

Machine Learning Classification of High Antibiotic Consumption Using the Drug Resistance Index: A Comparison of Random Forest, Support Vector Machine and Logistic Regression in Indian Tertiary-Care Hospitals

Gayatri Prashant Sapkale¹, Dr. Prashant Kumar Dhakad^{2*}, Dr. Anita Arora³, Dr. Bishnu Panigrahi⁴, Sachin Gulia⁵

¹PhD Scholar, Gyan Vihar School of Pharmacy, Suresh Gyan Vihar University, Jaipur, Rajasthan 302017 Email Id: Sapkalegayatri64@gmail.com

^{2*} Professor, Gyan Vihar School of Pharmacy, Suresh Gyan Vihar University, Jaipur, Rajasthan 302017 Email Id: prashantkr.dhakad@mygyanvihar.com

³Group Head-Infection Prevention & Control, Director Medical Operation-MSOG, Fortis Healthcare Ltd., Gurgaon, Haryana, India-122002 Email Id: anita.arora@fortishealthcare.com

⁴Group Head - Medical Strategy & Operations Group, Fortis Healthcare Ltd., Gurgaon, Haryana, India-122002 Email Id: bishnu.panigrahi@fortishealthcare.com

⁵School of Pharmacy, Plot No. 32, 34, Knowledge Park III, Greater Noida, Uttar Pradesh 201310, India Email Id: guliasachin47@gmail.com

DOI: [https://doi.org/10.63001/tbs.2026.v21.i04.S.I\(4\).pp475-492](https://doi.org/10.63001/tbs.2026.v21.i04.S.I(4).pp475-492)

KEYWORDS

*Antimicrobial stewardship;
Antimicrobial resistance;
Drug Resistance Index;
Antibiotic consumption;
Machine learning;
Surveillance; India*

Received: 20/09 /2026

Accepted :07/10/2026

Published: 17/10/2026

Abstract

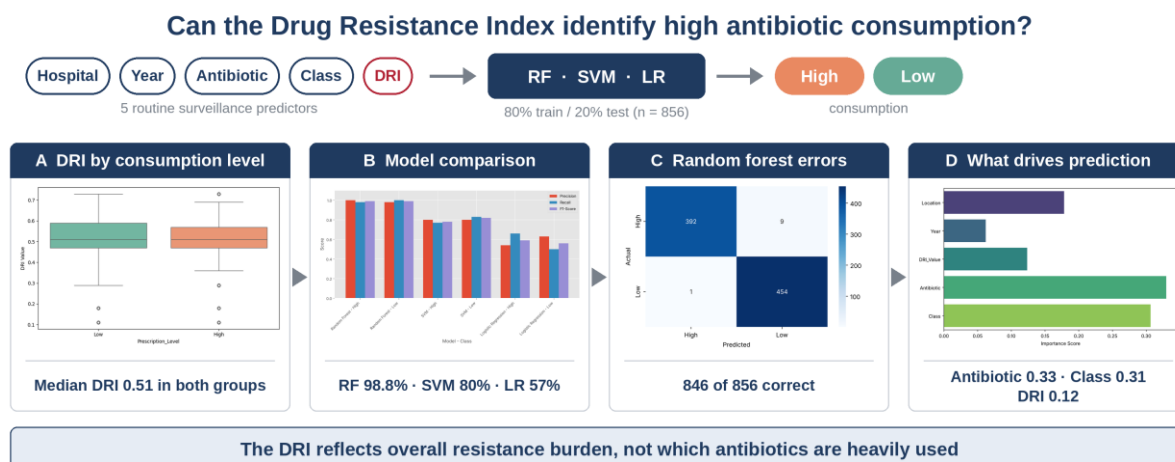
Background: Hospitals record antibiotic consumption and antimicrobial resistance but rarely analyse them together. The Drug Resistance Index (DRI) combines both by weighting resistance to each antibiotic by that drug's share of total use. This study compared three machine learning classifiers for identifying high antibiotic consumption from the DRI and routine surveillance variables.

Methods: Antibiotic consumption records with corresponding DRI values from tertiary-care hospitals in India were classified as high or low consumption. Random forest (RF), support vector machine (SVM) and logistic regression (LR) models were trained on five predictors: hospital, year, antibiotic, antibiotic class and DRI. Performance was evaluated on a hold-out test set of 856 records (401 high and 455 low consumption) using accuracy, precision, recall, F1-score and confusion matrices. The contribution of each predictor was estimated using RF impurity-based feature importance.

Results: RF correctly classified 846 of the 856 test records (accuracy 98.8%; macro-averaged F1-score 0.99). SVM and LR achieved accuracies of 80% and 57%, with macro-averaged F1-scores of 0.80 and 0.57, respectively. In the RF model, antibiotic (importance 0.33) and antibiotic class (0.31) were the strongest predictors, followed by hospital (0.18), DRI (0.12) and year (0.06). The median DRI was 0.51 in both the high- and low-consumption groups.

