Predicting antibiotic resistance in patients with postoperative infections using machine learning based models

F.J.N. (Friso) Engel

Master thesis Technical Medicine

Leiden University Medical Center

Department of Intensive Care Medicine

March 2023 – October 2023

This page intentionally left blank

Predicting antibiotic resistance in patients with postoperative infections using machine learning based models

F.J.N. (Friso) Engel, BSc.
Student Number: 4444000
1 October 2023

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Delft University of Technology – Leiden University Medical Centre - Erasmus Medical Centre

Master thesis project - TM30004 – 35 ECTS
Track: Sensing and stimulation
Faculty of Mechanical, Maritime and Materials Engineering (3mE), TU Delft
March 2023 – October 2023

Thesis committee members and supervisors

dr. M.S. (Sesmu) Arbous	LUMC	Chair & Medical supervisor
prof.dr.ir. M.J.T. (Marcel) Reinders	TU Delft	Technical supervisor
prof. dr. ir. J. (Jaap) Harlaar	TU Delft	External committee member
S.L. (Siri) van der Meijden, MSc.	LUMC	Daily supervisor
A.M. (Anna) van Boekel, MSc.	LUMC	Daily supervisor

TU Delft Delft University of Technology
Leids Universitair Medisch Centrum
Erasmus MC Universitair Medisch Centrum Rotterdam

This page intentionally left blank

Preface & Acknowledgements

In the ever-evolving landscape of healthcare, the battle against antibiotic resistance remains a critical and multifaceted challenge. The emergence and spread of antibiotic-resistant pathogens pose a significant threat to public health, particularly in patients recovering from postoperative infections. Addressing this issue requires innovative approaches, and this Master thesis represents my effort to contribute to the solution. I am both joyful and excited to present my Master thesis, titled "Predicting Antibiotic Resistance in Patients with Postoperative Infections Using Machine Learning Based Models," as the closing chapter of my Master of Science in Technical Medicine, specializing in Sensing and Stimulation techniques.

This research journey has been both intellectually stimulating and personally rewarding, offering a deep dive into the world of machine learning and its applications in healthcare. Over the past 12 months, I have dedicated myself to exploring the intricate interplay between data, algorithms, and clinical insights, striving to develop predictive models that can assist clinicians in making informed decisions regarding antibiotic treatment.

Writing this page concludes my Master thesis in Technical Medicine and marks the official end of my student years. I could never have reached this milestone on my one. I want to thank my supervisors Sesmu and Marcel who provided continuous guidance by challenging me in critical thinking throughout the process and by supporting me with valuable feedback on the content of the current paper. Their expertise in the fields of infectious diseases and machine learning not only shaped the direction of this research but also inspired me to reach higher and delve deeper into the subject matter. I'd like to thank my daily supervisors Siri and Anna whose unwavering support, responsiveness, and constructive feedback significantly contributed to the progress of this research. I extend my thanks to the members of HealthPlus.ai and the AI team at LUMC for their warm welcome and for fostering a collaborative environment. Mom, dad, Christine, without your everlasting support during the highs and lows of the last year, I couldn't have been where I am today.

Armed with the knowledge and experiences acquired during my academic journey and clinical internships, I am eager to channel my expertise towards innovating healthcare. I am driven to contribute to the strategic development and integration of new technologies that can enhance healthcare outcomes and patient well-being.

Gratefully reflecting on an memorable period of my life, but especially looking forward to new adventures and experiences.

Friso Engel
October 2023, Leiden

This page intentionally left blank

Abstract

Introduction

In the era of growing antimicrobial resistance, early detection and immediate treatment of antibiotic-resistant infections are crucial to ensuring successful outcomes in critically ill patients. The aim of this study is to apply machine learning (ML) to create classifiers that predict antibiotic resistance in postoperative infections caused by gram-negative pathogens, based on information retrievable from the Electronic Health Record. To determine if training prediction models on specific cultures improve the predictions' performance, eight sub-datasets have been created that only included specific culture sources or cultures containing specific bacteria.

Methods

All surgical procedures in the LUMC between January 2015 and May 2023 after which gram-negative bacteria had been cultured within 30 days after surgery were included. Logistic regression, random forest and support vector machine models were developed, evaluated using area under the receiver operating characteristics (AUROC) metric and calibrated afterwards. For each model, the most important predictors were determined using SHAP values.

Results

In total 27 models were created for the dataset and eight sub-datasets. The dataset containing all positive postoperative cultures within 30 days of surgery consisted of 5777 procedures with a resistance rate of 27.4%. The AUROC for the models developed for the whole dataset ranged between 0.65-0.68 on an unseen test set. The AUROC on unseen data for models developed on specific culture sources ranged between 0.63-0.79 and for those trained on specific bacteria between 0.63-0.75. The included features that were deemed most important are the presence of a previous resistant culture, abdominal surgery, the presence of an indwelling device after surgery, and the patient's sex, considering all (sub-)datasets.

Conclusions

This study shows that ML holds promise for predicting antimicrobial resistance. However, the current results are insufficient to support the implementation of these models in clinical practice to assist clinicians in choosing appropriate antibiotic therapy prior to receiving antibiogram results. With the current research, it cannot be proved with certainty that training models to particular postoperative infections enhances the predictions' performance. To further investigate the potential clinical benefit of applying ML prediction models in the context of antibiotic resistance, further research is necessary on extracting more features, with increased quality, which are available at the time the culture is taken.

Keywords: Antimicrobial, Resistance, Machine Learning, Prediction, Postoperative infections.

List of content

List of figures & tables

List of abbreviations

Abbreviation	Definition
2GC	Second Generation Cephalosporins
3GC	Third Generation Cephalosporins
AMR	Antimicrobial Resistance
AUROC	Area Under the Receiver Operating Characteristic
DPA	Data Protection Authority
EHR	Electronic Health Record
E. coli	Escherichia coli
EPV	Events Per Variable
K. pneumoniae	Klebsiella pneumoniae
LUMC	Leiden University Medical Centre
LR	Logistic Regression
ML	Machine Learning
P. aeruginosa	Pseudomonas aeruginosa
RF	Random Forest
SD	Standard Deviation
SHAP-values	Shapley Additive Explanations values
SVM	Support Vector Machine
TRIPOD-criteria	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis criteria

Introduction

Bacterial antimicrobial resistance (AMR) is a severe concern to public health. The prosperity of successful antibiotic treatment has been predicted to decline due to significant increases in antibiotic resistance rates [1]. Use of broad-spectrum antibiotics is a routine part of medical care for a significant proportion of hospitalized patients globally, making a major contribution to the emergence and global spread of multidrug-resistant pathogens [2]. WHO and numerous other groups and researchers agree that the spread of AMR is an urgent issue, as at this time, the death toll from AMR is thought to be a startling 700,000 every year [3-7]. Additionally, patients with AMR pathogens are more likely to experience lengthier hospital stays and worse outcomes, which also comes with a large financial burden [8-10].

Gram-negative infections are especially dangerous because they are highly effective at acquiring and upregulating the genes that code for antimicrobial treatment resistance [11]. Since gram-negative infections are most frequently treated with these antibiotics, resistance to second- and third-generation cephalosporins (2GC and 3GC), aminoglycosides, and carbapenems is clinically significant [12]. Susceptibility testing allows for the detection of resistance in infections, but the process can take up to multiple days [13]. See *Figure 1* for more information on the clinical timeline.

Before the results from susceptibility testing, i.e., antibiogram, become available, clinicians might prescribe empiric antibiotic treatment. This is the practice where therapy is started before a patient's precise bacterial aetiology, source of infection, or antibiotic resistance profile of the infecting bacteria are determined [14]. Empiric therapy is essential since it may be required to take rapid action, however its inherent reliance on empirical clinical judgement illustrates its constraints, given the limited accessible to doctors by that time.

The field of machine learning (ML) has emerged as a promising approach to the challenging task of assisting clinicians in selecting the appropriate antibiotic medication [15, 16]. ML is a subfield of artificial intelligence, focusing on creating algorithms that can learn from data and anticipate certain outcomes with little to no human intervention [17]. The interest for using ML in healthcare has increased during the last ten years. This is brought on by the exponential expansion in computing power, the availability of biological and medical data, and the considerable advancements in algorithm development.

Within the field of ML, selecting the appropriate scope and corresponding data is an important aspect in the development of useful algorithms. Narrowing the focus to address more specific prediction issues might be beneficial, as by doing

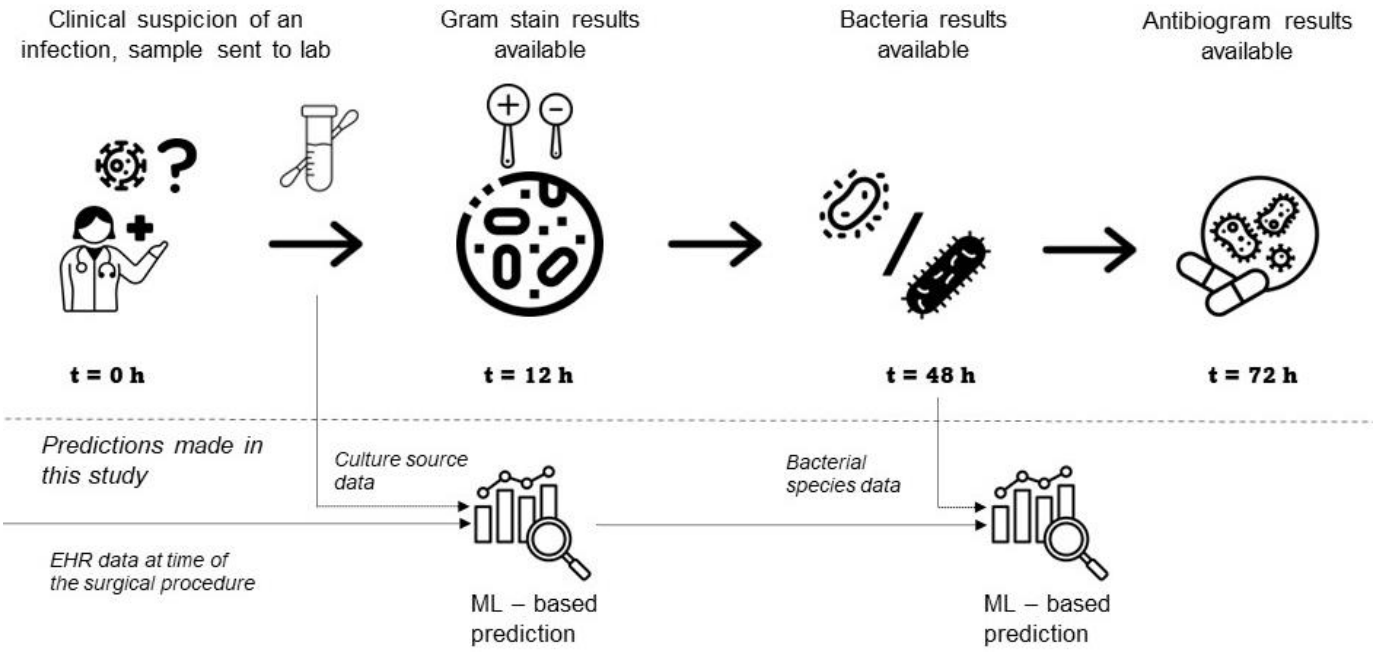


Figure 1. Clinical timeline of obtaining antibiogram results. It can take up to 72 hours before results of the susceptibility testing are available to clinicians. In this study, machine learning (ML)-based predictions of results are predicted at the time of gram stain availability (t = 12 h), and at the time of bacteria results availability (t = 48 h), using data from the EHR at time of the surgical procedure, the culture source, and data from about the bacterial species cultured.

this, researchers are able to tailor their models to the specific challenges and capture unique patterns and intricacies within the selected, more comparable, data. This approach enables the creation of specialized models that can produce more accurate predictions and valuable insights.

Previous studies have demonstrated the potential of ML-based prediction models to produce precise predictions in the context of AMR in hospitalized patients [18-20].
A recent systematic review by the authors of the current study found that earlier research on the use of ML in AMR prediction was done in a variety of patient populations [21]. Interestingly however, no previous research has been found that investigated how ML might be used to predict antibiotic resistance in infections in postoperative patients, although roughly 25–30% of all hospital acquired infections could be regarded as postoperative infections [22, 23]. In order to close this knowledge gap, this study investigates the ability of multiple ML algorithm techniques to create prediction models for antibiotic resistance to 2GC, 3GC, aminoglycoside, and/or carbapenem in postoperative patients with a gram-negative infection. To research whether narrowing the predictions' scope to more comparable patients influences the predictions' performance, separate models were developed for specific sub-populations. These groups were defined by focusing on specific infection sources, or infections brought on by certain pathogens.

Methods

In this single-centre study, ML models were developed using retrospective data from the Leiden University Medical Centre (LUMC), an 880-bed tertiary academic hospital in the Netherlands, to predict resistance to 2GC, 3GC, aminoglycoside, and/or carbapenem in postoperative patients that had a postoperative culture with gram-negative bacteria within 30 days after surgery. To present modelling techniques, the TRIPOD criteria (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) were adhered to [24].

2.1 Ethics statement

The LUMC's Medical Research Ethics Committee, the data protection authority (DPA) waived approval for the study protocol.

2.2 Source of data

The pseudo-anonymized database used for this study was created retrospectively. Data was requested from the LUMC Data platform based on the Electronic Health Record (EHR, HiX, ChipSoft, Amsterdam, The Netherlands). The data was collected as part of the PERISCOPE project (protocol number: G18.129) and consisted of health records from every surgical patient at the LUMC between January 2015 and May 2023. Data was collected per surgical procedure, and multiple procedures per patient have been included.

2.3 Study population

Surgical procedures for all adult (≥ 18 years of age) patients who had a culture containing gram-negative bacteria within 30 days after their procedure for which an antibiogram has been performed were considered for this research. Procedures were excluded if the surgery performed was non-invasive (e.g., biopsies, punctions), radiotherapy, psychiatric, procedures in pregnant patients, cardiological, outpatient, or sole anaesthetic.

To determine if training separate prediction models on specific types of infections improved the predictions' performance, sub-datasets were created for eight sub-populations. Five sub-populations were defined based on the source of the postoperative cultures: ascites, blood, respiratory, urine or wound cultures, and three sub-populations were defined based on the presence of specific bacteria in the cultures: *Escherichia* (E.) *coli*, *Klebsiella* (K.) *pneuomniae* and *Pseudomonas* (P.) *aeruginosa*. Procedures may be included in multiple sub-datasets, as they are not mutually exclusive.

The procedures were included in the (sub-)datasets using culture results and bacterial data as primary sources of information. This was done as addressing clinical infections or causative pathogens directly was retrospectively unfeasible due to limitations in the data available for this research.

2.4 Outcome

The predicted outcome in this study was resistance to 2GC, 3GC, aminoglycoside, and/or carbapenem antibiotics in patients that had a positive gram-negative culture, cultured within 30 days postoperatively. More information on the antibiotics in these categories is provided in the *Supplementary materials*.

Culture and susceptibility testing was processed at the LUMC using standard microbiology guidelines. During culture testing, a patient's swab is cultured and assessed for possible pathogenic bacteria by medical microbiologists. When the microbiologists

regard the bacteria found to be pathogenic, the culture is classified as positive. During susceptibility testing, bacterial strains are classified as susceptible (S), resistant (R), or intermediate resistant (I) to the tested antibiotics. For more information on these guidelines see the *Supplementary materials.*
In this study, the procedures were labelled as being resistant when susceptibility testing on any positive gram-negative culture taken within 30 days postoperative showed resistance (R) or intermediate resistance (I) to the antibiotics of interest. All other procedures were labelled as being non-resistant. It was decided to include cultures with intermediate resistance in the resistant category, because when intermediate resistance is found in a bacterial strain, similar to resistant strains, the patient commonly requires alternate antibiotic therapy or higher antibiotic dosages then when no resistance is found. Also, isolation is required for both patients in which resistance and intermediate resistance is found [25].

2.5 Variables and pre-processing

For both the dataset for the total study population, and the sub-datasets, data was split into two groups: a training set made up of 80% of the data and a test set made up of 20%, to prevent data leaking and overfitting. Splitting was done at random, except for stratification for the outcome, which kept the proportions of the binary (resistant, not resistant) outcomes in each set the same as they were in the full dataset. The test set was kept separate until final model evaluation, whereas the training dataset was used for pre-processing, such as missing data imputation, and model development.

The available extracted features could be described as variables regarding *patient characteristics, the*

medical status, medication, laboratory results, procedure characteristics, or culture characteristics. These features were partly made available in the context of the PERISCOPE project for predicting postoperative infections, and partly specifically extracted for the current study. For more details about both the available features, and the used methods for feature extraction see the *Supplementary materials*.

All features were checked for missing data. Missing data in the numerical variables were imputed by the median value. To prevent (biologically) impossible values to be included in the analyses, outlier thresholds were determined. The values exceeding these thresholds were handled as missing data as well. Numerical variables regarding procedure characteristics (e.g., planned procedure length) were imputed by the median of the corresponding medical specialty. Missing data in the categorical variables was handled by creating a new additional category: '*unknown'*, as this prevents bias from non-random missing data. Variables were excluded when missing data was more than 50% of the dataset.

After missing data was handled, one-hot encoding was used to convert the categorical data to binary features, which is necessary for an efficient implementation of ML algorithms. Standard scaling was applied to ensure all variables had a mean of zero and a standard deviation of one.

2.6 Model development and evaluation

2.6.1 Model development

In this study, three ML classifier techniques (logistic regression, random forest, and support vector machine) were compared to predict antibiotic resistance in the postoperative cultures. Logistic regression (LR), a classical statistical approach for prediction modelling, was chosen as a fundamental baseline model [26]. It is important to note that while LR has its roots in classical statistics, it is now also considered a ML technique due to its adaptability to various applications. Random forest (RF) was selected as the second classifier because it is commonly used in healthcare-related prediction models and has demonstrated relatively good performance in such contexts [27]. Lastly, support vector machine (SVM) was selected since it represents a fundamentally other approach to ML compared to the other two techniques [28]. SVM can uncover complex patterns in data by changing them into a higher-dimensional form and creating optimal decision boundaries to uncover subtle relationships in the antibiotic resistance data that the other techniques might miss.

These model techniques were optimized by applying three development steps: feature selection, hyperparameter tuning, and SMOTE oversampling to correct for class imbalances. The best combination of feature selection method and hyperparameter settings was sought by optimizing the Area Under the Receiver Operating Characteristic (AUROC) using a grid search [29] (LR), or Bayesian optimalisation [30] (RF and SVM) with a five-fold cross-validation scheme on

the training set of the dataset regarding the total study population, containing all procedures with positive postoperative cultures. The AUROC metric demonstrates the relationship between the true positive rate and false positive rate, independent of a chosen threshold at what predicted probability a sample should be considered as resistant. A description of the tested feature selection methods and hyperparameter settings is available in *Supplementary materials*. An overview of the complete model development setup is shown in *Figure 2.*

2.6.2 Model evaluation

After optimizing the model techniques, LR, RF, and SVM models were trained on the dataset for the total study population and all sub-datasets separately. The robustness of the models was assessed by addressing the mean AUROC for a 10-fold cross-validation within the training set. The performance of each model was determined by the AUROC on the unseen test set.

2.6.3. Model interpretation & explainability

Shapley Additive exPlanations (SHAP) values were used to visualize the impact and the relationship to the outcome of the most important features for each model [31]. The top 10 features, ranked by their mean absolute SHAP value, were used to assess feature importance for each of the models. To compare the most important features for the total study population with those for the different sub-populations, a heatmap is created containing the importance of features that appeared in the top five of one of the developed models. The scoring method used for the heatmap can be found in the *Supplementary materials.*

2.6.4 Model calibration

To correctly interpret the output of a prediction value in terms of a probability, a model's predictions might need to be calibrated. To determine what calibration method provided the best calibration, the models' output on the test set were calibrated using Platt scaling and Isotonic regression. Calibration plots were created for each model, also including the uncalibrated output. The calibration method resulting in the combination of slope and intercept closest to a perfectly calibrated model (a slope of 1.0 and an intercept of 0.0), was deemed best for the corresponding model.

2.6.5 Software

All analyses were performed using Python 3.8. and the following packages: Imbalanced-learn 0.11.0, Matplotlib 3.6.2, NumPy 1.23.4, Pandas 1.5.1, Seaborn 0.12.1, and Scikit-learn 1.1.3. The code to obtain the results can be found in the LUMC git environment using the following link: Engel / Thesis Antibiotic Resistance Prediction · GitLab (lumc.nl) (for access, contact the main author).

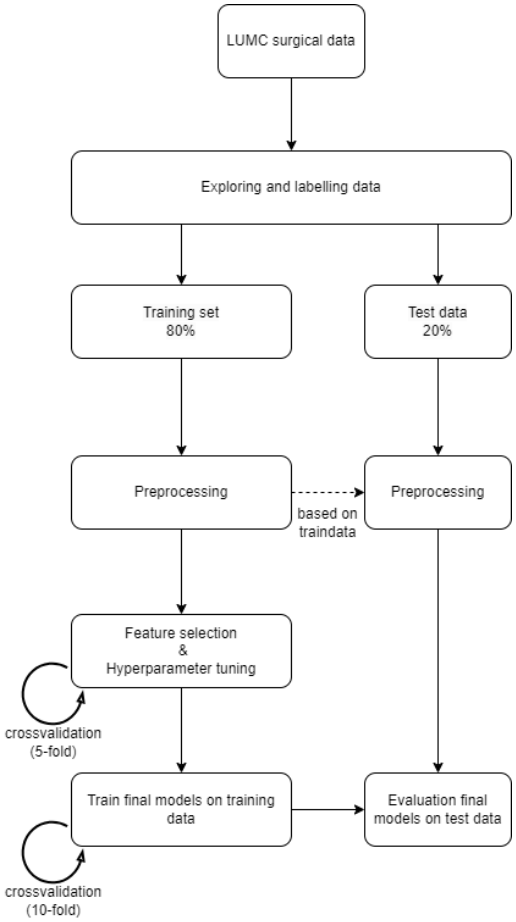


Figure 2. Setup for model development. Tuning of hyperparameters and selecting the optimal feature selection methods were based on the prediction models' results for the total study population.

