Rates Among Gram-Positive Pathogens

Category	Subgroup	GP Isolates (n)	MDR Yes (n)	MDR Rate (%)	p Value (χ^2)	Notes
All GP	—	393	328	83.5	—	Excludes "No growth"
Age group	< 20	20	16	80.0	0.82	vs reference (20–39)
	20–39	122	103	84.4	Ref	–
	40–59	89	73	82.0	0.61	–
	≥ 60	162	136	84.0	0.91	–
Gender	Female	192	159	82.8	0.67	vs male
	Male	201	169	84.1	Ref.	–
Age × gender interaction	Female, < 20	7	5	71.4	0.63	Interaction not significant
	Male, < 20	13	11	84.6	–	–
	Female, 20–39	60	49	81.7	–	–
	Male, 20–39	62	54	87.1	–	–
	Female, 40–59	42	34	81.0	–	–
	Male, 40–59	47	39	83.0	–	–
	Female, ≥ 60	83	71	85.5	–	–
	Male, ≥ 60	79	65	82.3	–	–

Machine Learning Input Features and Data Preprocessing

A detailed summary of variables selected for supervised ML, including feature engineering, encoding approaches, and completeness, is shown in Table 6. This level of detail supports reproducibility and rigorous model development for MDR detection among GP pathogens.

All 458 clinical specimens, including culture-negative cases, were retained for machine learning analysis to improve model generalisability and capture clinically relevant patterns associated with multidrug resistance.

MDR: Multidrug resistance, defined as resistance to three or more classes of antibiotics, based on local laboratory criteria. Organism Identified: "No growth" cases are included as a distinct category to support comprehensive modelling, but may be excluded for culture-positive–only analyses. Specimen: Other: Includes endotracheal tube (ET), stool, and pleural fluid specimens; all non-zero values were standardised to 1 to indicate the presence of any non-core specimen type. Gender: Minor inconsistencies (e.g., "M") were corrected to ensure uniform binary encoding (F = 0, M = 1). Age: Used as a continuous variable; centring or normalisation may be applied during model training, but was not performed in preprocessing. Missing data: No missing values were detected in any of the included features; all records are complete. One-hot encoding: Applied to organism categories to avoid implied ordinal relationships and support flexible modelling. Target variable: MDR status (Yes/ No) is directly encoded as a binary outcome for supervised classification. Data source: All features and preprocessing steps are derived from the hospital's electronic records and laboratory information system, ensuring clinical relevance and traceability.

The dataset was structured and preprocessed to support supervised machine learning for MDR prediction. Age and gender were included as demographic variables, while specimen type (blood, pus, sputum, urine, and others) captured the clinical source of infection. Organism identification (*Enterococcus, Streptococcus*, MRSA, and "no growth") was encoded using one-hot encoding to allow independent assessment of each category. Gender and specimen variables were binary encoded, and data entry inconsistencies were standardised to ensure data quality. No missing data were identified in the selected features, eliminating the need for imputation.

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202618

Performance Metrics of Supervised Machine Learning Models for MDR Detection

This section compares leading supervised ML models designed to predict MDR among GP pathogens, using a hospital-derived dataset (Table 7). Models evaluated include Logistic Regression, Random Forest, XGBoost, and an MLP neural network. Key metrics for clinical prediction included accuracy, precision, recall, F1-score, AUC-ROC, and interpretability.

Accuracy: Fraction of correct predictions on the test set. Precision: Fraction of predicted MDR-positive that are truly MDR. Recall: Fraction of true MDR that were correctly predicted. F1-score: Harmonic mean of precision and recall, balancing both. AUC-ROC: Area under the Receiver Operating Characteristic curve (discrimination for MDR). Interpretability: High = model provides clear feature importance; Low = less transparent/ black box.