Conclusions: RF distinguished high from low antibiotic consumption considerably better than SVM and LR, but it relied mainly on the antibiotic and its class rather than on the DRI. Grouped and external validation are needed before such models are used to guide antimicrobial stewardship.

Graphical abstract



Graphical abstract. Five routine surveillance predictors (hospital, year, antibiotic, antibiotic class and DRI) were used to classify antibiotic consumption records as high or low with random forest (RF), support vector machine (SVM) and logistic regression (LR). Panels A–D reproduce Figures 1B, 2, 3A and 3B. RF achieved the highest accuracy, but its predictions relied mainly on the antibiotic and its class, and DRI values did not differ between consumption groups.

1. Introduction

Antimicrobial resistance (AMR) is one of the leading threats to global health. In 2021, an estimated 1.14 million deaths were directly attributable to bacterial AMR, and 4.71 million deaths were associated with it. Without further action, these figures are forecast to reach 1.91 million and 8.22

million deaths, respectively, in 2050 [1]. Surveillance data from 104 countries showed that about one in six laboratory-confirmed bacterial infections worldwide were resistant to antibiotic treatment in 2023, rising to one in three in the South-East Asia and Eastern Mediterranean regions [2]. Antibiotic consumption is a major driver of resistance, and higher use is associated with higher resistance [3]. Across 67 countries, antibiotic consumption increased by 16.3% between 2016 and 2023 to 49.3 billion defined daily doses (DDDs), and it could reach 75.1 billion DDDs by 2030 without further intervention [4]. India alone consumed 5,071 million DDDs of systemic antibiotics in 2019, and Watch group antibiotics accounted for 54.9% of this total [5]. Hospitals therefore need practical ways to

identify where antibiotic use is concentrated.

Antimicrobial stewardship programmes improve antibiotic prescribing for hospital inpatients [6]. These programmes rely on two data sources that are usually reported separately: the cumulative antibiogram from the microbiology laboratory and antibiotic consumption data from the pharmacy. The Drug Resistance Index (DRI) combines these into a single measure by weighting resistance to each antibiotic by that drug's share of total use, giving a value between 0 and 1 [7]. The DRI has been used to track the effectiveness of antibiotic therapy across countries and over time [8]. In a multicentre analysis of 20 Indian hospitals, the DRI correlated with AMR and declined after stewardship implementation [9]. Because the DRI already incorporates use weights, its relationship with the intensity of antibiotic prescribing is of direct interest to stewardship teams.

Machine learning (ML) is increasingly applied to AMR, mainly to predict resistance in individual isolates or patients from clinical, microbiological or genomic data [10,11]. Supervised algorithms such as logistic regression (LR), support vector machine (SVM) [12] and random forest (RF) [13] are among the most widely used [10]. Reported performance is often high,

but most models have been developed retrospectively, data processing is rarely standardised, and few models have been validated in real-world settings [11]. Moreover, ML models do not consistently outperform LR in clinical prediction [10,14]. Their performance can also be overestimated when closely related records appear in both the training and test sets, a problem known as data leakage [15].

By contrast, ML has rarely been applied to aggregated stewardship data, such as hospital-level antibiotic consumption and the DRI. We are not aware of any study that has tested whether the DRI, combined with routine surveillance variables, can identify antibiotics with high consumption. It is also unclear how much the DRI adds beyond knowing which antibiotic was used and at which hospital. Answering these questions would show whether the DRI can support targeted stewardship review, or whether consumption is driven mainly by other factors.

This study therefore compared RF, SVM and LR for classifying antibiotic consumption records from Indian tertiary-care hospitals as high or low consumption, using hospital, year, antibiotic, antibiotic class and DRI as predictors. We also examined the contribution of the DRI relative to the other predictors.

2. Materials and Methods

2.1. Study design

This was a retrospective, multicentre analysis of aggregated antimicrobial stewardship surveillance data from tertiary-care hospitals in India. Three supervised machine learning classifiers were compared for their ability to classify antibiotic consumption records as high or low consumption. Reporting was guided by the TRIPOD+AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + Artificial Intelligence) statement [16].

2.2. Data source and variables

Data were drawn from routine antimicrobial stewardship surveillance in the participating hospitals. This surveillance combines antibiotic consumption data from the pharmacy with antibiogram data from the microbiology laboratory. Each record contained the hospital, year, antibiotic, antibiotic class, antibiotic consumption and the corresponding DRI value. Only aggregated data were used, and no patient-level or identifiable information was accessed.

2.3. Calculation of the Drug Resistance Index

The DRI was calculated using the method described by Laxminarayan and Klugman [7]. For each antibiotic k , the quantity

dispensed was first converted to DDDs using the WHO DDD [17]:

$$DDD_k = \frac{A_k}{D_k} \quad (1)$$

where A_k is the total quantity of antibiotic k dispensed (g) and D_k is the WHO DDD assigned to that antibiotic (g). Consumption was then expressed per 100 bed-days:

$$U_k = \frac{DDD_k}{BD} \times 100 \quad (2)$$

$$BD = 365 \times N \times O \quad (3)$$

where U_k is the use of antibiotic k in DDDs per 100 bed-days, BD is the number of bed-days in the year, N is the number of hospital beds and O is the bed occupancy index.