Results

3.1 Participants

The initial PERISCOPE database contained a total of 69,326 procedures from 49,497 patients at the time of this study. Of these procedures 5,777 were included in the dataset for the total study population, containing positive postoperative cultures with gram-negative bacteria, which had an antibiogram result. The incidence of resistance in the study population was 27.4%. *Table 1* provides an overview with detailed information on the included study population.

3.2 Model development

Data pre-processing, including encoding of categorial features, resulted in a maximum of 72 features for the dataset regarding the total study population. The sub-populations regarding the specific culture sources were defined based on the *culture source* features,

Table 1. Detailed information on the dataset for the total study population

		Training set	Test set	Total	Missing values n, % (train, test)
Total	Procedures, n (% of total)	4621 (80)	1156 (20)	5777 (100)	-
	Resistant procedures, n (%)	1265 (27.4)	317 (27.4)	1265 (27.4)	-
	Patients, n	3525	1083	4178	-
Patient characteristics	Sex, n (%)				-
	Male	2635 (57)	669 (58)	3304 (57)	
	Female	1986 (43)	487 (42)	2473 (43)	
	Age, median (25-75th percentile)	65 (53-73)	65 (52-73)	65 (53-73)	-
	BMI, mean (SD)	26.4 (5.5)	26.4 (5.5)	26.5 (5.7)	333, 6% (266, 67)
Medical status	ASA classification, n (%)				1719, 30%
	ASA 1	254 (5)	58 (5)	312 (5)	(1375, 344)
	ASA 2	1535 (33)	393 (34)	1928 (33)	
	ASA 3	1263 (27)	314 (27)	1577 (27)	
	ASA 4	183 (4)	40 (3)	223 (4)	
	ASA 5	11 (0)	7 (1)	18 (0)	
	Source of admission, n (%)				390, 7% (317,73)
	Residential environment	3931 (85)	980 (85)	4911 (85)	
	Transferred	373 (8)	103 (9)	476 (8)	
	Previously admitted to LUMC, n (%)	1411 (31)	362 (31)	1773 (31)	-
	Previously admitted to ICU, n (%)	559 (12)	136 (12)	695 (12)	-
	Previously presence of indwelling device, n (%)	1233 (27)	333 (29)	1566 (27)	-
	Previous resistant pathogen, n (%)	413 (9)	105 (9)	518 (9)	-
Medication	Currently used medication, n (%)				-
	Anti-arrythmics	290 (6)	81 (7)	371 (6)	
	Beta blockers	1127 (24)	275 (24)	1402 (24)	
	Ca-channel blockers	706 (15)	170 (15)	876 (15)	
	Anti-thrombotics	948 (21)	244 (21)	1192 (21)	
	Diabetes medication	1009 (22)	265 (23)	1274 (22)	
	Chemotherapy	322 (7)	83 (7)	405 (7)	
	Immunosuppressants	941 (20)	266 (23)	1207 (21)	
	Diuretics	1124 (24)	303 (26)	1427 (25)	
	Raas inhibitors	761 (16)	204 (18)	965 (17)	
Laboratory results	CRP measured last week	1102 (24)	293 (25)	1395 (24)	-
Procedure characteristics	Specialty, n (%)				-
	General surgery	2573 (56)	651 (56)	3224 (56)	
	Urology	555 (12)	166 (14)	721 (12)	
	Neurosurgery	483 (10)	96 (8)	579 (10)	
	ENT	301 (7)	70 (6)	371 (6)	
	Orthopaedics	250 (5)	63 (5)	313 (5)	
	Gynaecology	253 (5)	56 (5)	309 (5)	
	Cardiothoracic surgery	83 (2)	23 (2)	106 (2)	
	Plastic surgery	64 (1)	19 (2)	83 (1)	
	Otolaryngology surgery	59 (1)	12 (1)	71 (1)	
	Area of surgery, n (%)				-
	Abdomen	2222 (48)	596 (52)	2818 (49)	
	Extremity	524 (11)	133 (12)	657 (11)	
	Head	351 (8)	69 (6)	420 (7)	
	Thorax	297 (6)	73 (6)	370 (6)	
	Genitals	275 (6)	66 (6)	341 (6)	
	Brain	240 (5)	54 (5)	294 (5)	
	Hip	128 (3)	32 (3)	160 (3)	
	CNS	44 (1)	4 (0)	48 (1)	
	Other	540 (12)	129 (11)	669 (12)	
	Type of procedure, n (%)				-
	Laparotomic	582 (13)	128 (11)	710 (12)	
	Laparoscopic	176 (4)	43 (4)	219 (4)	
	Implant	111 (2)	32 (3)	143 (2)	
	Other	3752 (81)	953 (82)	4705 (81)	
	Priority procedure, n (%)				-
	Emergency	515 (11)	109 (9)	624 (11)	
	Urgent	1488 (32)	388 (34)	1876 (32)	
	Elective	2618 (57)	659 (57)	3277 (57)	
	Procedure count, median (range)	1 (1-4)	1 (1-4)	1 (1-4)	-

(Continued on next page)

Table 1. continued.

		Training set	Test set	Total	Missing values n, % (train, test)
Procedure characteristics *(continued)*	Planned procedure length in minutes, mean (SD)	120 (117)	128 (114)	122 (117)	1036, 18% (831, 205)
	Actual procedure length in minutes, mean (SD)	164 (119)	167 (121)	165 (119)	1703, 29% (1363, 340)
	Difference between planned and actual length in minutes, mean (SD)	44 (127)	39 (126)	43 (126)	- [3]
	Expected days preop, median (range)	0 (0-2)	0 (0-2)	0 (0-2)	-
	Expected days postop, median (range)	0 (0-14)	0 (0-14)	0 (0-14)	25, 0% (21, 4)
	Procedure risk, n (%)				-
	Low	242 (5)	49 (4)	291 (5)	
	Medium	1629 (35)	376 (33)	2005 (35)	
	High	2750 (60)	731 (63)	3481 (60)	
	Antibiotic prophylaxis, n (%)	2412 (52)	582 (50)	2994 (52)	-
	Presence of indwelling device after procedure, n (%)	799 (17)	196 (17)	995 (17)	-
Culture characteristics	Source [1]				1328, 23% (1063, 265)
	Urine	1304 (28)	319 (28)	1623 (28)	
	Wound	465 (10)	120 (10)	585 (10)	
	Respiratory	320 (7)	79 (7)	399 (7)	
	Blood	246 (5)	66 (6)	312 (5)	
	Ascites	250 (5)	55 (5)	305 (5)	
	Other: Fluid	339 (7)	96 (8)	435 (8)	
	Other: Tissue	312 (7)	72 (6)	384 (7)	
	Other: Undefined [2]	322 (7)	84 (7)	406 (7)	

[1] The number of these cultures do not match the numbers in the subsets, as only the first (resistant) culture per procedure is included as feature.

[2] Undefined was used as a category to describe when the cultures description was available, but it was too ambiguous to determine the exact source of the culture.

[3] This feature does not have missing data, as it is created after imputing the missing data in the *planned procedure length* and *actual procedure length* features.

so, these features were not included in the corresponding sub-datasets. A total of 27 classifiers have been developed and trained in this study: i.e. three different ML model techniques (LR, RF, and SVM) for the total study population and eight sub-populations (five regarding specific culture sources: *ascites, blood, respiratory, urine,* and *wound*, and three regarding specific bacteria of interest: *E. coli, K. pneumoniae, and P. aeruginosa*).

The *Elastic net* penalty produced the best mean AUROC on the cross-validation sets of the complete dataset using the LR model. For both the RF and the SVM models the *recursive feature elimination selection* method resulted in the best performance. An overview of the selected hyperparameters is shown in *Table 2*.

Table 2. Applied hyperparameters for all classifiers after optimization

	Feature selection method & hyperparameters	
Logistic Regression	Penalty	Elastic Net
	L1-ratio	0.3
	C	0.01
	Max iterations	2000
	Solver	Saga
Random Forest	Feature selection	RFE
	Max depth	20
	Max features	Log2
	Min samples split	25
	Number of estimators	80
Support Vector Machine	Feature selection	RFE
	C	0.001
	Kernel	Linear
	gamma	n/a
	Degree	n/a

Abbreviations: RFE = recursive feature elimination

3.3 Model performance & Feature importance

For the models that predicted antibiotic resistance regarding the total study population, the mean AUROCs from the 10 cross-validation folds within the training set are plotted in *Figure 3.*

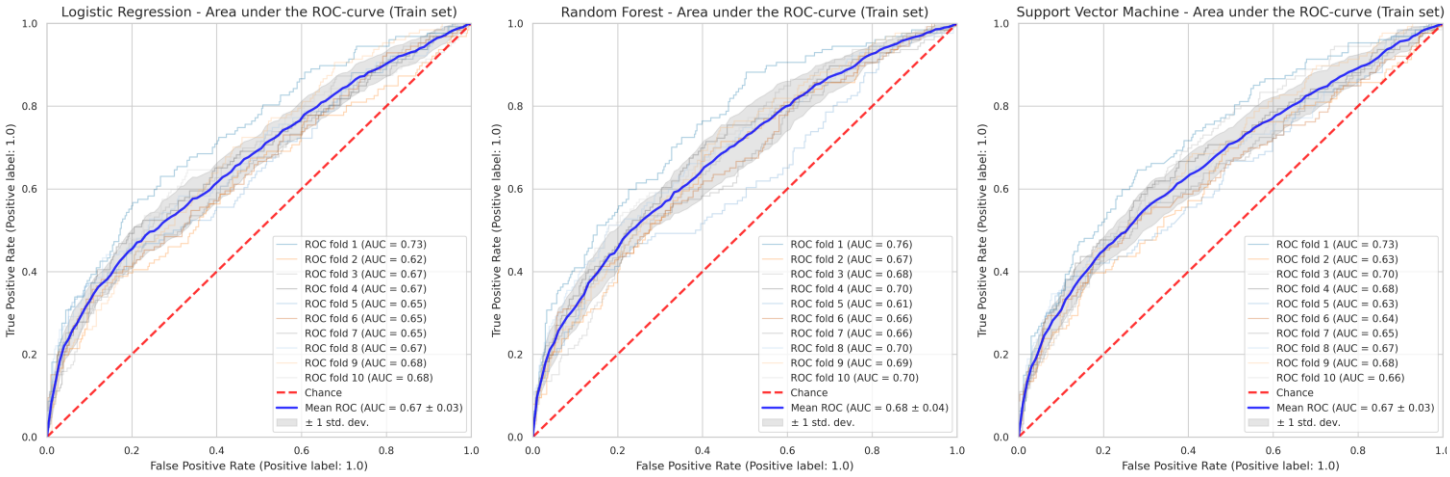


Figure 3. Mean AUROCs for the cross-validation folds in the training set for the three models regarding the prediction of antibiotic resistance in the total study population

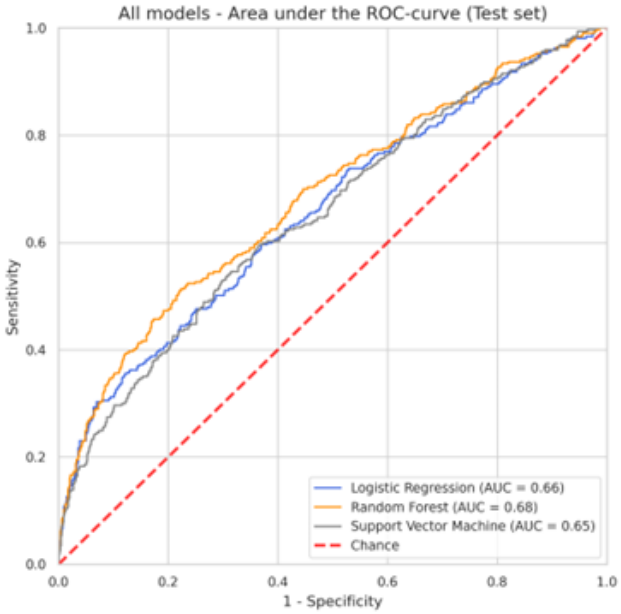


Figure 4. AUROCs for the test set for the three models regarding the prediction of antibiotic resistance in the total study population

The AUROCs on the unseen data in the test set for those models are shown in *Figure 4.*

Table 3 shows the achieved AUROCs for all developed models, including those based on the different sub-datasets. The figures including the AUROC plots for the models based on the sub-datasets can be found in the *Supplementary materials.*

Figure 5 shows the 10 features with the highest absolute mean SHAP value for the three classifiers regarding antibiotic resistance in the total study population. *Table 4* shows the created heatmap containing the importance of the 25 features that appeared in the top five most important features within any of the developed models. The figures showing the top 10 features for each model for every sub-population can be found in the *Supplementary materials.*

3.3.1 Models regarding the total study population

The LR classifier achieved an AUROC of 0.66 on the test set, with a mean AUROC of 0.67 (SD 0.03) on the training set. The RF classifier achieved on both the test set and the trainset an AUROC of 0.68, and the SVM classifier scored an AUROC of 0.65 on the test set while achieving a mean AUROC of 0.67 (SD respectively 0.04 and 0.03) on the train data.

Table 3. AUROCs for both the training set (mean of cross-validation folds) and the test set for all models and datasets

		Logistic Regression	**Random Forest**	**Support Vector Machine**
	(sample size, % events)	(training set (SD), test set)	(training set (SD), test set)	(training set (SD), test set)
Total study population	All procedures (5777, 27.4 %)	0.67 (0.03) 0.66	0.68 (0.04) 0.68	0.67 (0.03) 0.65
Sub-populations based on culture source	Ascites (495, 28.5%)	0.63 (0.08) 0.63	0.66 (0.13) 0.65	0.62 (0.12) 0.65
	Blood (608, 25.3%)	0.55 (0.07) 0.65	0.65 (0.08) 0.63	0.56 (0.12) 0.66
	Respiratory (536, 19.0%)	0.62 (0.10) 0.74	0.68 (0.10) 0.79	0.65 (0.09) 0.70
	Urine (1872, 24.7%)	0.63 (0.04) 0.67	0.63 (0.03) 0.68	0.61 (0.03) 0.68
	Wound (947, 26.8%)	0.69 (0.07 0.70	0.71 (0.07) 0.73	0.68 (0.08) 0.70
Sub-populations based on bacteria	*E. coli* (2247, 37.0%)	0.63 (0.03) 0.66	0.68 (0.03) 0.70	0.64 (0.03) 0.68
	K. pneumoniae (467, 31.7%)	0.69 (0.09) 0.63	0.71 (0.07) 0.68	0.67 (0.08) 0.67
	P. aeruginosa (1203, 12.6%)	0.72 (0.09) 0.75	0.74 (0.07) 0.73	0.72 (0.08) 0.73

Abbreviations: AUROC = Area under the Receiver Operation Characteristic, E. = Escherichia, K. = Klebsiella, P. = Pseudomonas

Five features were included in the most important features for all the three classifiers: Abdominal surgery, the presence of an earlier resistant culture, the use of prophylactic antibiotics during the procedure, laparotomic surgery, and the number of expected days of pre-operative admission. In all three classifiers, a positive correlation with the outcome was found for the first four features, while the number of expected days of pre-operative admission was found to have a negative correlation.

3.3.2 Models regarding the sub-populations based on culture source

Regarding the models that predicted antibiotic resistance in the sub-populations regarding specific culture sources, the best AUROCs on the test set were achieved by the models for the respiratory cultures (with AUROCs for LR, RF, and SVM being equal to 0.74, 0.79, 0.70, respectively). The ascites cultures' models achieved the lowest AUROCs for the test set (0.63, 0.65, and 0.65). The performances on the test sets are close to the performances achieved on the training sets (shown in *Table 3)*, which indicates low risk of bias due to overfitting. It is important to note that the standard deviation on the AUROCs within the trainset was quite large in most of these models, showing a lack of robustness.

Different combinations of features are shown in *Table 4* to be important when comparing the predictions for antibiotic resistance between the various sub-populations based on culture sources. Some features (e.g., the use of diabetes medication) are deemed especially important for certain culture sources, but not important at all for other sources. The only feature with high importance in all the researched sub-populations based on culture sources is the presence of a previous resistant culture, with a positive correlation to the outcome in all sub-populations. The male sex had a positive correlation to the outcome for ascites, blood, and urine cultures, while it had a negative correlation for respiratory cultures.

3.3.3 Models regarding the sub-populations based on bacteria

The models that predicted antibiotic resistance using postoperative cultures containing *E. coli* strains achieved AUROCs on the test set of 0.66, 0.70 and 0.68 for respectively the LR, RF, and SVM models. The AUROCS for the test sets of the *K. pneumoniae* models were 0.63, 0.68, and 0.67.

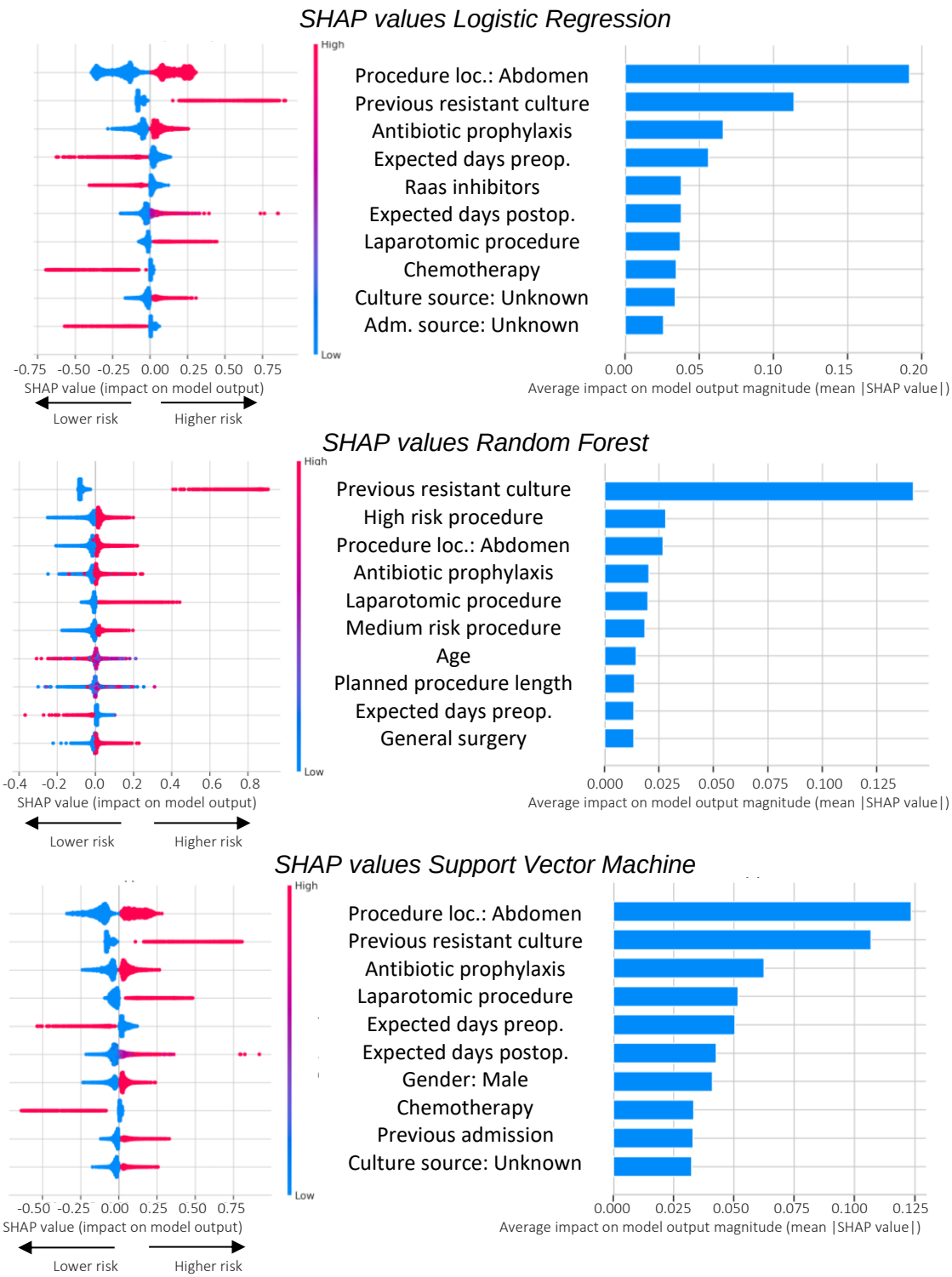


Figure 5. SHAP summary plots for the three models regarding the prediction of antibiotic resistance in the total study population.

***Abbreviations**:* SHAP = Shapley additive explanations, loc. = location, preop = preoperative, postop = postoperative, Adm. = Admission.

The best results were obtained by the models regarding *P. aeruginosa* strains, with AUROCS on the test set of 0.75, 0.73, and 0.73. Just as with the models for specific culture sources, are the test set results comparable to the training set results in *Table 3*, indicating a low likelihood of bias brought on by overfitting. The standard deviation of the mean AUROCs for the training set in the three *E. coli* models were 0.03, while those for the *K. pneumoniae* and *P. aeruginosa* were 0.09, 0.07 and 0.08 for the LR, RF, and SVM models, respectively.