Table 6.Machine Learning Input Features and Data Preprocessing

Feature Name	Type / Nature	Summary / Distribution	Encoding / Preprocessing	Missing Data n (%)
Age (years)	Continuous (numeric)	N = 458; Mean = 46.0; SD = 22.5; Range: 2–87	Used as continuous variable; optional normalisation/ centering	0 (0)
Gender	Categorical (binary)	Female: 231 (50.4%); Male: 227 (49.6%)	Binary encoding (Female = 0, Male = 1)	0 (0)
Specimen: Blood	Binary (0/1)	47 specimens (10.3%)	Used as-is (0 = absent, 1 = present)	0 (0)
Specimen: Pus	Binary (0/1)	109 specimens (23.8%)	Used as-is	0 (0)
Specimen: Sputum	Binary (0/1)	84 specimens (18.3%)	Used as-is	0 (0)
Specimen: Urine	Binary (0/1)	106 specimens (23.1%)	Used as-is	0 (0)
Specimen: Others (endotracheal tube aspirate, stool, pleural fluid)	Binary (0/1)	76 specimens (16.6%)	All non-zero values binarised to 1	0 (0)
Organism: Enterococcus	One-hot encoded	307 (67.0% of culture-positive isolates)	1 = Enterococcus; 0 = otherwise	0 (0)
Organism: Streptococcus	One-hot encoded	71 (15.5% of culture-positive isolates)	1 = Streptococcus; 0 = otherwise	0 (0)
Organism: MRSA	One-hot encoded	15 (3.3% of culture-positive isolates)	1 = MRSA; 0 = otherwise	0 (0)
Organism: No Growth	One-hot encoded	65 (14.2% of total specimens)	1 = No growth; 0 = otherwise	0 (0)
MDR Status (Target Variable)	Binary classification	Yes: 297 (64.8%); No: 161 (35.2%)	MDR = 1; Non-MDR = 0	0 (0)

MRSA: Methicillin-resistant Staphylococcus aureus

Table 7.Performance Metrics of Supervised Machine Learning Models for MDR Detection

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Interpretability
Logistic Regression	0.56	0.84	0.57	0.68	0.51	Moderate
Random Forest	0.69	0.85	0.76	0.80	0.58	High (feature importance)
XGBoost	0.66	0.86	0.70	0.77	0.59	High (feature importance)
MLP (Neural Net)	0.83	0.83	1.00	0.91	0.46	Low (black box)

Accuracy: Fraction of correct predictions on the test set. Precision: Fraction of predicted MDR-positive that are truly MDR.

Key Predictors of Multidrug Resistance Identified using Machine Learning

Table 8 summarises the most influential predictors of MDR among GP pathogens, as determined by a Random Forest classifier—one of the best-performing supervised ML models on this dataset. Each feature's importance reflects its relative impact on correctly classifying MDR cases. The table also includes clinical interpretations to guide application in practice.

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202618

Feature importance values are normalised and reflect relative contributions to MDR prediction in the Random Forest. All predictors above the threshold of 0.02 importance are shown. "No growth" as a predictive feature may indicate the complicating effects of prior antimicrobial exposure or fastidious organisms. Age and organism type are consistently leading drivers of MDR risk. Blood and atypical specimen types provide important context for model stratification and should be considered in clinical risk assessment.

The analysis shows that older age and organism type (especially *Enterococcus* and *Streptococcus*) are the strongest independent predictors of MDR among hospitalised patients with GP infections. Inclusion of "No growth" as a predictor highlights the importance of culture-negative results—often associated with prior antibiotic exposure or complex infections—as a signal for MDR risk. Male gender and blood or atypical specimen sites contribute moderately to MDR probability, underscoring the need for focused infection control in these settings. Empirical antimicrobial strategies and stewardship interventions should pay particular attention to older patients, those with *Enterococcus* or *Streptococcus* infections, culture-negative scenarios, and severe or invasive infection presentations. These features inform model-driven approaches to MDR prediction and can be adapted for clinical scoring and decision-support tools in similar healthcare environments.

Table 8.Key Predictors of Multidrug Resistance Identified using Machine Learning Models

Rank	Model Feature	Feature Importance	Clinical Interpretation
1	Age (years)	0.35	Older age
2	Organism: No growth	0.32	No organism cultured (often reflects prior antibiotics or atypical infection)
3	Organism: Enterococcus	0.13	*Enterococcus* isolate
4	Organism: Streptococcus	0.05	*Streptococcus* isolate
5	Male gender	0.03	Male patients
6	Blood specimen	0.02	Blood as an infection source
7	Other specimen (ET tube, stool, pleural)	0.02	Specimens from the endotracheal tube, stool, or pleural fluid
8	Organism: MRSA	0.02	MRSA isolate