The resistance rate and the use weight of each antibiotic were then calculated as follows:

$$\rho_k = \frac{R_k}{T_k} \quad (4)$$

$$q_k = \frac{U_k}{\sum_{j=1}^n U_j} \quad (5)$$

where R_k is the number of isolates resistant to antibiotic k , T_k is the number of isolates tested against it, and n is the number of antibiotics included in the index. The DRI

was the sum of the use-weighted resistance rates:

$$\text{DRI} = \sum_{k=1}^n \rho_k q_k \quad (6)$$

Because the use weights sum to 1, the DRI ranges from 0 to 1. A value of 0 indicates that all tested isolates were susceptible to the antibiotics in use, and a value of 1 indicates that they were resistant to all of them [7].

2.4. Outcome definition

The outcome was the consumption level of each record, a binary variable (high or low) derived from the recorded antibiotic consumption. High consumption was treated as the positive class.

2.5. Predictors and data preprocessing

Five predictors were used: hospital, year, antibiotic, antibiotic class and DRI. Antibiotic consumption itself was not included as a predictor because the outcome was derived from it. The categorical predictors (hospital, antibiotic and antibiotic class) were label-encoded as integers, so each predictor entered the models as a single feature. Year and DRI were entered as numerical variables.

2.6. Data partitioning

The dataset was divided into a training set (80%) and a hold-out test set (20%). The test set comprised 856 records, of which

401 were high consumption and 455 were low consumption. All models were fitted on the training set only, and the test set was used solely for performance evaluation.

2.7. Classification algorithms

Random forest (RF). RF is an ensemble of decision trees. Each tree is grown on a bootstrap sample of the training data, and a random subset of predictors is considered at each split. The final class is assigned by majority vote across trees, which allows RF to capture nonlinear relationships and interactions between predictors [13].

Support vector machine (SVM). SVM identifies the boundary that separates the two classes with the widest possible margin. Kernel functions allow this boundary to be nonlinear in the original predictor space [12].

Logistic regression (LR). LR models the log-odds of high consumption as a linear function of the predictors. It was included as an interpretable baseline because it is the standard comparator for machine learning models in clinical prediction [14].

All models were implemented in Python using scikit-learn [18]. Figures were produced with matplotlib [19] and seaborn [20].

2.8. Performance evaluation

Model performance on the test set was summarised in a confusion matrix of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The following metrics were calculated [21]:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (7)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

Recall is also known as sensitivity. Precision, recall and F1-score were calculated separately for each class, with that class treated as positive. Overall performance was summarised as the macro average (the unweighted mean of the two classes) and the weighted average (the mean weighted by the number of records in each class).

2.9. Feature importance

The contribution of each predictor to the RF model was estimated using impurity-based feature importance, defined as the mean decrease in Gini impurity across all trees and normalised to sum to 1. This measure can favour predictors with many distinct

values [22]. It was therefore interpreted as a relative ranking of predictors within the fitted model, not as evidence of a causal effect.

2.10. Descriptive analysis of the DRI

The distribution of DRI values was examined with a histogram and kernel density estimate. DRI values were compared between the high- and low-consumption groups using box plots. The relationship between DRI and antibiotic consumption was examined with a scatter plot and a fitted least-squares regression line.

2.11. Ethical considerations

The study used only aggregated, anonymised surveillance data and involved no patient contact or identifiable patient information.

3. Results

3.1. Distribution of DRI values and relationship with antibiotic consumption

DRI values ranged from 0.11 to 0.73 (Figure 1A). Most values lay between 0.40 and 0.65, with three small, separate clusters near 0.11, 0.18 and 0.29. The distribution of DRI values was similar in the high- and low-consumption groups (Figure 1B). The median DRI was 0.51 in both groups, and the interquartile range (IQR) was 0.47–0.59 in the low-consumption group and 0.47–

0.57 in the high-consumption group. DRI values of 0.11 and 0.18 appeared as outliers in both groups.

High-consumption records spanned a wide range of consumption values, up to approximately 40,000 DDDs, whereas low-consumption records were clustered at the lowest consumption values (Figure 1C).

Records from both groups occurred across the full DRI range. The fitted regression line was nearly flat, rising from approximately 2,000 DDDs at a DRI of 0.11 to approximately 2,700 DDDs at a DRI of 0.73. DRI alone was therefore a weak indicator of antibiotic consumption.