According to the heatmap shown in *Table 4* is the presence of a previous resistant culture also an

Table 4. Heatmap of features with greatest importance per (sub-)population.

Feature\(sub-)dataset	**Total population**	**Ascites**	**Blood**	**Respi-ratory**	**Urine**	**Wound**	***E. coli***	***K. pneu.***	***P. aeru.***
Antibiotic prophylaxis									
ASA score 1									
ASA score 3									
Culture source: Unknown									
Expected postop days									
Expected preop days									
Sex (Male)									
Indwelling device after procedure									
Infection type: wound	n/a	n/a	n/a	n/a	n/a	n/a			
Laparotomic procedure									
Medication: Antiarrhythmics									
Medication: Antithrombotics									
Medication: Ca-channel blockers									
Medication: Diabetes									
Medication: Diuretics									
Medication: Immunosuppressants									
Medication: Raas inhibitors									
Prev. admission									
Prev. ICU admission									
Prev. resistant culture									
Prev. presence of Indwelling device									
Procedure length									
Procedure length difference									
Procedure location: Abdomen									
Procedure risk: High									

Abbreviations: E. = Escherichia, K. pneu. = Klebsiella pneumoniae, P. aeru. = Pseudomonas aeruginosa n/a = not applicable, prev = previous, (-) = negative correlation with outcome, (+) = positive correlation with outcome

important feature for the predictions of antibiotic resistance regarding cultures containing the bacteria of interest. Interestingly, the presence of an indwelling device after the procedure is deemed as important in both *E. coli* and *K. pneumonia* models, but not in the models regarding *P. aeruginosa*.

3.4 Calibration

To show what a calibration plot looks like, the calibration curves for the LR classifier for antibiotic resistance prediction in the total study population are shown in *Figure 6,* as an example. *Table 5* shows the slope and the intercept for the outcomes for all 27 models, both uncalibrated and after applying the calibration methods. The calibration plots for the other models can be found in the *Supplementary materials*. The Sigmoid calibration method provided the best calibration the most often (13 out of 27 models). The uncalibrated models had the best calibration in 10 out of 27 models. The remaining four models were best calibrated by the isotonic method. It is noteworthy that the RF models were best calibrated without any extra calibration method (nine out of 10 models).

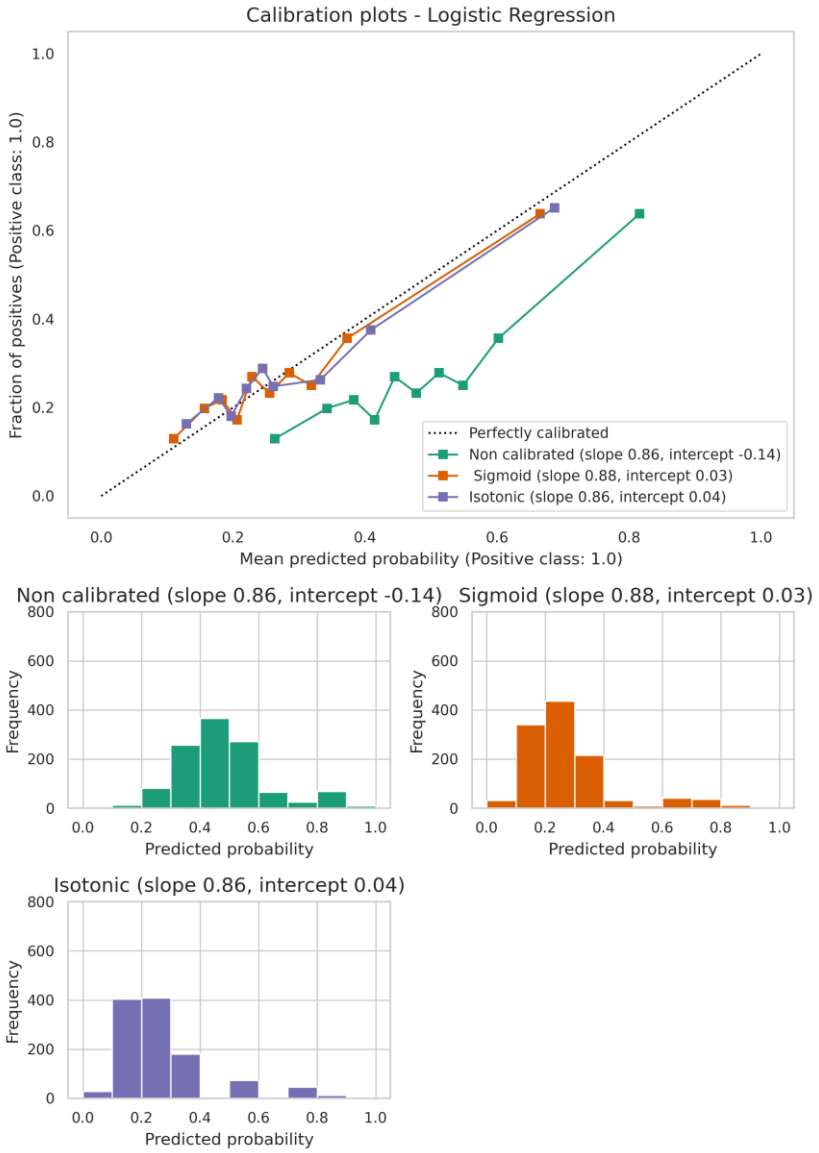


Figure 6. Calibration curve and prediction histograms for the logistic regression model regarding the prediction of antibiotic resistance in the total study population. The shown histograms represent the distribution of the predicted probabilities for the calibration methods.

Discussion

4.1 Main findings

In this study LR, RF, and SVM classifiers were built to predict antibiotic resistance in postoperative cultures, both in general and while focusing on sub-populations either based on specific sources of cultures or based on cultures containing specific bacteria. The AUROC performances for the LR, RF, and SVM classifiers regarding the total study population are respectively 0.66, 0.68, and 0.65 on the test set. For the classifiers trained on data using specific culture sources the AUROCs ranged between 0.63-0.79 on their test sets, with the models for predicting antibiotic resistance in respiratory cultures performing best. The classifiers trained on data regarding specific bacteria achieved AUROCs ranging between 0.63-0.75 on their test sets, with the models for the *P. aeruginosa* bacteria performing best. The RF models outperformed the other ML techniques for the separate prediction cases, but not substantially.

According to generally used standards, AUROCs can be considered moderate between 0.70-0.80, good between 0.80-0.90, and outstanding >0.90 [32]. The current results suggest that resistance in postoperative cultures can be predicted by ML with a moderate performance. The large standard deviation (up to 0.13 in terms of AUROC) found in the cross-validated performance suggest that the robustness of the classifiers should be improved before any claims on optimizing the performance by focusing on specific bacteria or culture sources may be validated.

Table 5. Calibration scores for the prediction models (slope, intercept) [1]

			Logistic Regression	Random Forest	Support Vector Machine
Total study population					
		Uncalibrated	0.86, -0.14	**1.12, -0.13**	0.73, -0.07
		Sigmoid	**0.88, 0.03**	0.46, 0.15	**0.93, 0.02**
		Isotonic	0.86, 0.04	0.44, 0.16	0.86, 0.05
Sub-populations based on culture source	Ascites				
		Uncalibrated	1.48, -0.40	**0.68, 0.04**	1.51, -0.45
		Sigmoid	**0.70, 0.12**	0.26, 0.22	**0.99, 0.00**
		Isotonic	0.62, 0.17	0.32, 0.23	0.56, 0.28
	Blood				
		Uncalibrated	2.32, -0.90	**0.86, -0.07**	1.47, -0.44
		Sigmoid	**0.75, 0.06**	0.26, 0.19	**0.60, 0.09**
		Isotonic	0.61, 0.06	0.08, 0.20	0.48, 0.12
	Respiratory				
		Uncalibrated	1.98, -0.76	1.38, -0.24	0.56, -0.05
		Sigmoid	1.27, -0.06	0.60, 0.09	**0.81, 0.04**
		Isotonic	**0.94, 0.01**	**0.63, 0.09**	0.59, 0.08
	Urine				
		Uncalibrated	1.13, -0.29	**1.30, -0.21**	0.85, -0.14
		Sigmoid	**0.93, 0.02**	0.49, 0.13	**1.00, 0.01**
		Isotonic	0.89, 0.03	0.45, 0.14	0.96, -0.04
	Wound				
		Uncalibrated	1.08, -0.24	**1.10, -0.14**	0.67, -0.04
		Sigmoid	**1.04, -0.01**	0.46, 0.15	**0.92, 0.02**
		Isotonic	1.05, -0.03	0.42, 0.15	0.74, 0.08
Sub-populations based on bacteria	*E. coli*				
		Uncalibrated	1.28, -0.27	**1.46, -0.25**	0.99, -0.12
		Sigmoid	**0.92, 0.02**	0.50, 0.19	0.86, 0.05
		Isotonic	0.84, 0.06	0.45, 0.21	**0.98, -0.00**
	K. pneumoniae				
		Uncalibrated	**1.45, -0.39**	**0.93, -0.11**	**0.96, -0.09**
		Sigmoid	0.55, 0.15	0.52, 0.06	0.75, 0.11
		Isotonic	0.53, 0.20	0.43, 0.05	0.64, 0.14
	P. aeruginosa				
		Uncalibrated	0.56, -0.11	**1.23, -0.17**	0.36, -0.02
		Sigmoid	**0.77, 0.02**	0.43, 0.19	0.58, 0.05
		Isotonic	0.70, 0.04	0.46, 0.16	**0.64, 0.03**

Abbreviations: E. = Escherichia, K. = Klebsiella, P. = Pseudomonas

[1] The scores in bold depict the best calibration method for the corresponding classifier, i.e., closest to the perfect calibration scores (slope 1.00, intercept 0.00).

Out of the 72 included features, the most impactful features according to the SHAP values while predicting antibiotic resistance in the total study population were Abdominal surgery, the presence of an earlier resistant culture, the use of prophylactic antibiotics during the procedure, laparotomic surgery, and the number of expected days of pre-operative admission. Narrowing the scope to specific sub-populations suggest that certain variables play a role in some populations, but not in others. Regarding the key features for the total study population and all sub-populations, the presence of a previous resistant pathogen, abdominal surgery, the presence of an indwelling device after surgery and the patient's sex were most influential during the predictions.

Regarding calibration methods, the sigmoid calibration method generally achieved the best calibration, while the RF models were better left uncalibrated.

4.2 Comparison with literature

This study is the first in predicting antibiotic resistance by using ML in the context of postoperative infections. Earlier literature by the

authors on predicting resistance in hospitalized patients has shown a wide range of results in terms of AUROC (0.63-0.93) [21]. The fact that RF outperformed the other classifiers is consistent with an earlier systematic review of healthcare related classifiers [27], although this finding is not supported specifically in literature regarding antibiotic resistance predictions [21, 33].

The current study used SHAP values to provide insights on variables that may affect a patient's projected probability of developing resistance. The most significant predictors found while making predictions for antibiotic resistance in postoperative cultures in the total study population were: *Abdominal surgery, the presence of an earlier resistant culture, the use of prophylactic antibiotics during the procedure, laparotomic surgery*, *and the number of expected days of pre-operative admission*. These features overlap with risk factors for antibiotic resistance reported in earlier literature [34-36].

4.3 Interpretations

The AUROCs from the models concerning the sub-populations based on respiratory cultures, wound cultures and cultures containing *P. aeruginosa* were higher than the AUROCs from the models concerning the total study population. This could be interpreted as that those sub-datasets contained more homogenous samples, enabling the capture of the more subtle patterns in the selected data. The fact that developing prediction models specifically for these sub-populations seem to improve the predictions could be interpreted as the gram-negative infections corresponding to these cultures being more predictable than those in the postoperative infections in general. However, an important sidenote would be that the standard deviations found in AUROCs for the validation folds for these sub-datasets are quite large (2-3 times larger than those for the complete dataset). The high standard deviation indicates a lack of robustness within the models and might be caused by the (smaller) sample sizes.

The features that have been found to be most important in this research are known risk factors and relate directly or indirectly to the patient's health status. Resistance to antibiotics is a known risk in patients that must stay in the hospital for a long period. Abdominal surgery, especially laparotomic surgery, is demanding for patients and often result in long hospital stay. The presence of previous resistant cultures was consistently identified as a key factor in our predictions, both for the total study population and for all sub-populations. It's worth noting that clinical practice also confirms the importance of this factor, as it is currently already a strong indicator for deciding if a patient needs broad-spectrum antibiotics. Importantly, the developed prediction models combine the collective strength of the included features in multivariate predictions, illustrating that their impact together is greater than any single feature on its own.

4.4 Strengths & Limitations

The current study had a number of strengths. First, additionally to developing ML-based models to predict antibiotic resistance in postoperative cultures in general, also were models developed specifically trained on sub-populations regarding specific culture sources or found bacteria. Comparing these models' performances provides information on how and when ML-based models might help clinicians in the best viable way. Second, to avoid the limitations of each specific classifier technique, three independent techniques with various underlying prediction methods were utilized and compared rather than depending just on one classifier technique. Third, compared to previous studies, a sizable number of positive (resistant) instances were included (1582 in the complete dataset versus other studies ranging from 177 to 1804) [24], enabling the models to capture underlying patterns and correlations in the data. Fourth, in addition to model development, model explainability was also investigated using SHAP values. An ML-based prediction model is less of a "black box" and more open about the features that affect each anticipated result when these SHAP values are shown. By presenting SHAP results, it is possible to evaluate potential biases and investigate what features correlate with the predicted outcome. This is likely to strengthen the clinicians' confidence in the model, increasing the chances of successful implementation into clinical practice.

There were also some limitations to this study. First, predictions on antibiotic resistance have been made per surgical procedure. Because the way the PERISCOPE database was set up, it was only possible to connect patient data from the EHR to procedures. This also resulted in the fact that all features, except for one, have been determined based on information at the time of the procedure. The only exception is the *culture source* feature, as it is determined from the culture's description. As a *postoperative infection* is defined in this study as any positive culture within 30 days of surgery, the information used to make the predictions could be up to 30 days old. Further research should be done to determine the performance of predictions made

with EHR information up to date to the moment of the culture. An extra benefit of this would be that it enables the addition of possibly relevant features, such as (volatile) blood levels that are included in other studies [37, 38].

A second limitation would be that, as mentioned in the methods section, it was unfeasible to retrospectively determine accurately which patients actually experienced an infection. Therefore, it was decided to define the chosen outcome as the presence of a positive culture. To minimize the effect of accidentally including contaminated or solely colonized cultures, it was also required that an antibiogram was determined on the positive culture. In practice, the combination of a positive culture with a determined antibiogram, typically only occurs when the clinician suspects an infection that needs pharmacological treatment.

A third limitation would be the quality of the EHR data which was used to extract the features for the classifiers. Some potential features concerning the patients' medical history were not included in this study because of the amount of missing data. Furthermore, as seen in *Figure 5*, the one-hot encoded categorial features describing the missing culture sources and admission sources even ended up in the top 10 most important features. Nonetheless, the lack of quality in medical data is not unique to this study, and is described more often as a limitation within literature regarding using artificial intelligence in the healthcare sector [39-41].

A last mentionable limitation would be that the Events Per Variable (EPV) ratio might be considered low in the smaller sub-datasets used in this study, thus increasing the risk of overfitting while training the models. Earlier research has shown that to minimize the effect of overfitting, an EPV of at least 10-20 would be advised [42]. To overcome the risk of overfitting in this study, the robustness of each model was also evaluated on the trainset using cross-validation.

4.5 Clinical implications

Accurate resistance predictions from ML-based models could help clinicians in determining a fitting antibiotic therapy in the timespan before the antibiogram results are available from the lab. At this point, the performances of the models developed in this study (AUROCs 0.63-0.79) need to be improved before the classifiers might be considered for use in the clinical practice. Literature is not decisive in at what performance would actually be sufficient for implementation, and for this reason some researchers suggest to look further than AUROC and focus more on achieved net benefit [43]. Further research could look into performing a net benefit analysis, researching at what predicted probabilities clinicians would change their antibiotic therapeutic strategy [44, 45].

Training models specifically on certain sub-populations might be expected to increase the performance in some cases, but it will always come with a trade-off. Not only becomes it more challenging to find enough data to train on, but training a model focused on a specific type of infection by definition reduces the usability of the model, as it can only be used in situations where those particular infections are clinically suspected. Furthermore, when training models on infections caused by specific pathogens, the potential time savings is reduced further as the predictions can only be made at the time that the lab reports what pathogen is causing the infection. These implications should be considered while conducting further research on the topic of predicting antibiotic resistance in infections. With the current findings, it cannot be conclusively demonstrated that limiting the models' application to particular postoperative infections enhances their performance.

It is crucial to remember that creating a high-performing prediction model is merely the first step in deploying the prediction models. The prediction models must first go through a number of stages of validation, impact assessment, and software implementation before they are made clinically relevant. Even at that time, more research is required to determine the impact of using an antibiotic prescription behaviour resistance risk prediction score before the use of such models is justified in routine clinical practice [46].

4.6 Conclusion

To conclude, in the current study ML-based prediction models have been built to predict antibiotic resistance to 2nd and 3rd generation cephalosporines, aminoglycosides and carbapenems in postoperative cultures both in the general patient-population and in specific cases, with the aim of providing guidance for clinicians in determining the adequate antibiotic therapy before the antibiogram results are available. The performances achieved in this study with AUROCs ranging between 0.63-0.79, are not sufficient for the models to be considered for implementation in clinical practice. In addition, with the current results it cannot be decisively concluded that narrowing the scope of the models to specific postoperative infections improve the predictions' performances.

Further research on extracting more relevant features, with increased quality, documented at time of taking the culture is necessary to further explore the possible added clinical value of using ML prediction models in the context of antibiotic resistance.

References

1. Yelin, I., et al., *Personal clinical history predicts antibiotic resistance of urinary tract infections.* Nat Med, 2019. **25**(7): p. 1143-1152.
2. Antimicrobial Resistance, C., *Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis.* Lancet, 2022. **399**(10325): p. 629-655.
3. *GLOBAL ACTION PLAN ON ANTIMICROBIAL RESISTANCE*, W.H. Organization, Editor. 2015, World Health Organization: who.int.
4. *Antimicrobial resistance*. 2021 [cited 2023 February 23rd]; Available from: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.
5. O'Neill, J., *Tackling drug-resistant infections globally: final report and recommendations*. 2016: Government of the United Kingdom.
6. Prestinaci, F., P. Pezzotti, and A. Pantosti, *Antimicrobial resistance: a global multifaceted phenomenon.* Pathog Glob Health, 2015. **109**(7): p. 309-18.
7. Solomon, S.L. and K.B. Oliver, *Antibiotic resistance threats in the United States: stepping back from the brink.* Am Fam Physician, 2014. **89**(12): p. 938-41.
8. Cosgrove, S.E., *The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs.* Clin Infect Dis, 2006. **42 Suppl 2**: p. S82-9.
9. Nathwani, D., et al., *Clinical and economic consequences of hospital-acquired resistant and multidrug-resistant Pseudomonas aeruginosa infections: a systematic review and meta-analysis.* Antimicrob Resist Infect Control, 2014. **3**(1): p. 32.
10. Naylor, N.R., et al., *Estimating the burden of antimicrobial resistance: a systematic literature review.* Antimicrob Resist Infect Control, 2018. **7**: p. 58.
11. Peleg, A.Y. and D.C. Hooper, *Hospital-acquired infections due to gram-negative bacteria.* N Engl J Med, 2010. **362**(19): p. 1804-13.
12. R. Moehring, D.J.A., *Gram-negative bacillary bacteremia in adults*. 2023, uptodate: uptodate.com.
13. Goodman, K.E., et al., *A Clinical Decision Tree to Predict Whether a Bacteremic Patient Is Infected With an Extended-Spectrum beta-Lactamase-Producing Organism.* Clin Infect Dis, 2016. **63**(7): p. 896-903.
14. John, E.B., D. Raphael, and J.B. Martin, *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Eighth Edition)*. 2015, W.B. Saunders: Philadelphia. p. I1-I120.
15. Kollef, M.H., et al., *Timing of antibiotic therapy in the ICU.* Crit Care, 2021. **25**(1): p. 360.
16. Peiffer-Smadja, N., et al., *Machine learning for clinical decision support in infectious diseases: a narrative review of current applications.* Clin Microbiol Infect, 2020. **26**(5): p. 584-595.
17. Burkov, A., *TheHundred Page Machine Learning Book*. Vol. 1. 2019: Andriy Burkov. 160.
18. Gerontini, M., et al. *Predictions in antibiotics resistance and nosocomial infections monitoring*. in *2011 24th International Symposium on Computer-Based Medical Systems (CBMS)*. 2011.
19. Mohd Sazlly Lim, S., et al., *Clinical prediction models for ESBL-Enterobacteriaceae colonization or infection: a systematic review.* J Hosp Infect, 2019. **102**(1): p. 8-16.
20. Oonsivilai, M., et al., *Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia.* Wellcome Open Res, 2018. **3**: p. 131.
21. Engel, F.J.N.e.a., *[Unpublished] Antibiotic Resistance prediction in hospitalized patients, a Systematic Review*. 2023, LUMC.
22. Koek, M.B.G., et al., *Burden of surgical site infections in the Netherlands: cost analyses and disability-adjusted life years.* J Hosp Infect, 2019. **103**(3): p. 293-302.
23. Sattar, F., et al., *Frequency of Post-operative Surgical Site Infections in a Tertiary Care Hospital in Abbottabad, Pakistan.* Cureus, 2019. **11**(3): p. e4243.
24. Collins, G.S., et al., *Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement.* BMJ, 2015. **350**: p. g7594.
25. Yusuf, E., M. Zeitlinger, and S. Meylan, *A narrative review of the intermediate category of the antimicrobial susceptibility test: relation with dosing and possible impact on antimicrobial stewardship.* J Antimicrob Chemother, 2023. **78**(2): p. 338-345.
26. Couronne, R., P. Probst, and A.L. Boulesteix, *Random forest versus logistic regression: a large-scale benchmark experiment.* BMC Bioinformatics, 2018. **19**(1): p. 270.
27. Fernández-Delgado, M., et al., *Do we need hundreds of classifiers to solve real world classification problems?* J. Mach. Learn. Res., 2014. **15**(1): p. 3133–3181.
28. Zhang, Y. *Support Vector Machine Classification Algorithm and Its Application*. in *Information Computing and Applications*. 2012. Berlin, Heidelberg: Springer Berlin Heidelberg.
29. Simon, C. and T. Philip, *Chapter 5 - Continuous Model Selection for Large-Scale Recommender Systems*, in *Big Data Analytics*, G. Venu, V.R. Vijay, and C.R. Rao, Editors. 2015, Elsevier. p. 107-124.
30. Snoek, J., H. Larochelle, and R.P. Adams, *Practical Bayesian optimization of machine learning algorithms*, in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2*. 2012, Curran Associates Inc.: Lake Tahoe, Nevada. p. 2951–2959.
31. Marcia, L.B., G. Kai, and M.P.H. Elsa, *Relation between prognostics predictor evaluation metrics and local interpretability SHAP values.* Artificial Intelligence, 2022. **306**: p. 103667.
32. Hosmer, D.W., S. Lemeshow, and R.X. Sturdivant, *Applied Logistic Regression*. 2013: Wiley.
33. Tang, R., et al., *Machine learning in predicting antimicrobial resistance: a systematic review and meta-analysis.* Int J Antimicrob Agents, 2022. **60**(5-6): p. 106684.
34. Low, M., et al., *Association between urinary community-acquired fluoroquinolone-resistant Escherichia coli and neighbourhood antibiotic consumption: a population-based case-control study.* Lancet Infect Dis, 2019. **19**(4): p. 419-428.
35. Richter, S.E., et al., *Risk factors for development of aminoglycoside resistance among gram-negative rods.* Am J Health Syst Pharm, 2019. **76**(22): p. 1838-1847.