ET: Endotracheal tube, MRSA: Methicillin-resistant Staphylococcus aureus

Discussion

This hospital-based cross-sectional study examined the burden of multidrug resistance (MDR), its major predictors, and the performance of supervised machine learning models among Gram-positive pathogens in a tertiary care setting. The findings indicate a substantial resistance burden, with *Enterococcus* as the predominant isolate, followed by *Streptococcusand* methicillin-resistant *Staphylococcus aureus* (MRSA). MDR was highly prevalent across all three organism groups, with consistently elevated rates observed irrespective of age, gender, or specimen source.[13],[14] These observations suggest a broad resistance pressure within the hospital environment and reinforce the need for strengthened antimicrobial stewardship and infection prevention measures. The predominance of *Enterococcus* across blood, pus, sputum, urine, and other specimen categories further indicates that resistance is not confined to a single clinical niche. *Streptococcus* also showed substantial resistance, while MRSA, though less common, remained uniformly multidrug resistant in this cohort. The absence of statistically meaningful demographic variation in MDR suggests that the risk is widely distributed across the study population rather than concentrated within specific patient subgroups. This differs from some published reports in which age, prior hospital exposure, or invasive-device use have shown clearer associations with resistance.[16],[17]

The machine learning analysis identified age, organism type, and culture negativity ("No growth") as the most influential predictors of MDR. The importance of "No growth" as a predictive feature may reflect prior antibiotic exposure, low organism burden, or the presence of fastidious and non-culturable pathogens, making it a potentially useful indicator in routine clinical assessment. Bloodstream and other invasive or atypical specimen sources also appeared relevant to MDR prediction, supporting the need for careful empirical treatment strategies in these settings.[18],[19] These findings highlight the value of integrating demographic, microbiological, and specimen-related variables when developing predictive tools for antimicrobial resistance.

Among the evaluated models, the multilayer perceptron showed the highest sensitivity for detecting MDR, whereas random forest and XGBoost provided a more balanced combination of predictive performance and interpretability.[21],[22] Logistic regression, although easier to interpret, showed lower discriminatory ability, suggesting that MDR prediction in this setting is influenced by complex, potentially non-linear relationships among the available variables.[23],[24] From a practical perspective, the tree-based

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202618

models may offer greater utility for clinical translation because they combine acceptable predictive accuracy with clearer feature importance patterns.

Unlike many centres where *Staphylococcus aureus* or Gram-negative organisms are more frequently dominant, this dataset was characterised by *Enterococcus* predominance, and the associated MDR level appeared higher than that reported in several comparable studies.[25],[26] Likewise, the resistance burden observed for *Streptococcus* exceeded many published estimates from similar hospital settings. MRSA prevalence was relatively low in the present study, but all identified isolates were multidrug resistant. In contrast to reports describing age- or gender-related differences in resistance risk, the current analysis did not demonstrate meaningful demographic variation, supporting the view that broad rather than subgroup-specific precautions may be more appropriate in this hospital context.[27],[28] The predictive contribution of culture-negative cases is also noteworthy, as this factor is not commonly emphasised in comparable reports but may have practical relevance for early risk assessment and antimicrobial stewardship.

These findings have several implications for practice. In this setting, Gram-positive infections from blood, urine, pus, sputum, and other invasive sources should be approached with awareness of the high background prevalence of MDR. Antimicrobial stewardship measures should therefore be applied consistently across clinical units, with particular attention to culture-negative cases and invasive or device-associated infections. Regular updating of local antibiograms and predictive models may further strengthen empirical treatment decisions and stewardship programmes. In parallel, infection prevention measures such as hand hygiene, environmental cleaning, and catheter-care protocols remain essential to reduce the transmission of resistant organisms within the hospital.

This study has several limitations. Its single-centre design may limit the generalisability of the findings to other institutions and regions. In addition, molecular characterisation and detailed patient-level risk factors, including prior antibiotic exposure and comorbidities, were not available, which may have further refined the analysis of resistance determinants. The number of MRSA isolates was also relatively small, and this should be considered when interpreting organism-specific results. Despite these limitations, the dataset was complete, clinically relevant, and suitable for both resistance profiling and predictive modelling, supporting the usefulness of the findings for local decision-making.

Conclusion

This study highlights a high burden of multidrug resistance (MDR) among Gram-positive pathogens in a tertiary care hospital setting. *Enterococcus* was the predominant organism and contributed substantially to the observed resistance patterns. No significant variation in MDR was identified across demographic subgroups, indicating that the risk is broadly distributed within the hospital population. These findings emphasise the need for strengthened antimicrobial stewardship practices, appropriate empiric therapy guided by local resistance patterns, and reinforced infection prevention measures. In addition, machine learning–based predictive models, particularly tree-based approaches, demonstrated potential as supportive tools for identifying MDR risk and assisting clinical decision-making.Overall, the results underscore the importance of continuous local surveillance and the development of context-specific, data-driven strategies to effectively manage and mitigate antimicrobial resistance in hospital settings.