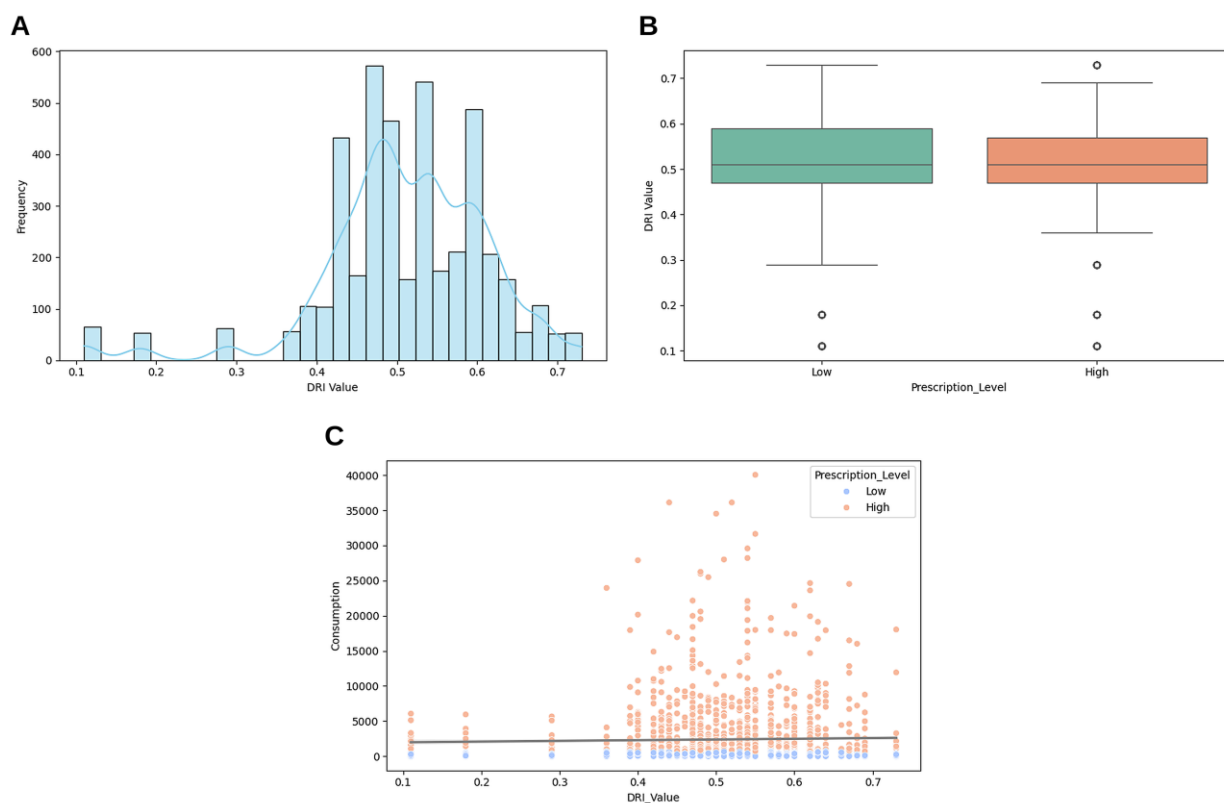


Figure 1. Distribution of Drug Resistance Index (DRI) values and their relationship with antibiotic consumption. (A) Histogram of DRI values with a kernel density estimate. (B) Box plots of DRI values by consumption level; boxes show the median and interquartile range, whiskers extend to 1.5 times the interquartile range, and circles indicate outliers. (C) Antibiotic consumption (DDDs) plotted against DRI and coloured by consumption level, with a fitted least-squares regression line (grey). In panels B and C, Prescription Level denotes the consumption level (high or low), and DRI Value denotes the DRI.

3.2. Comparative classification performance

The performance of the three classifiers on the hold-out test set of 856 records (401

high and 455 low consumption) is summarised in Table 1 and Figure 2. RF achieved the highest performance on every metric, with an accuracy of 0.99 and macro-averaged precision, recall and F1-score of 0.99. Its precision and recall ranged from 0.98 to 1.00 in both classes.

SVM achieved an accuracy of 0.80, with a precision of 0.80 in both classes. Its recall was lower for high-consumption records (0.77) than for low-consumption records (0.83), giving F1-scores of 0.78 and 0.82, respectively. LR performed least well, with an accuracy of 0.57 and a macro-averaged F1-score of 0.57. LR identified 66% of high-consumption records (recall 0.66) but only half of low-consumption records (recall 0.50), and its precision was 0.54 and 0.63 for the high and low classes, respectively.

Table 1. Classification performance of the three models on the hold-out test set (n = 856).

Model	Class / average	Precision	Recall	F1-score	Support
Random forest (accuracy 0.99)	High	1.00	0.98	0.99	401
	Low	0.98	1.00	0.99	455
	Macro average	0.99	0.99	0.99	856
	Weighted average	0.99	0.99	0.99	856
Support vector machine (accuracy 0.80)	High	0.80	0.77	0.78	401
	Low	0.80	0.83	0.82	455
	Macro average	0.80	0.80	0.80	856

	Weighted average	0.80	0.80	0.80	856
Logistic regression (accuracy 0.57)	High	0.54	0.66	0.59	401
	Low	0.63	0.50	0.56	455
	Macro average	0.58	0.58	0.57	856
	Weighted average	0.58	0.57	0.57	856

Precision, recall and F1-score are shown for each class, with that class treated as positive. Overall accuracy is given under each model name. Support is the number of test records in each class or, for the averages, in the whole test set. Values are rounded to two decimal places.