36. Richter, S.E., et al., *Risk Factors for Development of Carbapenem Resistance Among Gram-Negative Rods.* Open Forum Infect Dis, 2019. **6**(3): p. ofz027.
37. Lewin-Epstein, O., et al., *Predicting Antibiotic Resistance in Hospitalized Patients by Applying Machine Learning to Electronic Medical Records.* Clin Infect Dis, 2021. **72**(11): p. e848-e855.
38. Vazquez-Guillamet, M.C., et al., *Predicting Resistance to Piperacillin-Tazobactam, Cefepime and Meropenem in Septic Patients With Bloodstream Infection Due to Gram-Negative Bacteria.* Clin Infect Dis, 2017. **65**(10): p. 1607-1614.
39. Holmes, J.H., et al., *Why Is the Electronic Health Record So Challenging for Research and Clinical Care?* Methods Inf Med, 2021. **60**(1-02): p. 32-48.
40. Wu, J., J. Roy, and W.F. Stewart, *Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches.* Med Care, 2010. **48**(6 Suppl): p. S106-13.
41. Xiao, C., E. Choi, and J. Sun, *Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review.* J Am Med Inform Assoc, 2018. **25**(10): p. 1419-1428.
42. Austin, P.C. and E.W. Steyerberg, *Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models.* Stat Methods Med Res, 2017. **26**(2): p. 796-808.
43. de Hond, A.A.H., E.W. Steyerberg, and B. van Calster, *Interpreting area under the receiver operating characteristic curve.* Lancet Digit Health, 2022. **4**(12): p. e853-e855.
44. Vickers, A.J., B. Van Calster, and E.W. Steyerberg, *Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests.* BMJ, 2016. **352**: p. i6.
45. Vickers, A.J., B. van Calster, and E.W. Steyerberg, *A simple, step-by-step guide to interpreting decision curve analysis.* Diagn Progn Res, 2019. **3**: p. 18.
46. de Hond, A.A.H., et al., *Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review.* NPJ Digit Med, 2022. **5**(1): p. 2.

Supplementary materials

Antibiotics of interest

This study addresses resistance to carbapenems, aminoglycosides, second- and third generation cephalosporins (respectively 2GC and 3GC). These antibiotic groups consist of the following antibiotics[1,2]:

- 2GC: cefaclor, cefuroxime, and cefuroxime axetil.
- 3GC: ceftazidime(/avibactam), ceftibuten, cefotaxime, ceftriaxone
- Aminoglycosides: tobramycin, gentamycin, and amikacin
- Carbapenems: ertapenem, imipenem, meropenem.

Guidelines antibiotic susceptibility testing

The Leiden University Medical Centre (LUMC) managed culture and susceptibility tests in accordance with regulatory standards. Antibiotic susceptibility testing in this approach follows a positive culture, with a few exceptions. One example of such an exception is when multiple cultures of the same bacteria were taken from the same material of a single patient. In that case only one isolate must be assessed for susceptibility. Susceptibility tests may also be bypassed if a prior susceptibility profile based on earlier cultures was available and cultures collected from the same patient within a two-day period revealed the same bacteria. Finally, susceptibility testing is not necessary for isolates from non-human species or autopsies.

Disk diffusion, E-test, and/or Vitek techniques were used for the antibiotic susceptibility testing.
The antibiotic panels were chosen, and the findings were interpreted, in accordance with the guidelines of the European Committee on Antimicrobial Susceptibility Testing (EUCAST)[3]. These guidelines define a bacterium as susceptible (S) if it responds to a typical dose of antibiotics and as resistant (R) if treatment failure was highly expected, even with increased dosages. Infections might have been classified as intermediate susceptible (I) before January 1st, 2019, if the likelihood of treatment success with the appropriate antibiotic was questionable or ambiguous. Infections are only classified as intermediate susceptible as of January 1st, 2019, if susceptibility is seen, only with greater dosages of antibiotics.

[1] WHO Collaborating Centre for Drug Statistics Methodology., *Guidelines for ATC classification and DDD assignment 2023*. 2022: Oslo, Norway.
[2] Boomkamp, M.D., et al. *Farmacotherapeutisch Kompas*. Available from: https://www.farmacotherapeutischkompas.nl/.
[3] European Committee on Antimicrobial Susceptibility Testing, E. *Area of Technical Uncertainty (ATU) in antimicrobial susceptibility testing*. 2020 1 June 2020.

Description of available features

This supplement is confidential. Contact the authors for access.

Methods for feature extraction

Admission source

1. Query data from the LUMC Data platform, specifically from the Encounter database, containing information about the admissions' dates and corresponding sources for all patients.
2. Using the queried data, match the procedures included in this study to the registered admission sources. -> Use the admission date least prior to the procedure date
3. Create an overview of the different options are registered as admission source
4. Create a categorical feature in which the different options are compiled to prevent noise or potential overfitting:
    - Residential environment: admission from home or send by GP or outpatient clinic
    - Transferred: admission from other hospital or healthcare facility
    - Unknown: admission source unknown due to missing data or ambiguity in registered data

Previous admission

1. Query data from the LUMC Data platform, specifically from the Encounter database, containing information about the discharge dates for all patients.
2. Using the queried data, find out for the procedures included in this study if the corresponding patient had a discharge date within six months before the procedure date.
3. Create a binary feature:
    - 1: if discharge date was found
    - 0: if not

Previous ICU admission

1. Query data from the LUMC Data platform, specifically from the Encounter database, containing information about the encounters that took place at the ICU.
2. Using the queried data, find out for the procedures included in this study if an encounter at the ICU took place within six months before the procedure date.
3. Create a binary feature:
    - 1: if ICU encounter was found
    - 0: if not

Previous Indwelling devices & Indwelling devices after the procedure

1. Query data from the LUMC Data platform, specifically from the Devices database, containing information on the registered devices and the time of usage.
2. Classify the devices that are considered to be indwelling:
    - Urine catheters
    - Drains
    - Central venous catheters
    - Tubes used for mechanical ventilation
3. Using the queried and classified data, create a binary feature for if the use of an indwelling device was registered in the six months before the procedure
    - 1: if indwelling device was used within 6 months before the procedure
    - 0: if not
4. Using the queried and classified data, create a binary feature for if the use of an indwelling device was registered on the day of the procedure
    - 1: if indwelling device was used on day of procedure
    - 0: if not

Previous resistant pathogen

1. Label procedures with the same methods as being used for the regular labelling (see code screenshots)
2. Except the timeframe of interest is this time within 6 months before the procedure instead of 30 days after the procedure.
3. Create a binary feature:
    - 1: if resistant pathogen was found
    - 0: if not

Hyperparameter Optimization

The table below provides an overview of the examined hyperparameter settings that were used to find the optimal configuration for the three classifiers. As mentioned in the main part of the thesis. The hyperparameter tuning was done with grid search for the logistic regression model, and with Bayesian optimization for the other models because of computational limitations. Because the input for the optimization was a pipeline, it was possible to include different feature selection methods as well.

Classifier	Hyperparameter	Examined settings	Final configuration
LR	Type of regularization (penalty)	L1, L2, elasticnet, none	Elastic-Net
	If penalty = 'elasticnet': The Elastic-Net mixing parameter between L1 & L2 (L1 ratio)	0.1, 0.3, 0.5, 0.7, 0.9	0.3
	Algorithm used in optimization (solver) [1]	Liblinear, newton-cg, lbfgs, sag, s aga	saga
	Regularization strength (C-value)	0.001, 0.01, 0.1, 1, 10	0.01
	Maximum number of iterations to converge (max_iter)	50, 100, 400, 700, 1000, 2000	400
RF	Feature selection method [2]	None, RFECV, SelectFromModel	RFECV
	Number of trees (n_estimators)	20, 40, 60, 80	80
	Maximum depth of tree (max_depth)	10, 15, 20	20
	Maximum number of features to consider when looking for the best split (max_features)	sqrt, log2	Log2
	Minimum numbers of samples required to split an internal node (min_samples_split)	5, 15, 25, 35	25
SVM	Feature selection method [2]	None, RFECV, SelectFromModel	RFECV
	Regularization strength (C-value)	0.001, 0.01, 0.1, 1, 10	0.001
	Type of function used in algorithm (kernel	linear, rbf, poly, sigmoid	linear
	If kernel = 'poly': degree of polynomial kernel function (degree)	1, 2, 3, 4	n/a
	If kernel = 'poly', 'rbf' or 'sigmoid': kernel coefficient (gamma)	1, 0.1, 0.01, 0.001	n/a

Abbreviations: LR: logistic regression, RF: random forest, RFECV: recursive feature elimination with cross-validation, SVM: support vector machine

[1] Not all penalties are supported by each solver. For more details see: scikit-learn.org

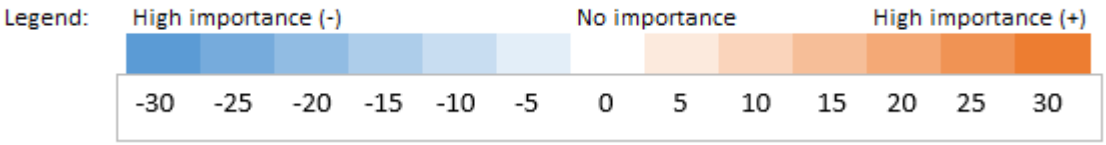
[2] The application of a feature selection method is not a hyperparameter, but as multiple methods were examined in combination with the hyperparameters, the method chosen is shown in this table.

Methods for scoring feature importance in heatmap

SHAP values were determined for all 27 developed models and the 10 feature with the highest importance were plotted. The features appearing in the top 10 of any of the models were scored and summed over the three models per corresponding (sub-)population. The most importance feature per model got assigned 10 points, the second most importance feature 9 points, etc. This means that every feature could get a maximum score of 30 points per (sub-)population. These importance scores reflect the significance of each feature in predicting (or influencing) the outcome. The five features per (sub-)population with the highest scores are shown and ranked in the table below.

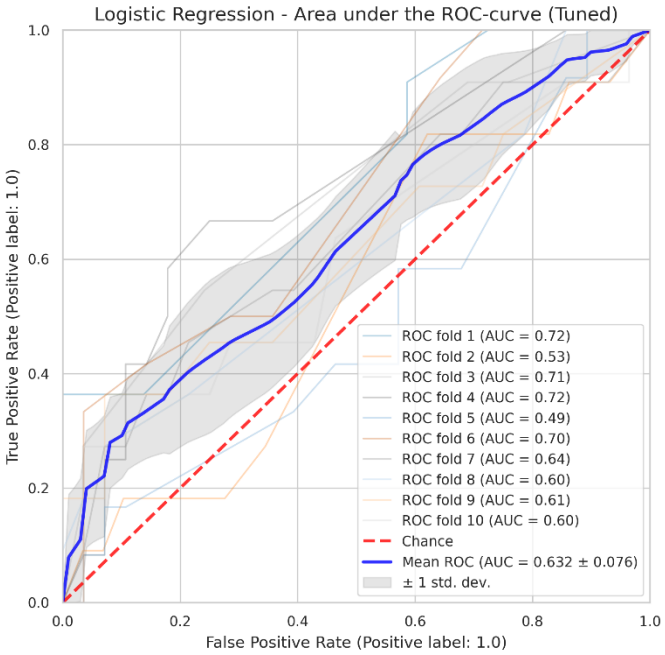
Population	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
Total study population	Area: Abdomen (+)	Previous Resistant culture (+)	Antibiotic Prophylaxis (?)	Type: Laparotomic (+)	Expected pre-op days (+)
Culture source: Ascites	Previous Resistant culture (+)	Area: Abdomen (+)	Previous Admission (+)	Procedure risk: high (+)	Medication: Antithrombotics (+)
Culture source: Blood	Previous Resistant culture (+)	Medication: diabetes (+)	Expected post-op days (+)	Gender: Male (+)	ASA score 3.0 (-)
Culture source: Respiratory	Gender: Male (-)	Previous Resistant culture (+)	Medication: Anti-arrhythmics (+)	Medication: Diuretics (+)	Medication: Ca-channel blockers (-)
Culture source: Urine	Previous Resistant culture (+)	Antibiotic Prophylaxis (+)	Indwelling device after procedure (+)	Procedure risk: high (+)	ASA score 1.0 (+)
Culture source: Wound	Area: Abdomen (+)	Difference between actual and planned surgery time (+)	Indwelling device after procedure (+)	Type: Laparotomic (+)	Procedure length (+)
Bacteria: E. coli	Previous Resistant culture (+)	Gender: Male (+)	Indwelling device after procedure (+)	Culture source: Wound (+)	Procedure length (+)
Bacteria: K. pneumoniae	Indwelling device after procedure (+)	Previous Resistant culture (+)	Medication: Immunosuppressants (+)	Culture source: unknown (-)	Antibiotic Prophylaxis (+)
Bacteria: P. Aeruginosa	Previous Resistant culture (+)	Previous ICU Admission (+)	Type: Laparotomic (+)	Medication: Diuretics (+)	Medication: Raas inhibitors (-)

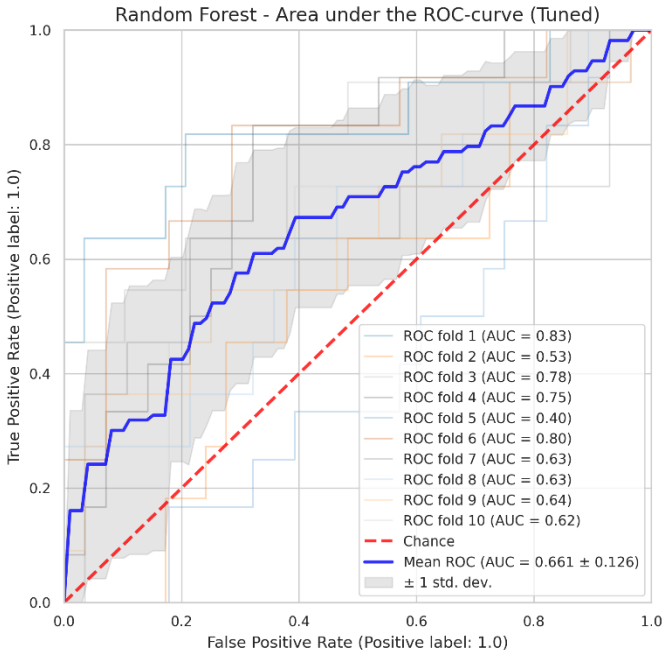
To visually represent the importance of the features appearing in the top-5 for each (sub-)population, a heatmap was generated using the created importance scores. Brackets were used to display distinct levels of importance. The scores were rounded to the next 5 and color-coded using the colours shown in the legend below. A final positive score indicates a feature positively correlates with the outcome, while a final negative score suggests a negative correlation.