Authors' Contribution: Umaarasu Thirunavukkarasu and Thirunavukkarasu Dharmalingam contributed to the study conception and design. Suprraja Subramani and Senthil Kumar S were involved in data collection and data curation. Loganathan Thangavel performed the data analysis and interpretation. Arbind Kumar Choudhary supervised the study, contributed to the interpretation of results, and drafted the manuscript. All authors critically reviewed the manuscript, approved the final version, and agree to be accountable for all aspects of the work.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process:

No generative AI or AI-assisted technologies were used in the writing, analysis, or preparation of this manuscript. (If you want to declare use, I can rewrite this accordingly.)

Source of Funding: This study did not receive any specific funding from public, commercial, or not-for-profit funding agencies.

Conflict of Interest: The authors declare that there are no conflicts of interest related to this study.

References

1. Bhatt P, Patel A, Sahni AK, Praharaj AK, Grover N, Chaudhari CN, Das NK, Kulkarni M. Emergence of multidrug resistant enterococci at a tertiary care centre. Med J Armed Forces India. 2015 Apr;71(2):139-44. [PubMed] [Google Scholar]
2. Sangwan J, Lohan K, Mane P, Kumar M. Emerging crisis of multidrug-resistant enterococci from a rural tertiary care hospital of North India. Indian J Health Sci Biomed Res (KLEU). 2021;14(2):260-4. [Google Scholar]
3. Taneja N, Sharma M. Antimicrobial resistance in the environment: the Indian scenario. Indian J Med Res. 2019;149(2):119-28. [PubMed] [Google Scholar]