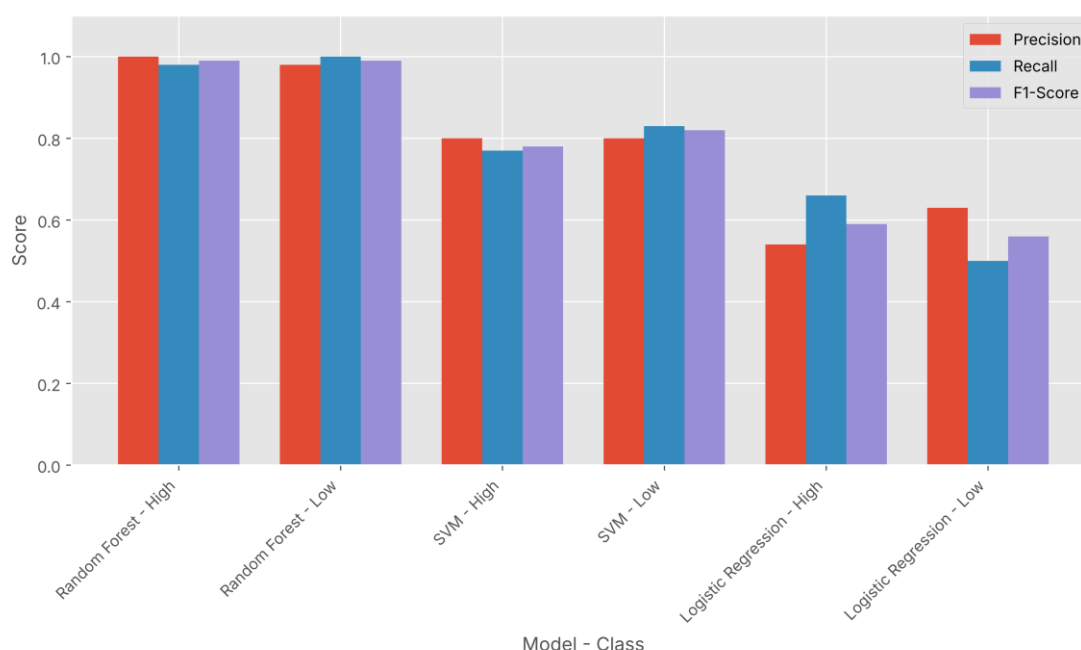


Figure 2. Class-specific precision, recall and F1-score of the random forest, support vector machine (SVM) and logistic regression models on the hold-out test set (n = 856).

3.3. Confusion matrix of the random forest model

The RF confusion matrix is shown in Figure 3A, and the measures derived from it are given in Table 2. Of the 401 high-consumption records, 392 were correctly classified and 9 were misclassified as low

consumption. Of the 455 low-consumption records, 454 were correctly classified and 1 was misclassified as high consumption. RF therefore correctly classified 846 of the 856 test records (accuracy 98.8%). Nine of the ten misclassifications were high-consumption records predicted as low, so

recall was slightly lower for the high-consumption class (97.8%) than for the low-consumption class (99.8%).

Table 2. Performance measures derived from the random forest confusion matrix on the hold-out test set (n = 856).

Measure	Formula	Calculation	Value
Recall, high consumption (sensitivity)	$TP / (TP + FN)$	392 / 401	0.978 (97.8%)
Recall, low consumption (specificity)	$TN / (TN + FP)$	454 / 455	0.998 (99.8%)
Precision, high consumption	$TP / (TP + FP)$	392 / 393	0.997 (99.7%)
Precision, low consumption	$TN / (TN + FN)$	454 / 463	0.981 (98.1%)
F1-score, high consumption	$2TP / (2TP + FP + FN)$	784 / 794	0.987
F1-score, low consumption	$2TN / (2TN + FN + FP)$	908 / 918	0.989
Macro-averaged F1-score	Mean of the two class F1-scores	$(0.987 + 0.989) / 2$	0.988
Accuracy	$(TP + TN) / N$	846 / 856	0.988 (98.8%)

High consumption was treated as the positive class. TP, true positives (392); FN, false negatives (9); FP, false positives (1); TN, true negatives (454); N, total test records (856). Values are given to three decimal places.

3.4. Contribution of predictors

The impurity-based feature importance of the RF model is shown in Figure 3B. Antibiotic (0.33) and antibiotic class (0.31) were the strongest predictors and together accounted for 0.64 of the total importance. Hospital ranked third (0.18), followed by DRI (0.12) and year (0.06).