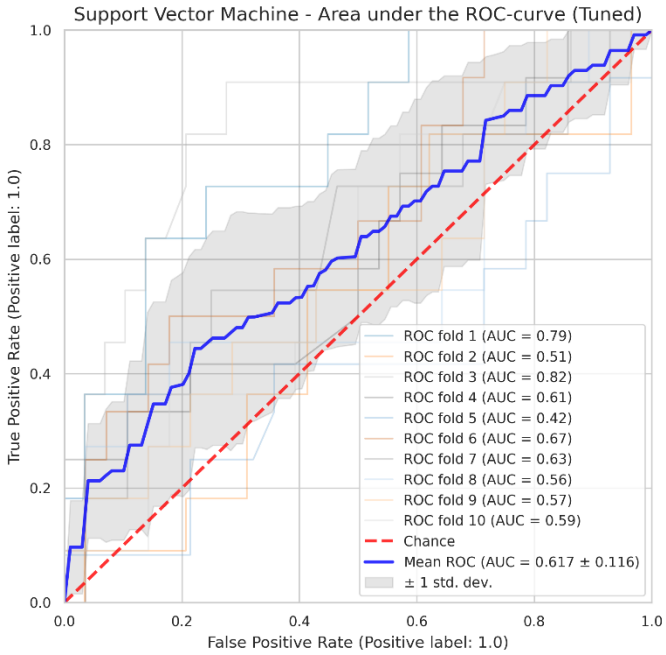


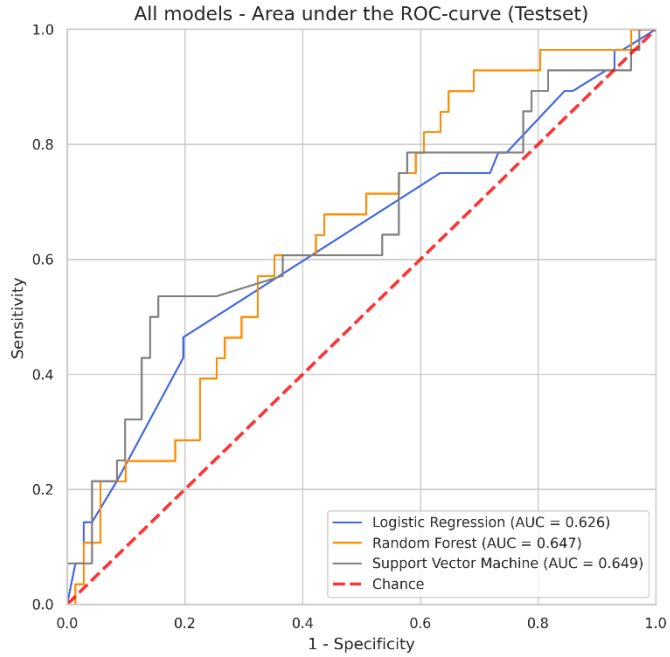
Results for model evaluation

Ascites

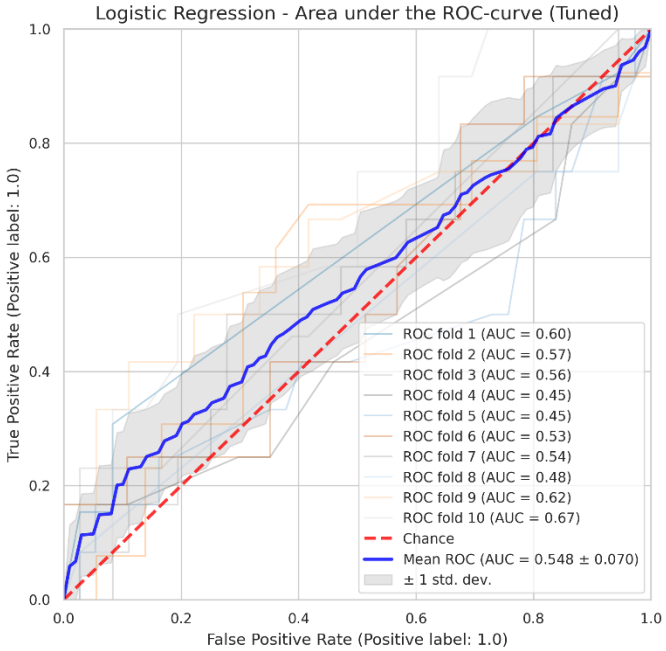


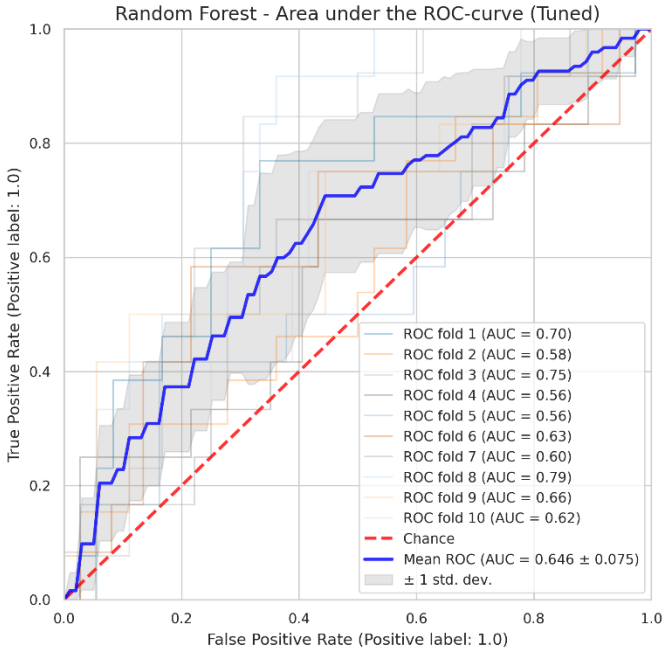


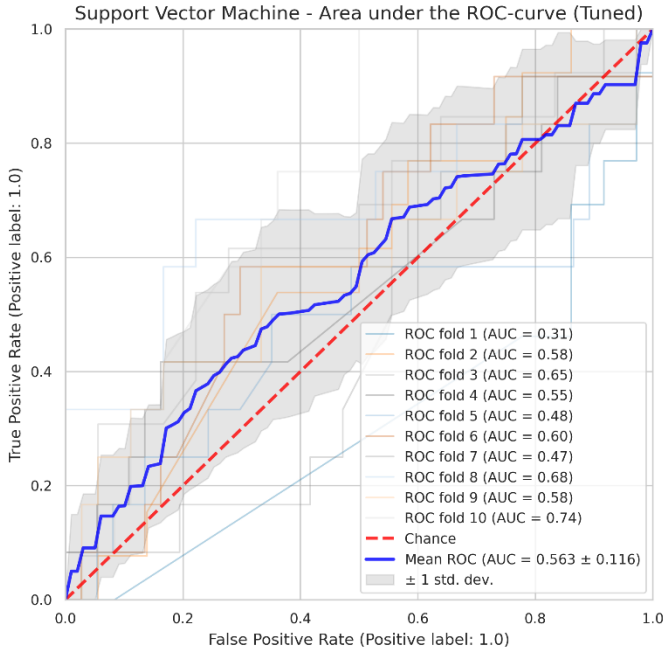


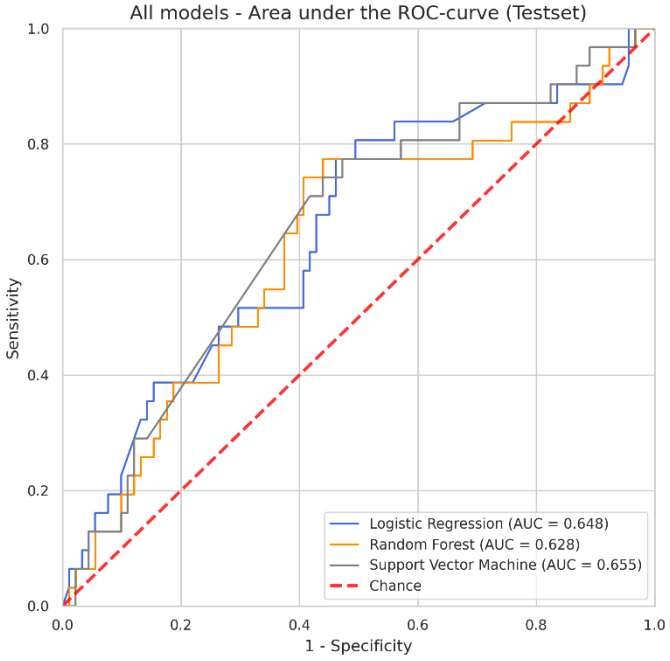


Blood

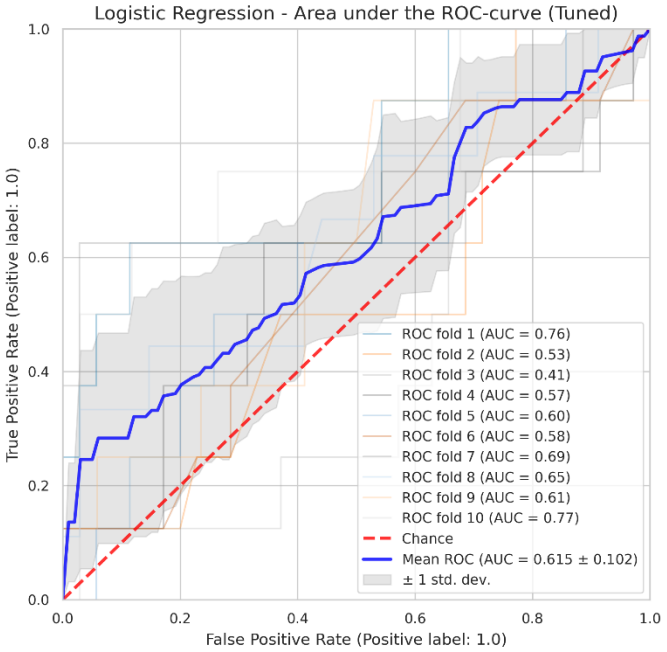

Logistic Regression - Area under the ROC-curve (Tuned)
True Positive Rate (Positive label: 1.0)
False Positive Rate (Positive label: 1.0)
ROC fold 1 (AUC = 0.60)
ROC fold 2 (AUC = 0.57)
ROC fold 3 (AUC = 0.56)
ROC fold 4 (AUC = 0.45)
ROC fold 5 (AUC = 0.45)
ROC fold 6 (AUC = 0.53)
ROC fold 7 (AUC = 0.54)
ROC fold 8 (AUC = 0.48)
ROC fold 9 (AUC = 0.62)
ROC fold 10 (AUC = 0.67)
Chance
Mean ROC (AUC = 0.548 ± 0.070)
± 1 std. dev.

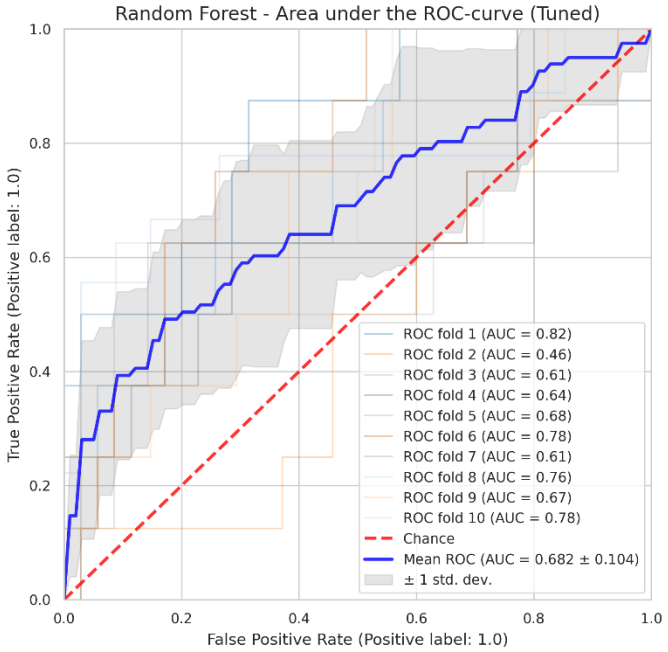

Random Forest - Area under the ROC-curve (Tuned)
True Positive Rate (Positive label: 1.0)
False Positive Rate (Positive label: 1.0)
ROC fold 1 (AUC = 0.70)
ROC fold 2 (AUC = 0.58)
ROC fold 3 (AUC = 0.75)
ROC fold 4 (AUC = 0.56)
ROC fold 5 (AUC = 0.56)
ROC fold 6 (AUC = 0.63)
ROC fold 7 (AUC = 0.60)
ROC fold 8 (AUC = 0.79)
ROC fold 9 (AUC = 0.66)
ROC fold 10 (AUC = 0.62)
Chance
Mean ROC (AUC = 0.646 ± 0.075)
± 1 std. dev.

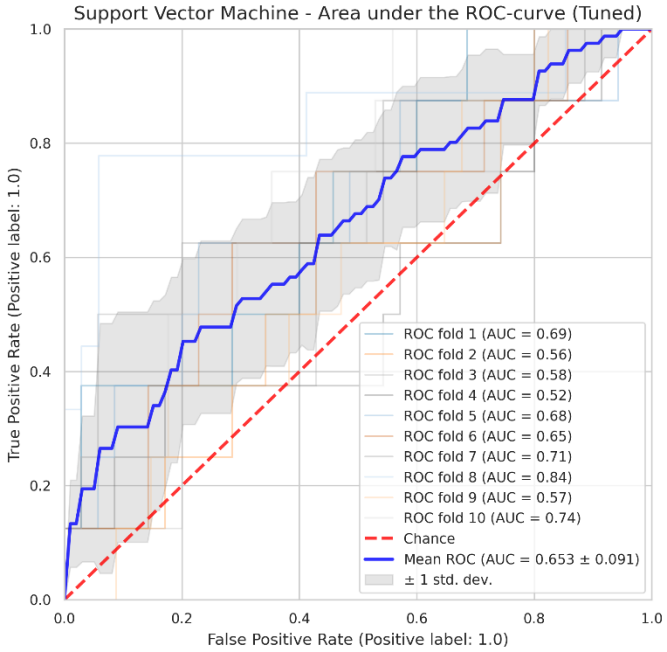

Support Vector Machine - Area under the ROC-curve (Tuned)
True Positive Rate (Positive label: 1.0)
False Positive Rate (Positive label: 1.0)
ROC fold 1 (AUC = 0.31)
ROC fold 2 (AUC = 0.58)
ROC fold 3 (AUC = 0.65)
ROC fold 4 (AUC = 0.55)
ROC fold 5 (AUC = 0.48)
ROC fold 6 (AUC = 0.60)
ROC fold 7 (AUC = 0.47)
ROC fold 8 (AUC = 0.68)
ROC fold 9 (AUC = 0.58)
ROC fold 10 (AUC = 0.74)
Chance
Mean ROC (AUC = 0.563 ± 0.116)
± 1 std. dev.

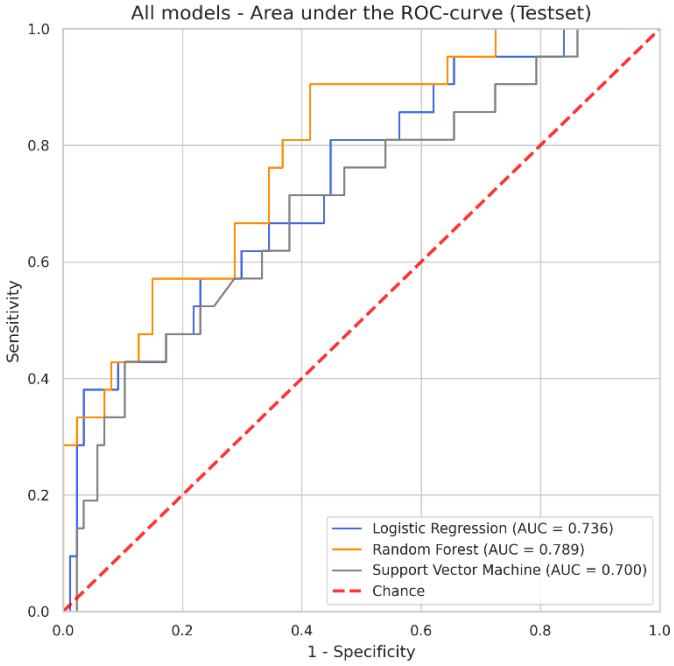

All models - Area under the ROC-curve (Testset)
Sensitivity
1 - Specificity
Logistic Regression (AUC = 0.648)
Random Forest (AUC = 0.628)
Support Vector Machine (AUC = 0.655)
Chance

Respiratory

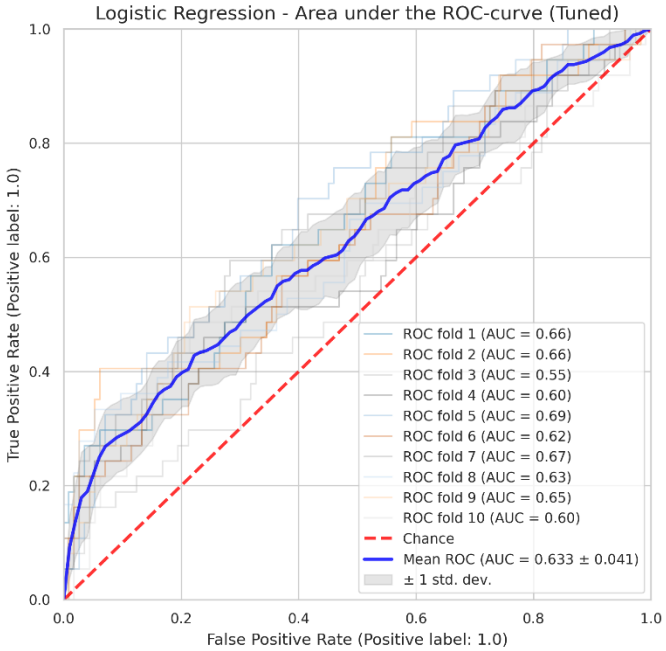


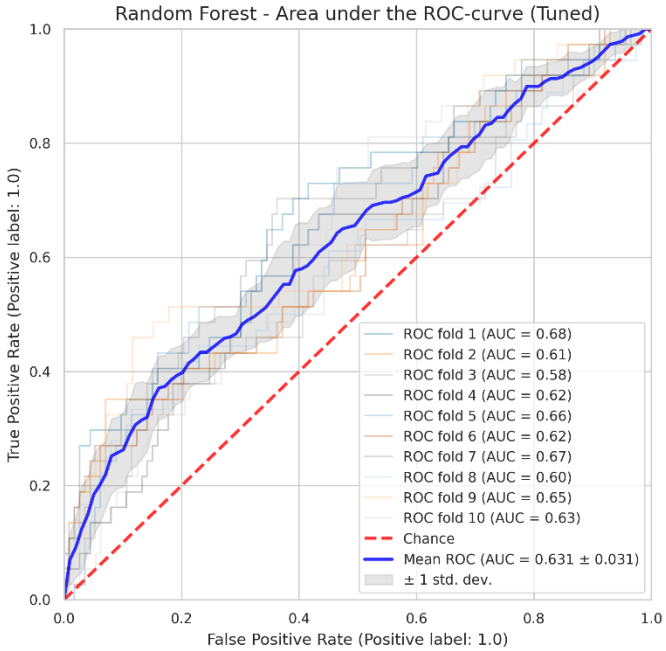


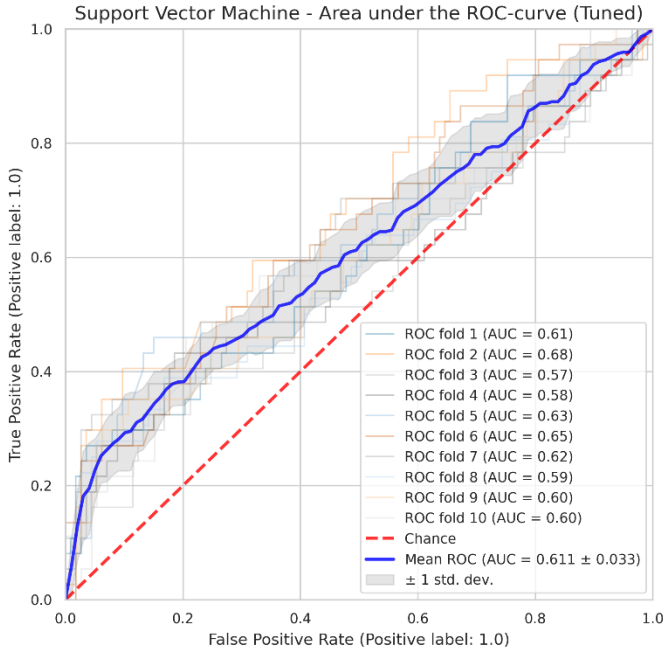


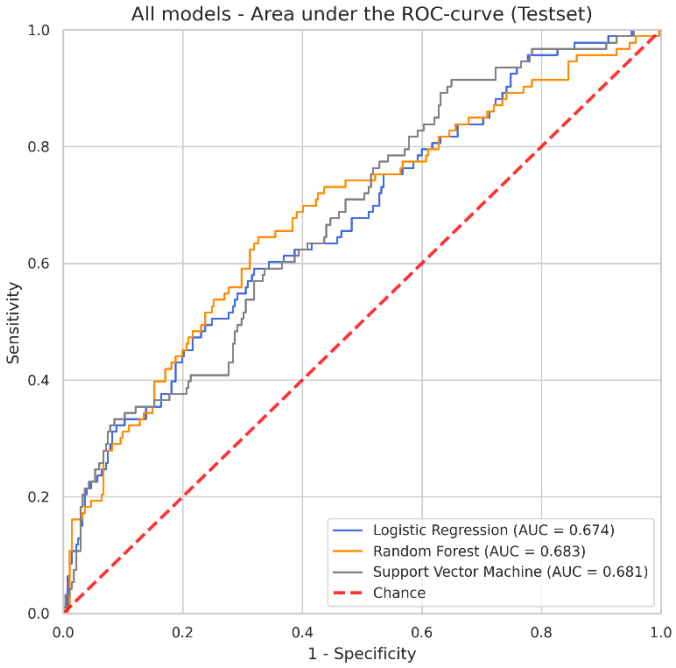


Urine

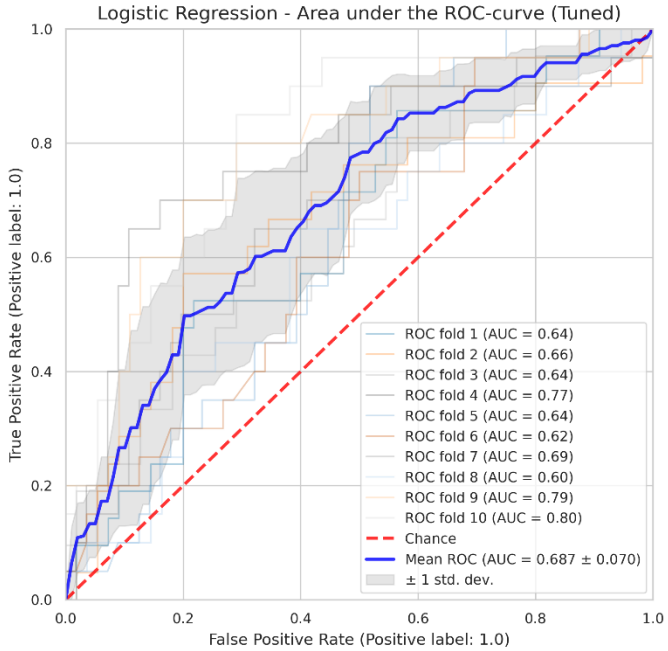

Logistic Regression - Area under the ROC-curve (Tuned)
True Positive Rate (Positive label: 1.0)
False Positive Rate (Positive label: 1.0)
ROC fold 1 (AUC = 0.66)
ROC fold 2 (AUC = 0.66)
ROC fold 3 (AUC = 0.55)
ROC fold 4 (AUC = 0.60)
ROC fold 5 (AUC = 0.69)
ROC fold 6 (AUC = 0.62)
ROC fold 7 (AUC = 0.67)
ROC fold 8 (AUC = 0.63)
ROC fold 9 (AUC = 0.65)
ROC fold 10 (AUC = 0.60)
Chance
Mean ROC (AUC = 0.633 ± 0.041)
± 1 std. dev.

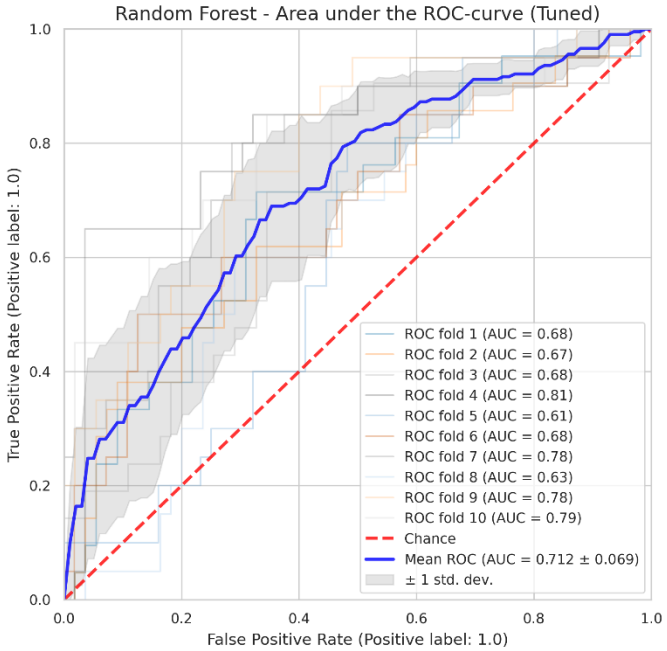

Random Forest - Area under the ROC-curve (Tuned)
True Positive Rate (Positive label: 1.0)
False Positive Rate (Positive label: 1.0)
ROC fold 1 (AUC = 0.68)
ROC fold 2 (AUC = 0.61)
ROC fold 3 (AUC = 0.58)
ROC fold 4 (AUC = 0.62)
ROC fold 5 (AUC = 0.66)
ROC fold 6 (AUC = 0.62)
ROC fold 7 (AUC = 0.67)
ROC fold 8 (AUC = 0.60)
ROC fold 9 (AUC = 0.65)
ROC fold 10 (AUC = 0.63)
Chance
Mean ROC (AUC = 0.631 ± 0.031)
± 1 std. dev.