4. Shah S, Rampal R, Thakkar P, Poojary S, Ladi S. The prevalence and antimicrobial susceptibility pattern of Gram-positive pathogens: three-year study at a tertiary care hospital in Mumbai, India. J Lab Physicians. 2022;14(2):109-14. [PubMed] [Google Scholar]
5. Mamtora D, Saseedharan S, Bhalekar P, Katakdhond S. Microbiological profile and antibiotic susceptibility pattern of Gram-positive isolates at a tertiary care hospital. J Lab Physicians. 2019;11(2):144-8. [PubMed] [Google Scholar]
6. Mendem SK, Gangadhara TA, Shivannavar CT, Gaddad SM. Antibiotic resistance patterns of *Staphylococcus aureus*: a multi center study from India. Microb Pathog. 2016 Sep;98:167-70. [PubMed] [Google Scholar]
7. Patil SS, Suresh KP, Shinduja R, Amachawadi RG, Chandrashekhar S, Pradeep S, Kollur SP, Syed A, Sood R, Roy P, Shivamallu C. Prevalence of methicillin-resistant *Staphylococcus aureus* in India: a systematic review and meta-analysis. Oman Med J. 2022 Jul 31;37(4):e440. [PubMed] [Google Scholar]
8. Rajkumar S, Sistla S, Manoharan M, Sugumar M, Nagasundaram N, Parija SC, Ray P, Bakthavatchalam YD, Veeraraghavan B, Kapil A, Walia K, Ohri VC. Prevalence and genetic mechanisms of antimicrobial resistance in *Staphylococcus* species: a multicentre report of the Indian Council of Medical Research Antimicrobial Resistance Surveillance Network. Indian J Med Microbiol. 2017;35(1):53-60. [PubMed] [Google Scholar]
9. Pai V, Rao VI, Rao SP. Prevalence and antimicrobial susceptibility pattern of methicillin-resistant *Staphylococcus aureus* [MRSA] isolates at a tertiary care hospital in Mangalore, South India. J Lab Physicians. 2010;2(2):82-4. [PubMed] [Google Scholar]
10. Khurana S, Bhardwaj N, Kumari M, Malhotra R, Mathur P. Prevalence, etiology, and antibiotic resistance profiles of bacterial bloodstream infections in a tertiary care hospital in Northern India: a 4-year study. J Lab Physicians. 2018;10(4):426-31. [PubMed] [Google Scholar]
11. Kannan R, Anil A, Thomas P, Raju NS, Reji SM. Antibiotic susceptibility profiling of Gram-positive and Gram-negative bacterial isolates in a tertiary care hospital: establishment of an antibiogram. Cureus. 2024 May 18;16(5):e60542. [PubMed] [Google Scholar]
12. Bhardwaj N, Mathur P, Behera B, Mathur K, Kapil A, Misra MC. Antimicrobial resistance in beta-haemolytic streptococci in India: a four-year study. Indian J Med Res. 2018;147(1):81-7. [PubMed] [Google Scholar]
13. Al-Zoubi MS, Al-Tayyar IA, Hussein E, Jabali AA, Khudairat S. Antimicrobial susceptibility pattern of Staphylococcus aureus isolated from clinical specimens in Northern area of Jordan. Iran J Microbiol. 2015 Oct;7(5):265-72. PMID: 26719783; PMCID: PMC4695508.
14. Gupta R, Malik A, Rizvi M, Ahmed M, Hashmi A. Multidrug resistant Gram positive pathogens with special reference to MRSA and biofilm production in ICU patients: recurrent challenge for clinicians. Int J Curr Microbiol App Sci. 2015; Special Issue 1:207–212.
15. Indian Network for Surveillance of Antimicrobial Resistance (INSAR) group, India. Methicillin resistant *Staphylococcus aureus* (MRSA) in India: prevalence & susceptibility pattern. Indian J Med Res. 2013 Feb;137(2):363-9. [PubMed] [Google Scholar]
16. Saxena D, Gwalani R, Yadav A, Shah R. Growing Concerns on Antimicrobial Resistance - Past, Present, and Future Trends. Indian J Community Med. 2025 Jan-Feb;50(1):4-8. doi: 10.4103/ijcm.ijcm_838_23. Epub 2025 Jan 23. PMID: 40124794; PMCID: PMC11927827.
17. Walia K, Madhumathi J, Veeraraghavan B, Chakrabarti A, Kapil A, Ray P, Singh H, Sistla S, Ohri VC. Establishing antimicrobial resistance surveillance & research network in India: journey so far. Indian J Med Res. 2019 Feb;149(2):164-79. [PubMed] [Google Scholar]
18. Kaur J, Singh H, Sethi T. Emerging trends in antimicrobial resistance in bloodstream infections: multicentric longitudinal study in India (2017-2022). Lancet Reg Health Southeast Asia. 2024 May 9;26:100412. doi: 10.1016/j.lansea.2024.100412. PMID: 38757091; PMCID: PMC11097075.
19. Farhat F, Athar MT, Ahmad S, Madsen DØ, Sohail SS. Antimicrobial resistance and machine learning: past, present, and future. Front Microbiol. 2023;14:1179312. [PubMed] [Google Scholar]
20. Sakagianni A, Koufopoulou C, Feretzakis G, Kalles D, Verykios VS, Myrianthefs P, Fildisis G. Using machine learning to predict antimicrobial resistance - a literature review. Antibiotics (Basel). 2023;12(3):452. [PubMed] [Google Scholar]
21. Ardila CM, Gonzalez-Arroyave D, Tobon S. Machine learning for predicting antimicrobial resistance in critical and high-priority pathogens: a systematic review considering antimicrobial susceptibility tests in real-world healthcare settings. PLoS One. 2025;20(2):e0319460. [PubMed] [Google Scholar]
22. Gandra S, Mojica N, Klein EY, Ashok A, Nerurkar V, Kumari M, Ramesh U, Dey S, Vadwai V, Das BR, Laxminarayan R. Trends in antibiotic resistance among major bacterial pathogens isolated from blood cultures tested at a large private laboratory network in India, 2008–2014. Int J Infect Dis. 2016;50:75-82. [PubMed] [Google Scholar]
23. Kaur J, Kaur J, Dhama AS, Jindal S, Walia K, Singh H. Strengthening the surveillance of antimicrobial resistance in India using integrative technologies. Front Public Health. 2022;10:861888. [PubMed] [Google Scholar]
24. Christina S, Praveena R, Saikumar C. Emerging resistance mechanisms in Gram-positive bacteria isolated from

septicemia cases in ICUs: a focus on genotypic insights. Cureus. 2025;17(1):e76979. [PubMed] [Google Scholar]
25. Kim JS, Jung BK, Kim JW, Kim GY. Prevalence and Antimicrobial Susceptibility of *Streptococcus pneumoniae* Isolated from Clinical Samples in the Past 8 Years in Korea. Biomed Res Int. 2021 Apr 28;2021:6615334. doi: 10.1155/2021/6615334. PMID: 33997025; PMCID: PMC8099532.
26. Golli, Andreea & Cristea, Oana & Zlatian, Ovidiu & Adina Dorina, Glodeanu & Balasoiu, Andrei & Ionescu, Mihaela & Popa, Simona. (2022). Prevalence of Multidrug-Resistant Pathogens Causing Bloodstream Infections in an Intensive Care Unit. Infection and Drug Resistance. 15. 5981-5992. 10.2147/IDR.S383285.