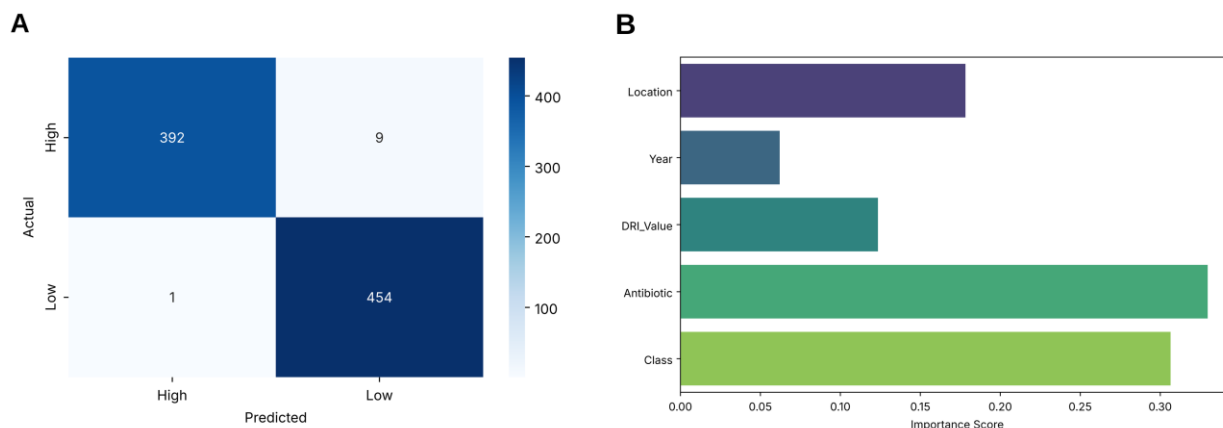


Figure 3. Performance and predictor contributions of the random forest model. (A) Confusion matrix on the hold-out test set ($n = 856$). (B) Impurity-based feature importance. Location denotes the hospital, Class denotes the antibiotic class, and DRI Value denotes the Drug Resistance Index.

4. Discussion

This study compared three machine learning classifiers for distinguishing high from low antibiotic consumption using the DRI and four routine surveillance variables. RF performed best, correctly classifying 98.8% of test records, compared with 80% for SVM and 57% for LR. However, RF relied mainly on antibiotic identity and antibiotic class, and the DRI made only a modest contribution. DRI values were also almost identical in the high- and low-consumption groups. Together, these findings indicate that antibiotic consumption in this dataset was explained largely by which antibiotic was used, rather than by the resistance burden summarised by the DRI.

The large gap between the three models is best explained by the structure of the predictors. Hospital, antibiotic and antibiotic class were label-encoded as integers. These codes have no natural order, but LR treats them as continuous values and fits a single linear slope across them. LR therefore cannot represent the separate consumption level of each antibiotic or hospital, which is consistent with its near-chance accuracy of 57%. In contrast, RF can isolate individual codes, or groups of codes, through successive splits. It can therefore learn antibiotic- and hospital-specific patterns from the same encoding [13]. SVM performed between these two extremes, and its performance depends on the choice of kernel and the scaling of predictors [12]. The comparison therefore reflects how each algorithm handles the

encoding as much as any inherent advantage of RF. This interpretation agrees with evidence that ML models do not consistently outperform LR in clinical prediction [14]. One-hot encoding represents each category as a separate binary variable. It would allow LR to estimate a separate effect for each antibiotic and hospital and would provide a fairer comparison [23].

The dominance of antibiotic and antibiotic class is clinically plausible. The amount of an antibiotic used in a hospital is shaped largely by its role in treatment, its position in the hospital formulary and local guidelines, and whether its use is restricted. In India, Watch group antibiotics accounted for 54.9% of national antibiotic consumption in 2019 [5]. The WHO AWaRe (Access, Watch, Reserve) classification was developed to monitor this pattern of use and to set stewardship targets [24]. Hospital ranked third in importance, which may reflect differences in size, case mix and prescribing policy between hospitals. Year contributed least, suggesting that it added little information once the antibiotic and hospital were known.

The weak contribution of the DRI follows from how the index is constructed. The DRI is a use-weighted average of resistance across a basket of antibiotics (Equation 6)

[7]. It therefore summarises the overall resistance burden of a setting and does not vary between antibiotics in the way that consumption does. This pattern was visible in the data. Many records shared the same DRI value (Figure 1C), the median DRI was 0.51 in both consumption groups (Figure 1B), and the regression line between DRI and consumption was almost flat (Figure 1C). Although the DRI contains use weights, these weights are proportions of total use rather than absolute quantities (Equation 5). A high DRI therefore does not imply that any particular antibiotic is heavily used. The DRI was designed as an aggregate indicator for communicating resistance trends over time and between locations [7,8]. A multicentre analysis of 20 Indian hospitals supports this role, showing that the DRI correlated with AMR and declined after stewardship implementation [9]. The present results indicate that this aggregate signal adds little when the task is to identify which individual antibiotics are heavily consumed.