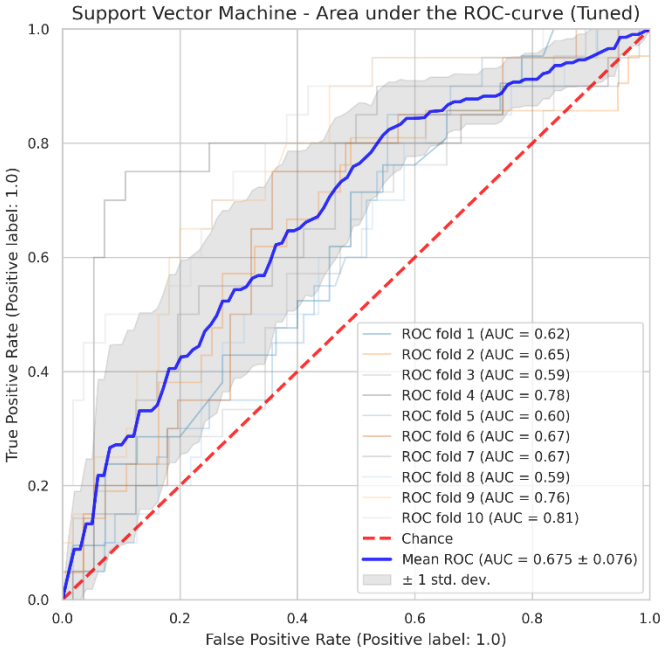

Support Vector Machine - Area under the ROC-curve (Tuned)
True Positive Rate (Positive label: 1.0)
False Positive Rate (Positive label: 1.0)
ROC fold 1 (AUC = 0.61)
ROC fold 2 (AUC = 0.68)
ROC fold 3 (AUC = 0.57)
ROC fold 4 (AUC = 0.58)
ROC fold 5 (AUC = 0.63)
ROC fold 6 (AUC = 0.65)
ROC fold 7 (AUC = 0.62)
ROC fold 8 (AUC = 0.59)
ROC fold 9 (AUC = 0.60)
ROC fold 10 (AUC = 0.60)
Chance
Mean ROC (AUC = 0.611 ± 0.033)
± 1 std. dev.

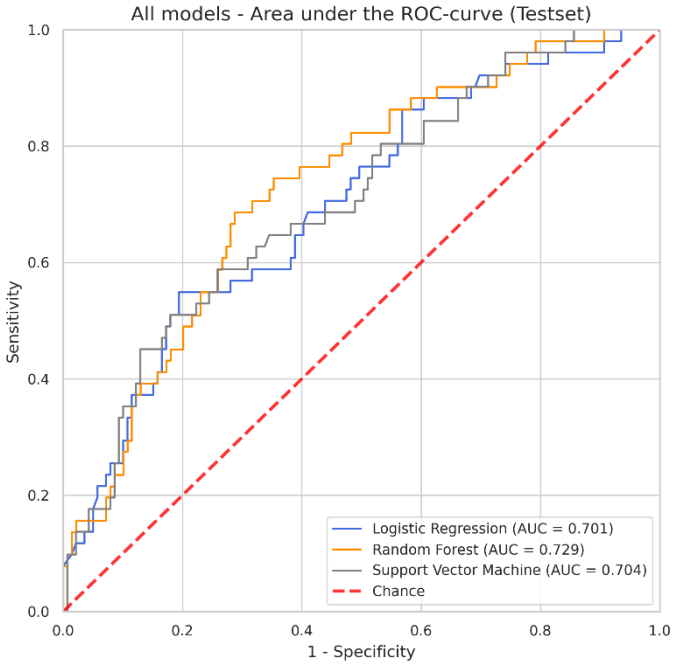

All models - Area under the ROC-curve (Testset)
Sensitivity
1 - Specificity
Logistic Regression (AUC = 0.674)
Random Forest (AUC = 0.683)
Support Vector Machine (AUC = 0.681)
Chance

Wound

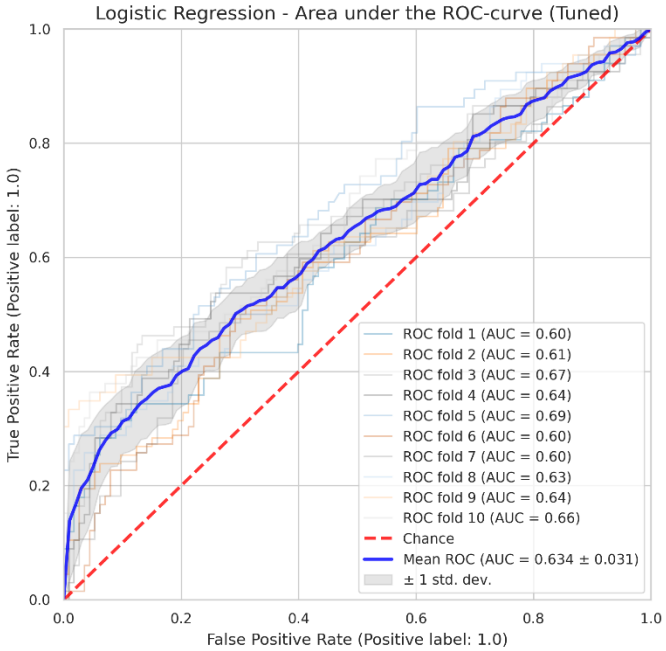

Logistic Regression - Area under the ROC-curve (Tuned)
True Positive Rate (Positive label: 1.0)
False Positive Rate (Positive label: 1.0)
ROC fold 1 (AUC = 0.64)
ROC fold 2 (AUC = 0.66)
ROC fold 3 (AUC = 0.64)
ROC fold 4 (AUC = 0.77)
ROC fold 5 (AUC = 0.64)
ROC fold 6 (AUC = 0.62)
ROC fold 7 (AUC = 0.69)
ROC fold 8 (AUC = 0.60)
ROC fold 9 (AUC = 0.79)
ROC fold 10 (AUC = 0.80)
Chance
Mean ROC (AUC = 0.687 ± 0.070)
± 1 std. dev.

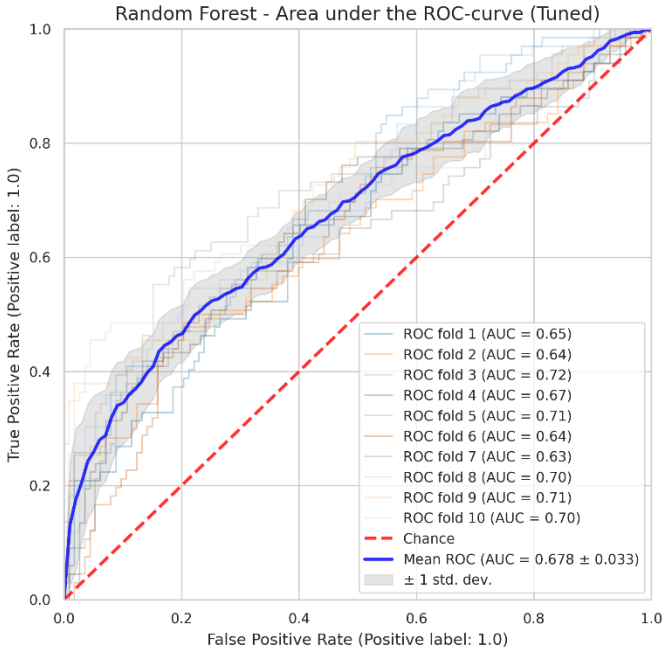

Random Forest - Area under the ROC-curve (Tuned)
True Positive Rate (Positive label: 1.0)
False Positive Rate (Positive label: 1.0)
ROC fold 1 (AUC = 0.68)
ROC fold 2 (AUC = 0.67)
ROC fold 3 (AUC = 0.68)
ROC fold 4 (AUC = 0.81)
ROC fold 5 (AUC = 0.61)
ROC fold 6 (AUC = 0.68)
ROC fold 7 (AUC = 0.78)
ROC fold 8 (AUC = 0.63)
ROC fold 9 (AUC = 0.78)
ROC fold 10 (AUC = 0.79)
Chance
Mean ROC (AUC = 0.712 ± 0.069)
± 1 std. dev.

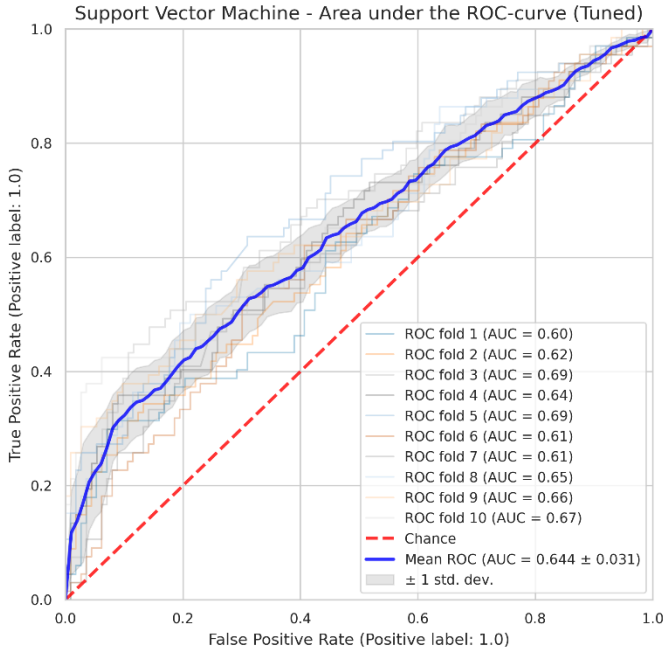

Support Vector Machine - Area under the ROC-curve (Tuned)
True Positive Rate (Positive label: 1.0)
False Positive Rate (Positive label: 1.0)
ROC fold 1 (AUC = 0.62)
ROC fold 2 (AUC = 0.65)
ROC fold 3 (AUC = 0.59)
ROC fold 4 (AUC = 0.78)
ROC fold 5 (AUC = 0.60)
ROC fold 6 (AUC = 0.67)
ROC fold 7 (AUC = 0.67)
ROC fold 8 (AUC = 0.59)
ROC fold 9 (AUC = 0.76)
ROC fold 10 (AUC = 0.81)
Chance
Mean ROC (AUC = 0.675 ± 0.076)
± 1 std. dev.

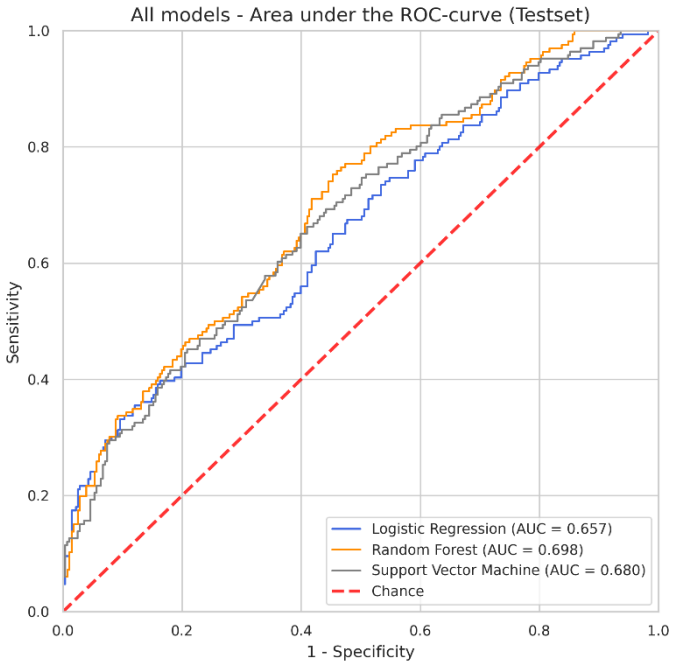

All models - Area under the ROC-curve (Testset)
Sensitivity
1 - Specificity
Logistic Regression (AUC = 0.701)
Random Forest (AUC = 0.729)
Support Vector Machine (AUC = 0.704)
Chance