The near-perfect accuracy of RF should be interpreted with caution. The test set was drawn from the same dataset as the training set and therefore contained the same hospitals and antibiotics. Records for the same antibiotic at the same hospital in different years are likely to have similar consumption. RF may therefore have

recognised combinations already seen during training rather than learned a pattern that generalises to new settings. This form of non-independence between training and test data is a recognised cause of over-optimistic performance [15]. Validation that keeps all records for a given hospital, antibiotic or year together, such as grouped or temporal cross-validation, gives a more realistic estimate [25,26]. Testing in hospitals outside the development dataset is also needed before any clinical or managerial use [16,25].

These findings have two practical implications for stewardship teams. First, the DRI should not be used on its own to flag antibiotics with high consumption. Antibiotic-specific measures, such as DDDs per 100 bed-days and AWaRe-based targets, remain more direct tools for this purpose [17,24]. Second, the DRI remains useful as a summary outcome for comparing hospitals and for tracking the effect of stewardship over time [8,9]. ML models intended to support stewardship should combine such indicators with richer predictors, such as previous consumption, AWaRe category and ward or specialty. They should also be developed and reported in line with the TRIPOD+AI statement [16].

5. Strengths and Limitations

This study has several strengths. It used routinely collected stewardship data from multiple tertiary-care hospitals, compared three widely used algorithms on the same hold-out test set, and reported class-specific metrics and a full confusion matrix rather than accuracy alone.

The study also has limitations. First, performance was assessed on a single hold-out split without cross-validation. Confidence intervals for the metrics could not therefore be estimated, and discrimination (area under the receiver operating characteristic curve) and calibration were not assessed. Second, the test set was not independent of the training data because it contained the same hospitals and antibiotics, and no external validation was performed. Third, label encoding of the categorical predictors disadvantaged LR and possibly SVM, so the comparison between algorithms may partly reflect predictor encoding. Fourth, impurity-based importance can overstate predictors with many distinct values [22]. Antibiotic and hospital have many categories, so their importance may be inflated. Fifth, dividing consumption into high and low levels discards information and makes the results dependent on the chosen threshold [27]. Finally, the data were aggregated rather than patient-level and came from tertiary-

care hospitals in India. The findings therefore cannot be applied to individual prescribing decisions and may not generalise to other settings.

6. Future Perspectives

Several steps would strengthen this work and show whether the DRI can support ML-based antimicrobial stewardship:

- **Validation.** Models should be validated with grouped or temporal cross-validation. In grouped validation, all records for a given hospital, antibiotic or year are kept together. In temporal validation, earlier years are used for training and later years for testing [25,26].
- **Encoding.** Categorical predictors should be one-hot encoded so that LR and SVM can be compared fairly with RF [23].
- **Reporting.** Performance should be reported with the area under the receiver operating characteristic curve, calibration and 95% confidence intervals, in line with the TRIPOD+AI statement [16].
- **The DRI.** Models with and without the DRI should be compared directly to measure the independent information it adds.
- **Feature importance.** Permutation importance [13] or Shapley additive

explanations [28] should be used, because they are less affected by the number of categories.

- **Outcome.** Consumption could be modelled as a continuous outcome, such as DDDs per 100 bed-days, rather than as a binary category [27].
- **Predictors.** Richer predictors, including AWaRe category, ward or specialty and previous consumption, should be tested [24].
- **External testing.** Models that perform well should be validated in other hospitals and then evaluated prospectively within routine stewardship review [25].

7. Conclusion

In this multicentre analysis of routine stewardship data from Indian tertiary-care hospitals, RF distinguished high from low antibiotic consumption considerably better than SVM and LR. Its predictions depended mainly on the antibiotic and its class. The DRI contributed only modest information and did not differ between high- and low-consumption records. The DRI is therefore better suited to tracking the overall resistance burden of a hospital than to identifying individual antibiotics with high consumption. The high accuracy of RF was obtained by testing on records closely related to the training data. It needs

confirmation with grouped validation and external data before such models are used to guide antimicrobial stewardship.

Author Contributions

Gayatri Prashant Sapkale: Conceptualization, methodology, data curation, formal analysis, and writing – original draft. Prashant Kumar Dhakad: Supervision, conceptualization, methodology, and critical revision of the manuscript. Anita Arora: Resources, investigation, and validation. Bishnu Panigrahi: Supervision, project administration, and resources. Sachin Gulia: Data curation, validation, interpretation of results, and writing – review and editing.

Declarations

Ethics approval and consent to participate: Ethics approval was not required for this study because it used only aggregated, anonymised antimicrobial stewardship surveillance data and involved no patient contact or identifiable patient information. For the same reason, informed consent was not applicable.

Consent for publication: Not applicable.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Competing interests: The authors declare that they have no competing interests.

Data availability: The aggregated data supporting the findings of this study are available from the corresponding author on reasonable request.

References

1. GBD 2021 Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *Lancet*. 2024;404(10459):1199–1226. doi:10.1016/S0140-6736(24)01867-1
2. World Health Organization. Global antibiotic resistance surveillance report 2025. Geneva: World Health Organization; 2025. ISBN 9789240116337. Available from: <https://www.who.int/publications/i/item/9789240116337>
3. Bell BG, Schellevis F, Stobberingh E, Goossens H, Pringle M. A systematic review and meta-analysis of the effects of antibiotic consumption on antibiotic resistance. *BMC Infect Dis*. 2014;14:13. doi:10.1186/1471-2334-14-13
4. Klein EY, Impalli I, Poleon S, Denoel P, Cipriano M, Van Boeckel TP, et al. Global trends in antibiotic consumption during 2016–2023 and

- future projections through 2030. *Proc Natl Acad Sci U S A*. 2024;121(49):e2411919121. doi:10.1073/pnas.2411919121
5. Koya SF, Ganesh S, Selvaraj S, Wirtz VJ, Galea S, Rockers PC. Consumption of systemic antibiotics in India in 2019. *Lancet Reg Health Southeast Asia*. 2022;4:100025. doi:10.1016/j.lansea.2022.100025
 6. Davey P, Marwick CA, Scott CL, Charani E, McNeil K, Brown E, et al. Interventions to improve antibiotic prescribing practices for hospital inpatients. *Cochrane Database Syst Rev*. 2017;2(2):CD003543. doi:10.1002/14651858.CD003543.pub4
 7. Laxminarayan R, Klugman KP. Communicating trends in resistance using a drug resistance index. *BMJ Open*. 2011;1(2):e000135. doi:10.1136/bmjopen-2011-000135
 8. Klein EY, Tseng KK, Pant S, Laxminarayan R. Tracking global trends in the effectiveness of antibiotic therapy using the Drug Resistance Index. *BMJ Glob Health*. 2019;4(2):e001315. doi:10.1136/bmjgh-2018-001315
 9. Sapkale GP, Dhakad PK, Arora A, Panigrahi B, Bhowmik B, Malik H, et al. Drug Resistance Index correlates with antimicrobial resistance and declines following stewardship implementation: a multicentre analysis of 20 Indian hospitals (2016–2023). *Nat Resour Hum Health*. 2026;6(8s):493–505. doi:10.53365/nrfhh.1407
 10. Sakagianni A, Koufopoulou C, Feretzakis G, Kalles D, Verykios VS, Myrianthefs P, et al. Using machine learning to predict antimicrobial resistance—a literature review. *Antibiotics (Basel)*. 2023;12(3):452. doi:10.3390/antibiotics12030452
 11. Tang R, Luo R, Tang S, Song H, Chen X. Machine learning in predicting antimicrobial resistance: a systematic review and meta-analysis. *Int J Antimicrob Agents*. 2022;60(5–6):106684. doi:10.1016/j.ijantimicag.2022.106684
 12. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–297. doi:10.1007/BF00994018
 13. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. doi:10.1023/A:1010933404324
 14. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van

- Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol.* 2019;110:12–22. doi:10.1016/j.jclinepi.2019.02.004
15. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* (N Y). 2023;4(9):100804. doi:10.1016/j.patter.2023.100804
 16. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. doi:10.1136/bmj-2023-078378
 17. WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC classification and DDD assignment 2026. Oslo: Norwegian Institute of Public Health; 2026.
 18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12:2825–2830.
 19. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng.* 2007;9(3):90–95. doi:10.1109/MCSE.2007.55
 20. Waskom ML. seaborn: statistical data visualization. *J Open Source Softw.* 2021;6(60):3021. doi:10.21105/joss.03021
 21. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag.* 2009;45(4):427–437. doi:10.1016/j.ipm.2009.03.002
 22. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics.* 2007;8:25. doi:10.1186/1471-2105-8-25
 23. Hancock JT, Khoshgoftaar TM. Survey on categorical data for neural networks. *J Big Data.* 2020;7:28. doi:10.1186/s40537-020-00305-w
 24. World Health Organization. The selection and use of essential medicines, 2025: WHO AWaRe (Access, Watch, Reserve) classification of antibiotics for evaluation and monitoring of use. Geneva: World Health Organization; 2025. Available from: <https://www.who.int/publications/i/item/B09489>

25. Steyerberg EW, Harrell FE Jr.
Prediction models need appropriate
internal, internal–external, and
external validation. *J Clin Epidemiol*.
2016;69:245–247.
doi:10.1016/j.jclinepi.2015.04.005
26. Roberts DR, Bahn V, Ciuti S, Boyce
MS, Elith J, Guillera-Aroita G, et al.
Cross-validation strategies for data
with temporal, spatial, hierarchical, or
phylogenetic structure. *Ecography*.
2017;40:913–929.
doi:10.1111/ecog.02881
27. Altman DG, Royston P. The cost of
dichotomising continuous variables.
BMJ. 2006;332(7549):1080.
doi:10.1136/bmj.332.7549.1080
28. Lundberg SM, Lee SI. A unified
approach to interpreting model
predictions. In: *Advances in Neural
Information Processing Systems 30*
(NIPS 2017). Long Beach, CA; 2017.
p. 4765–4774.