Escherichia coli

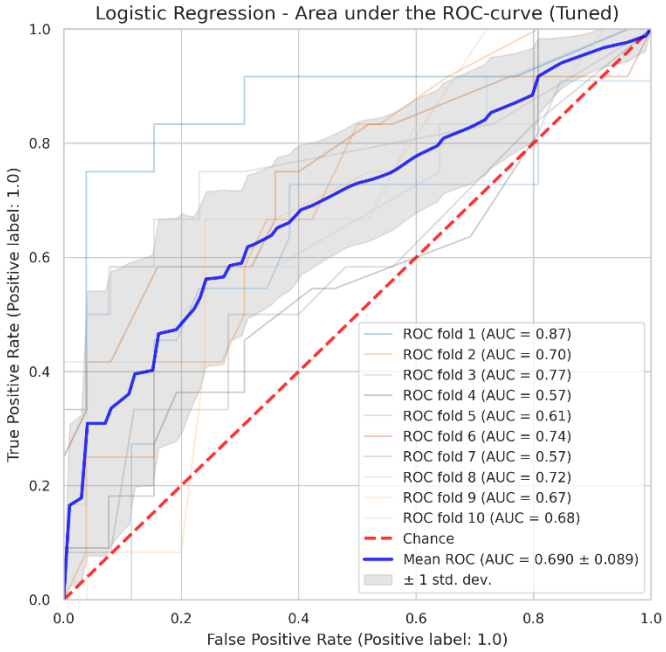


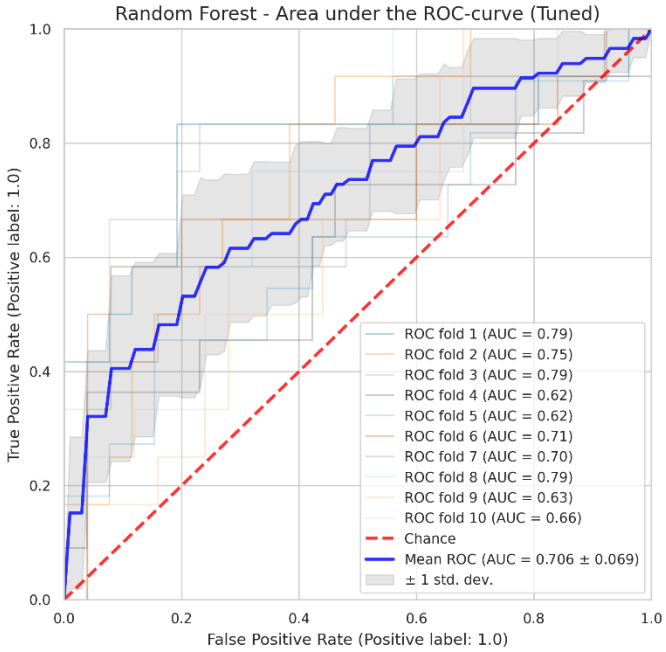


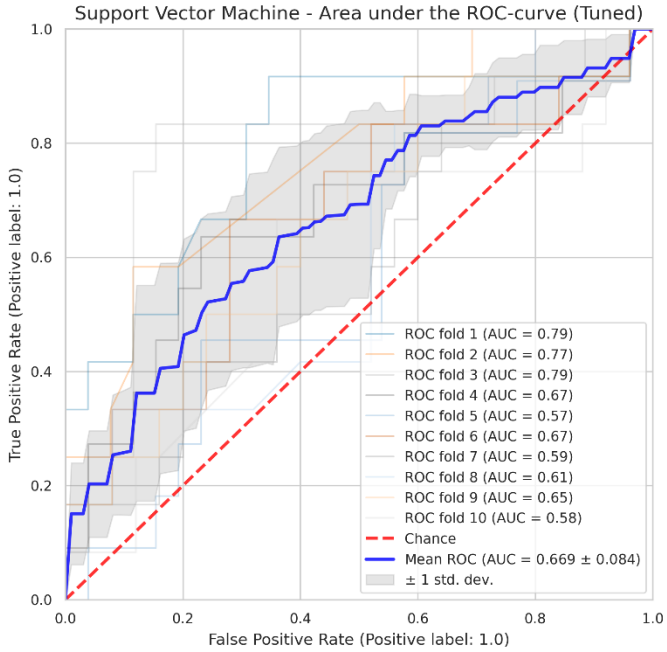


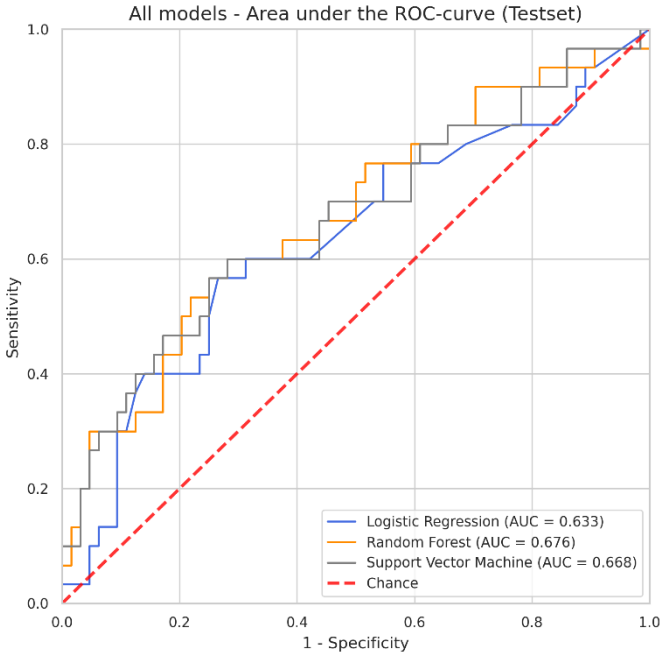


Klebsiella pneuomniae

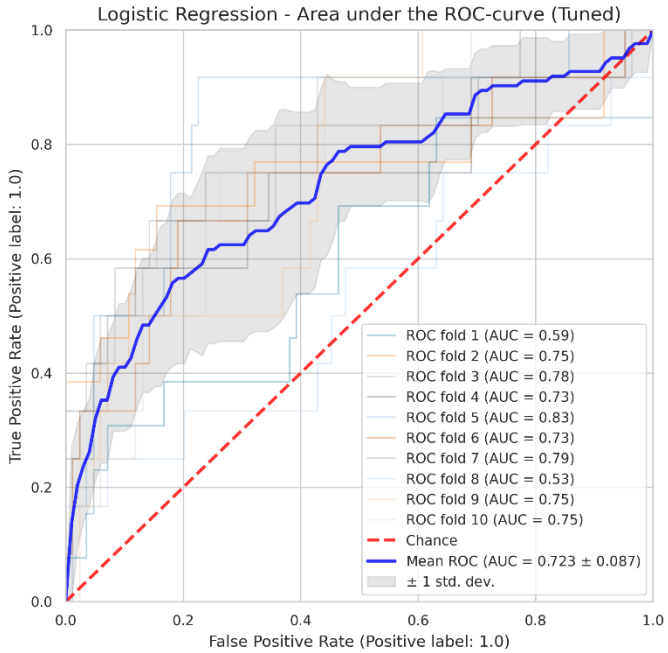


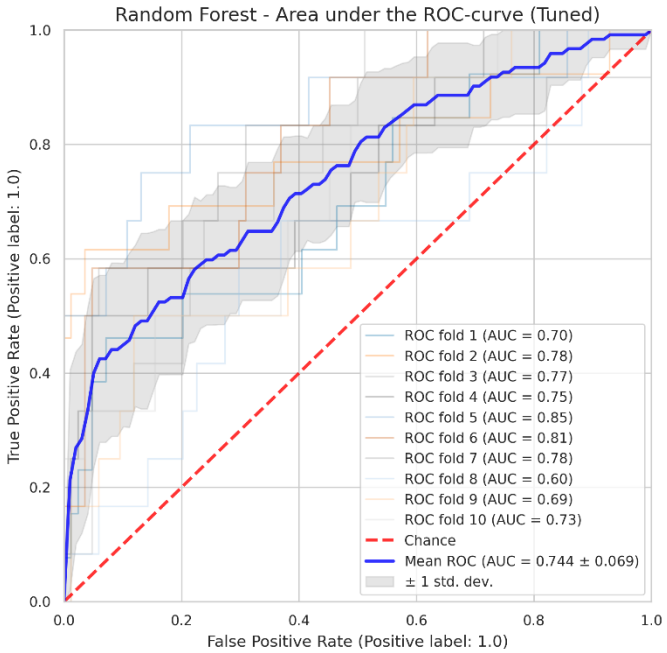


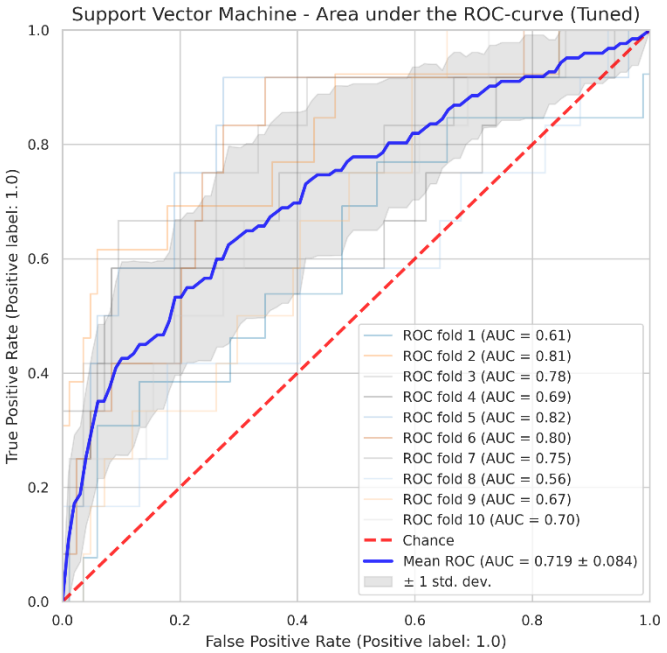


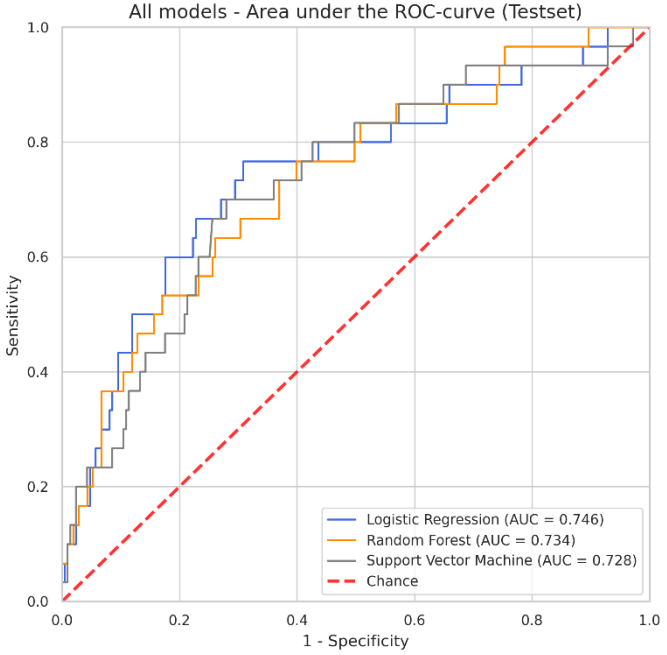


Pseudomonas aeruginosa





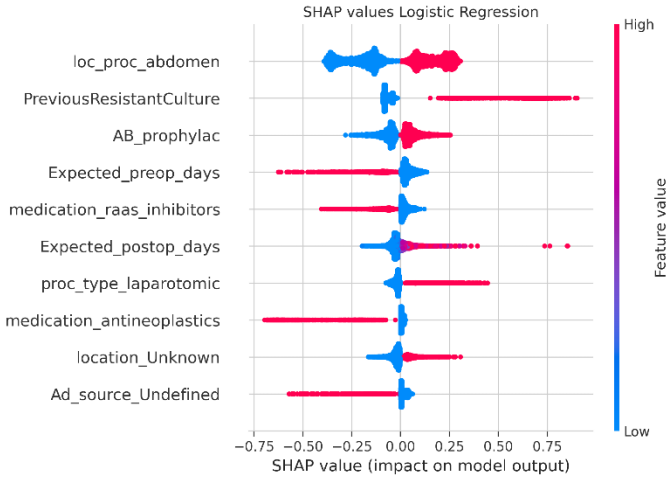


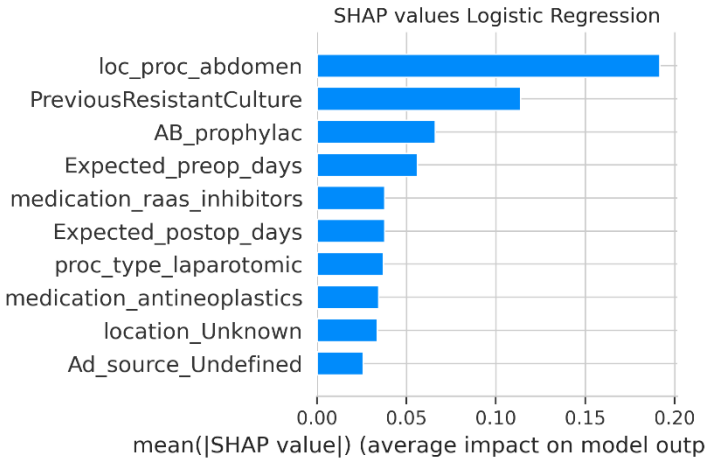


Results for feature importance

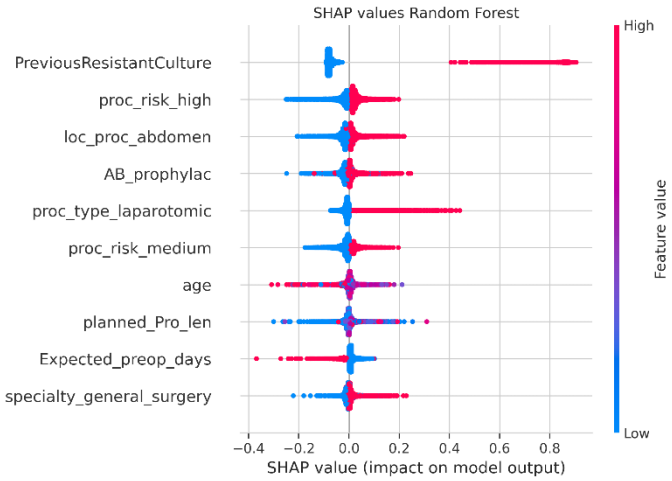
Ascites

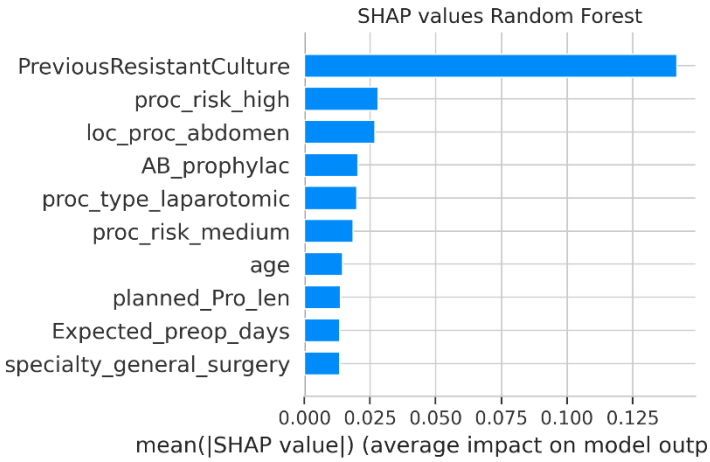
Logistic Regression



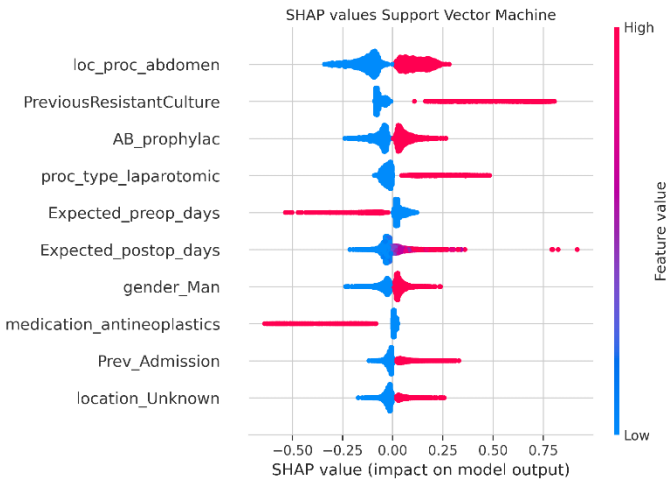


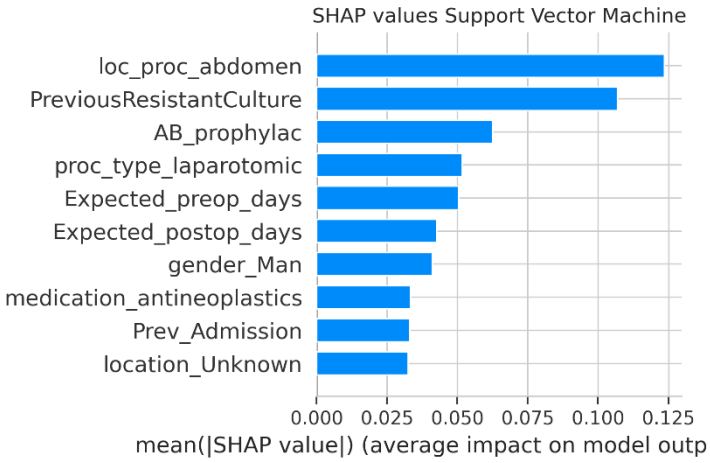
Random Forest





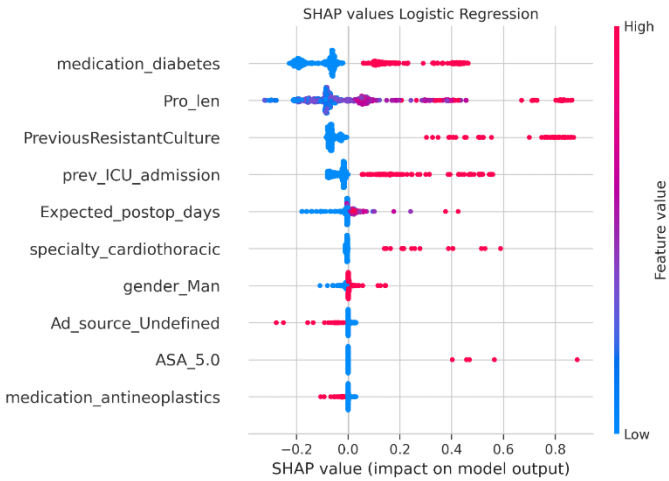
Support Vector Machine

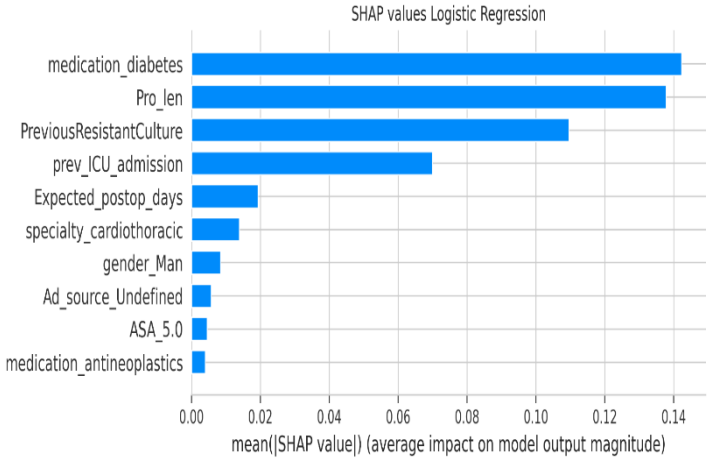




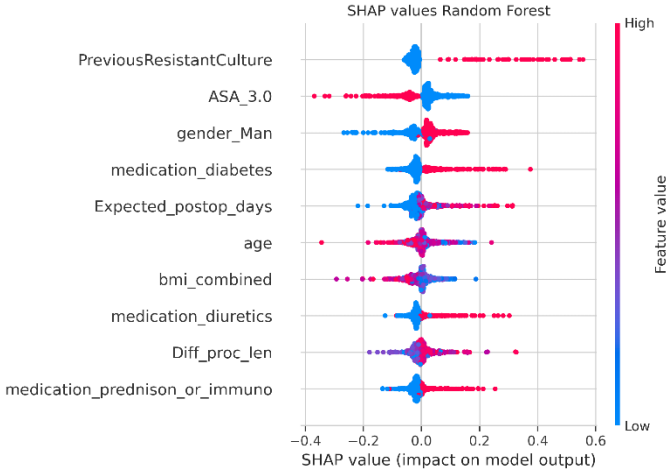
Blood

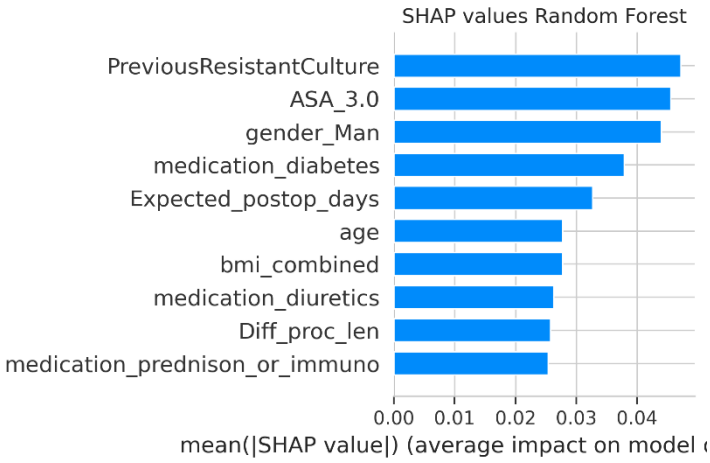
Logistic Regression



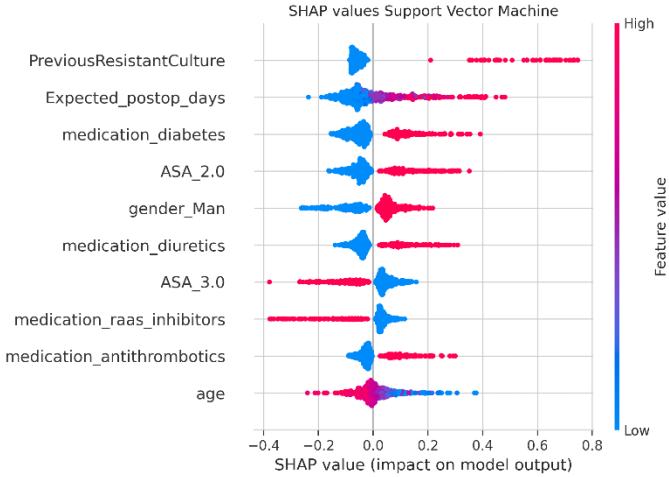


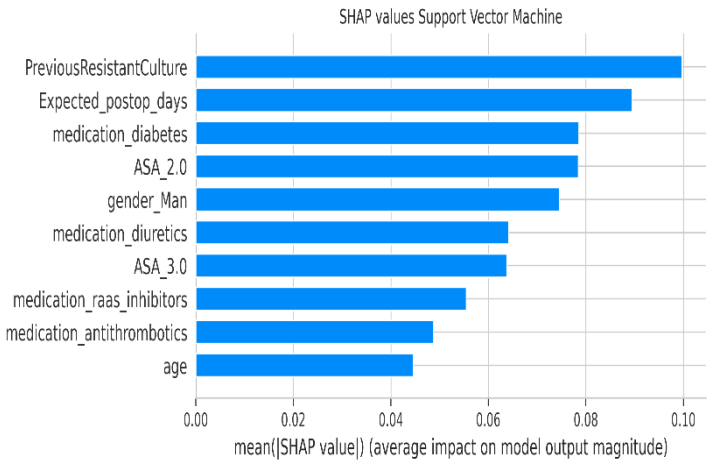
Random Forest





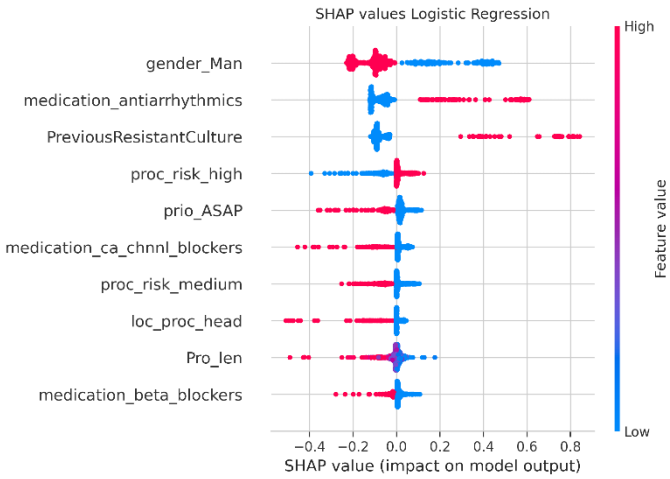
Support Vector Machine

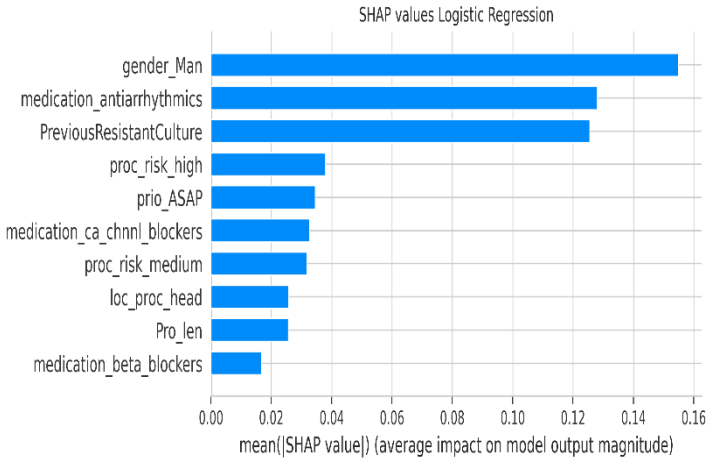




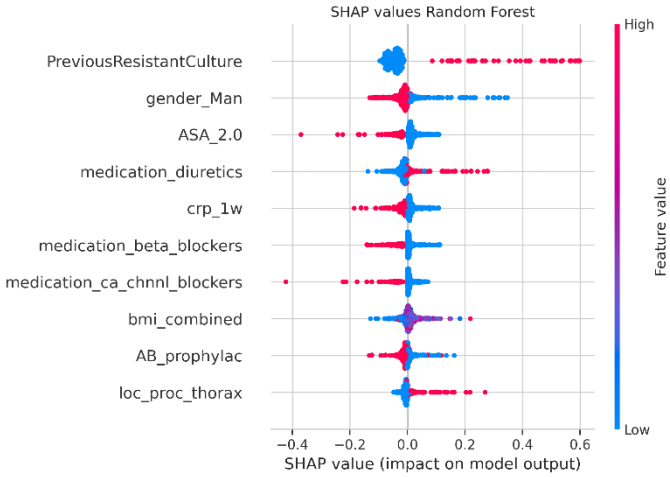
Respiratory

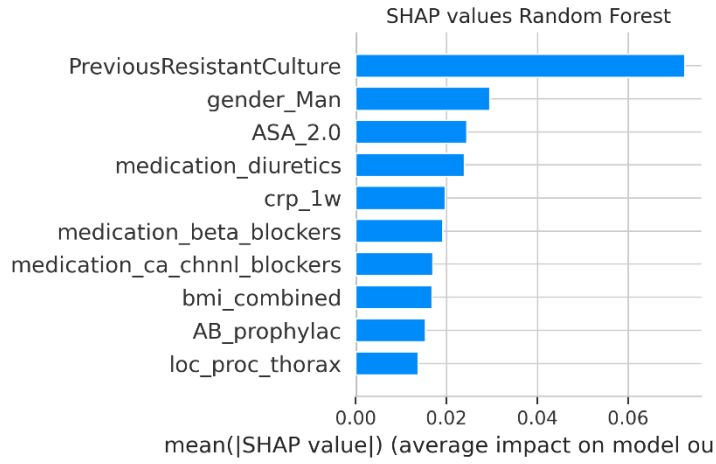
Logistic Regression



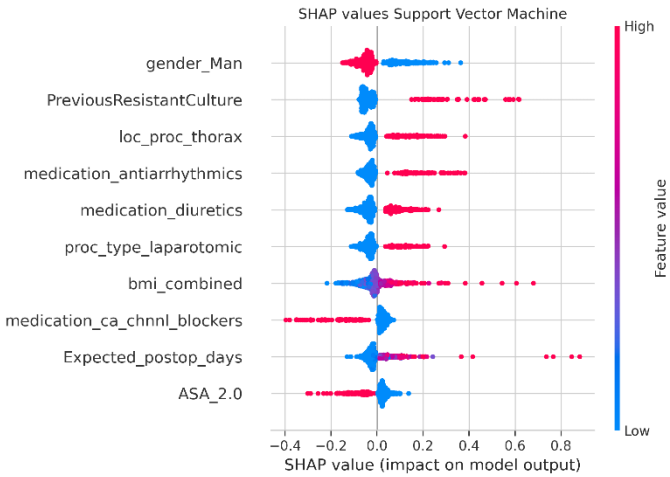


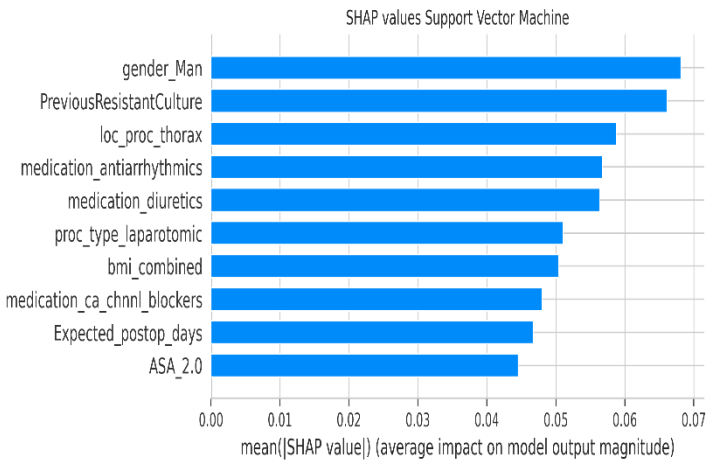
Random Forest





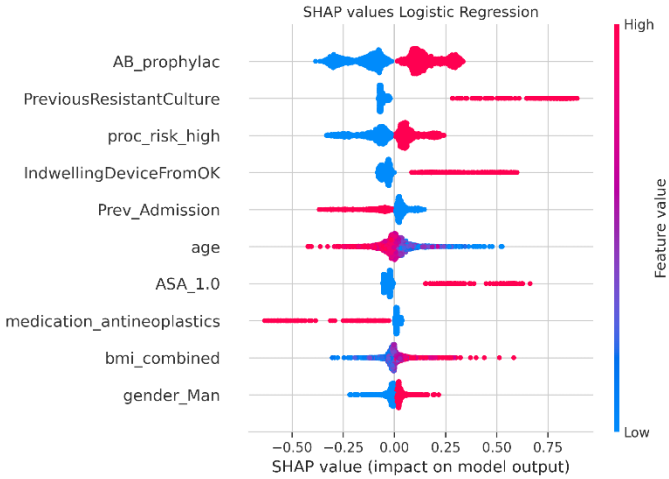
Support Vector Machine

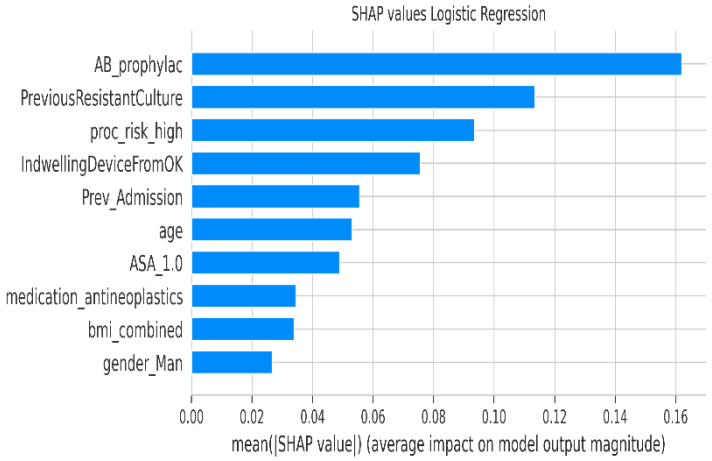




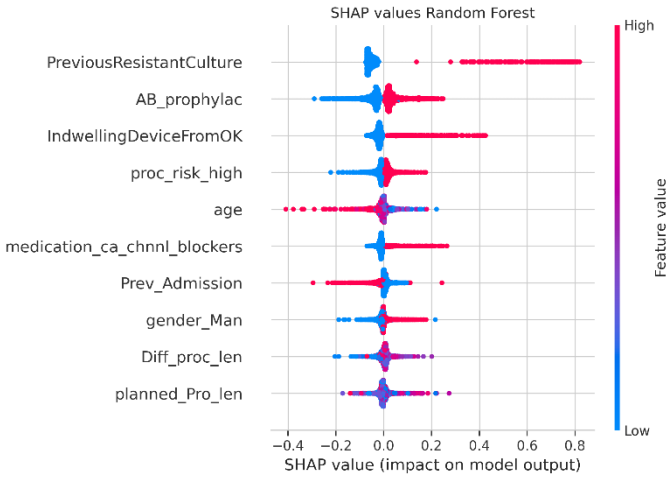
Urine

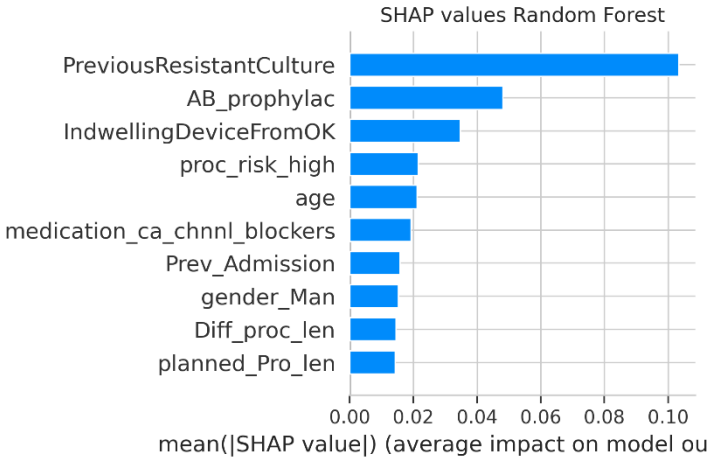
Logistic Regression



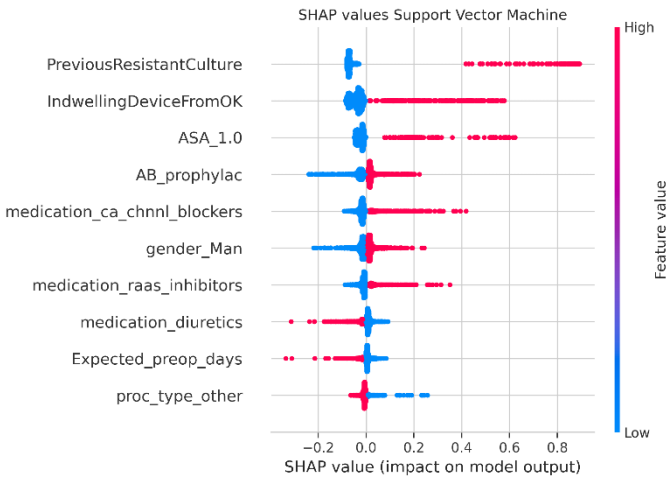


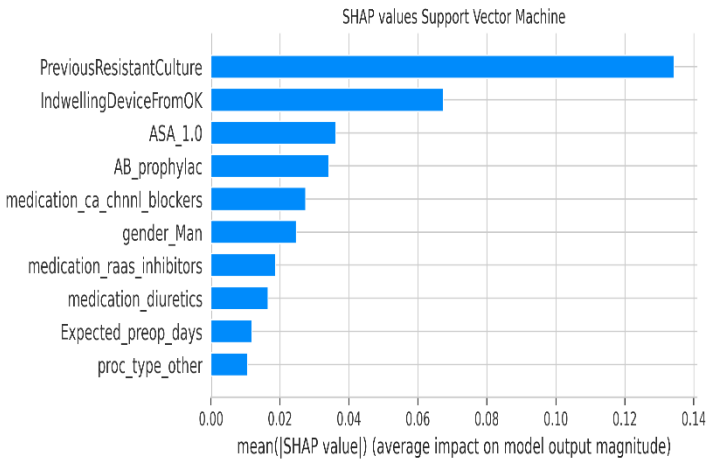
Random Forest





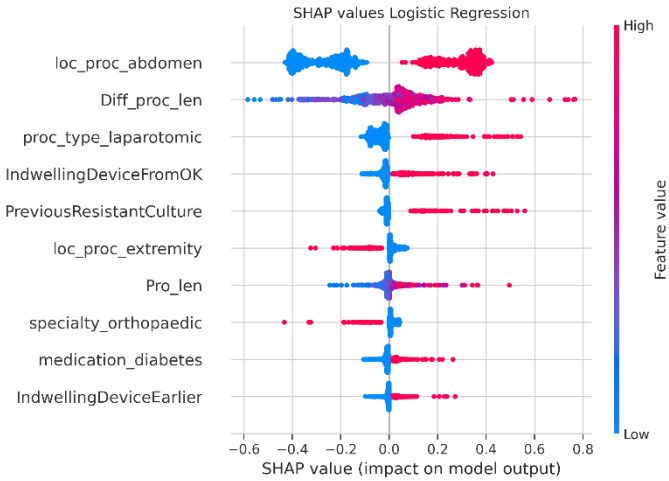
Support Vector Machine

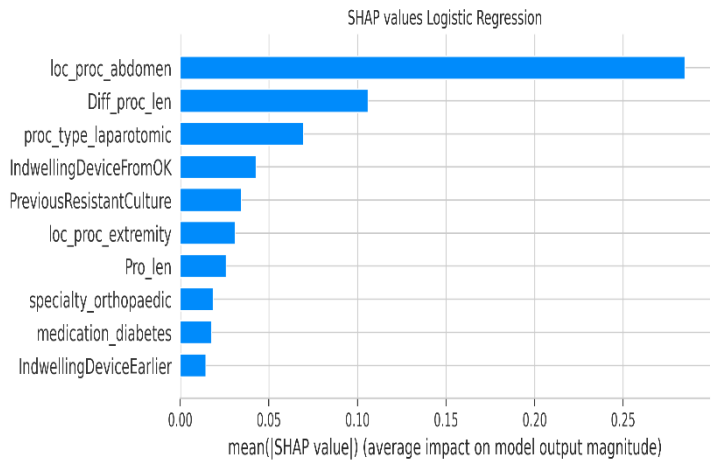




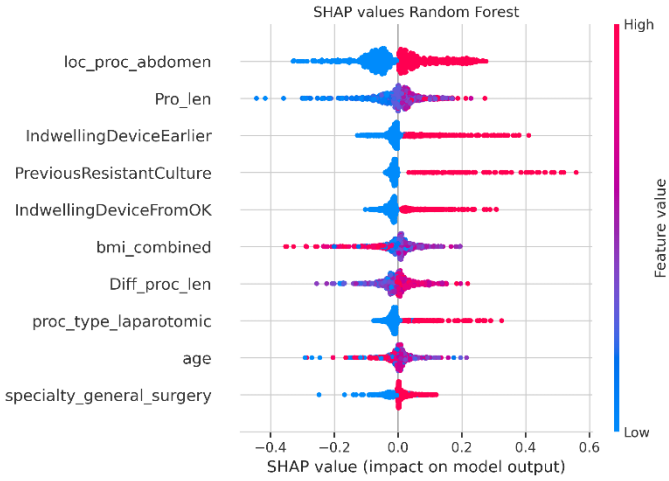
Wound

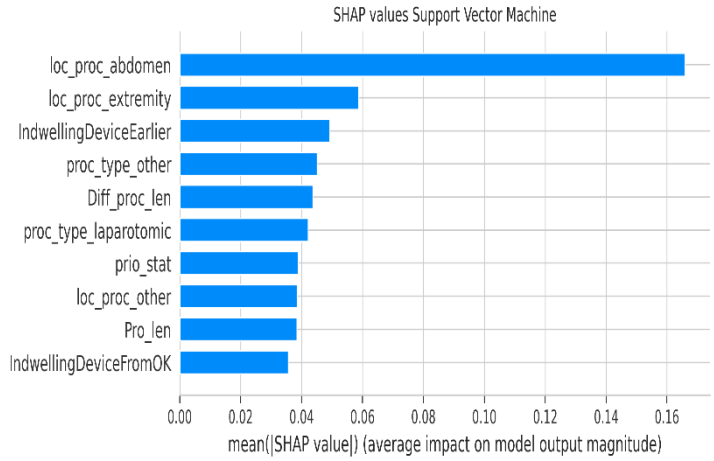
Logistic Regression



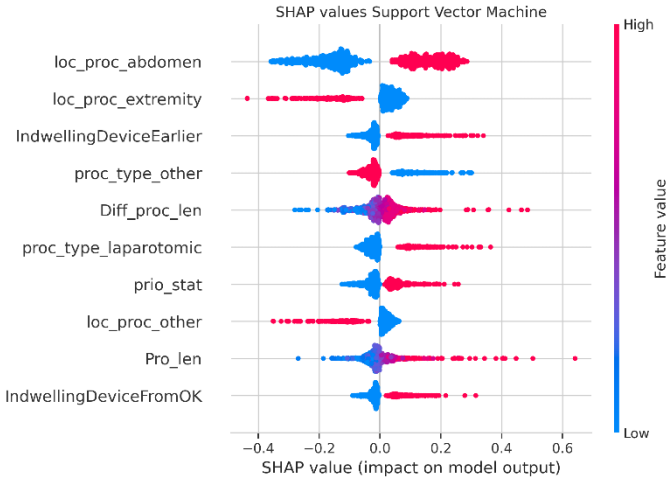


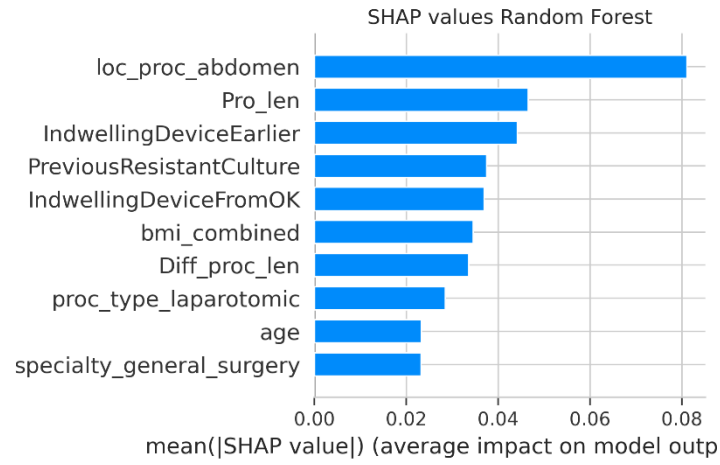
Random Forest





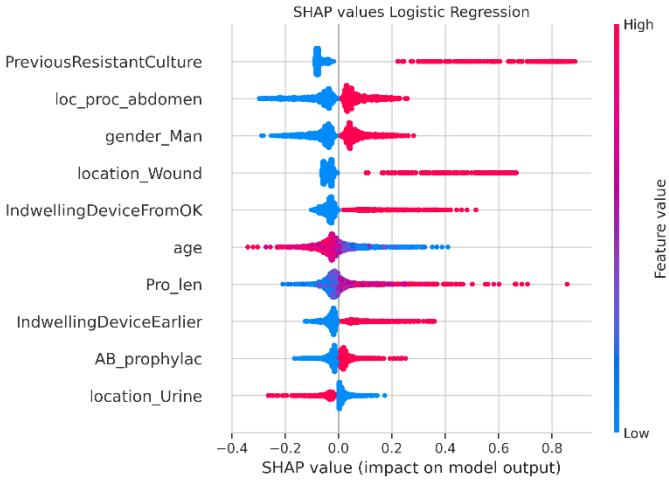
Support Vector Machine

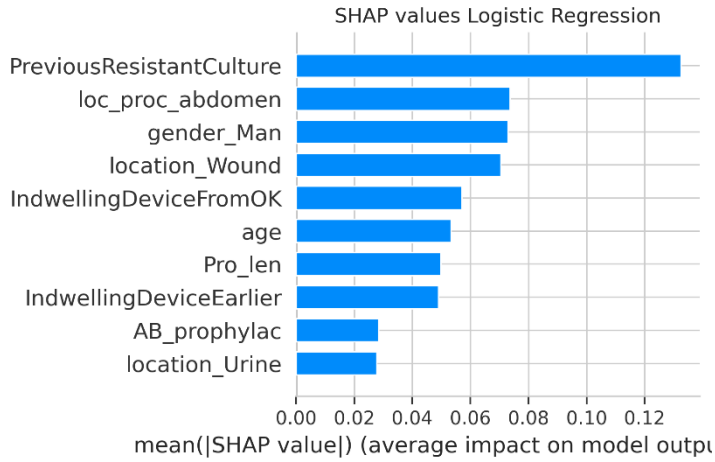




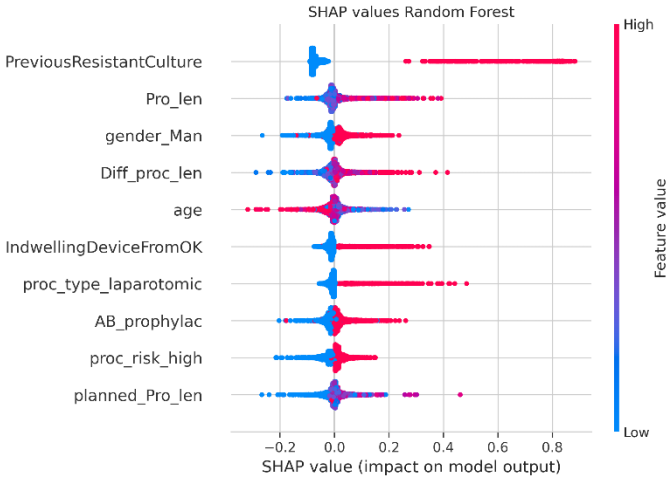
Escherichia coli

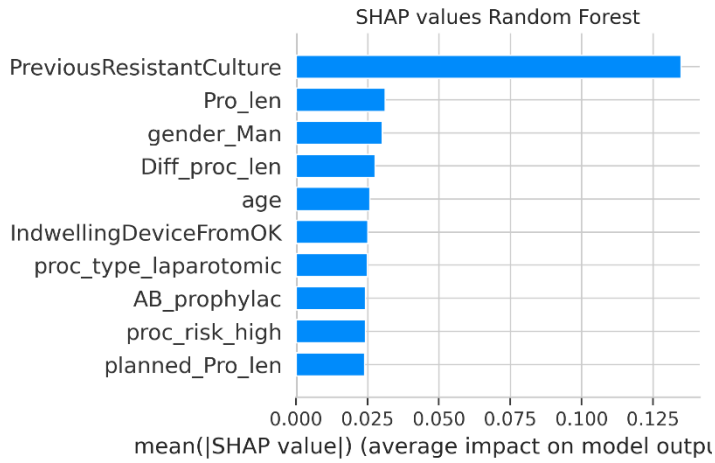
Logistic Regression



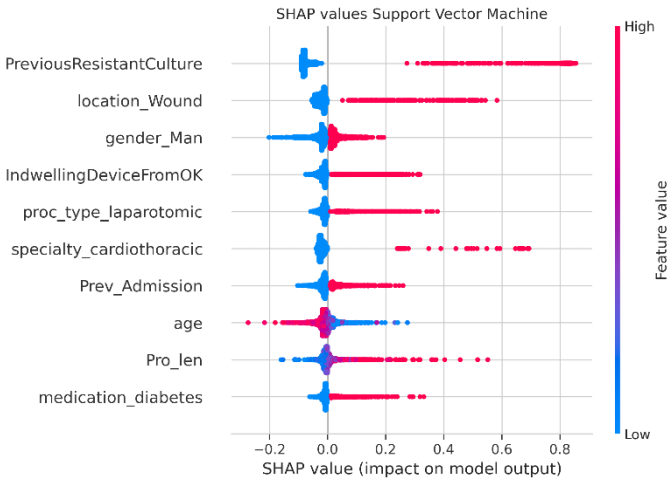


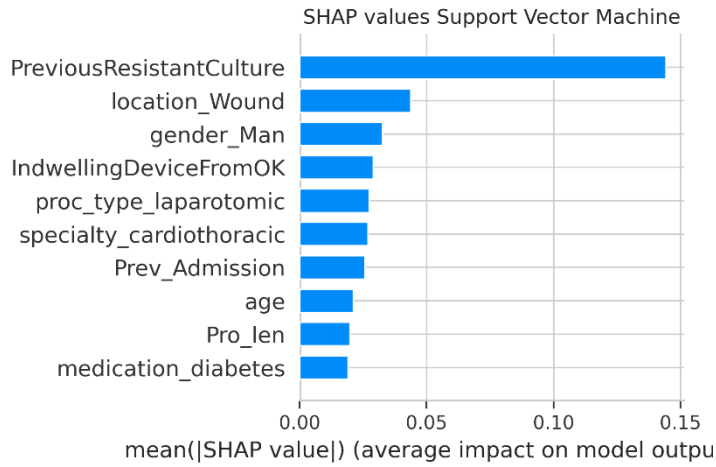
Random Forest





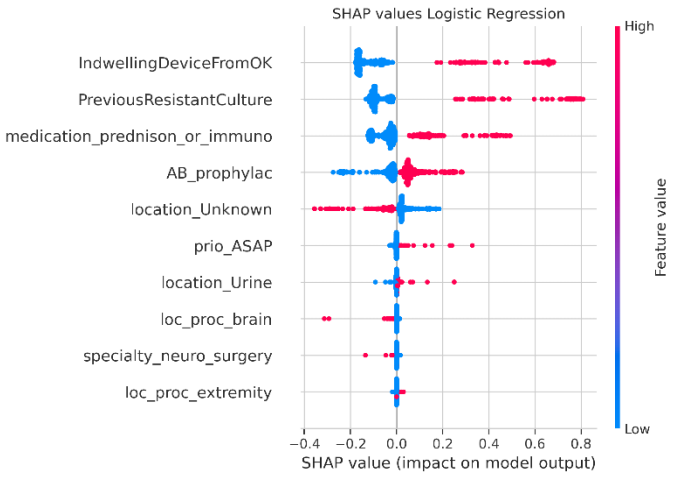
Support Vector Machine

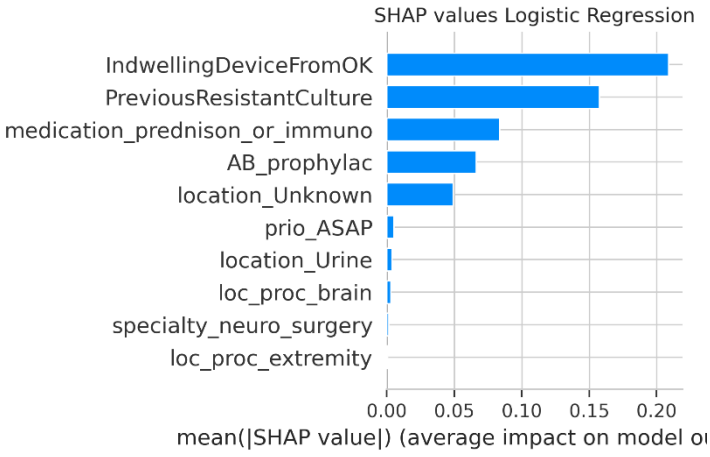




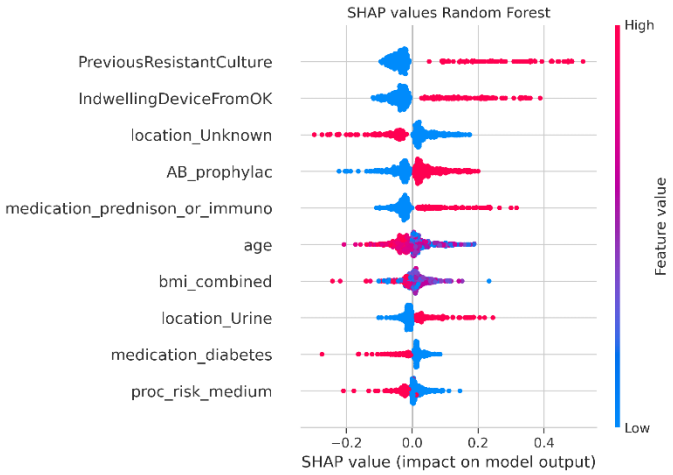
Klebsiella pneuomniae

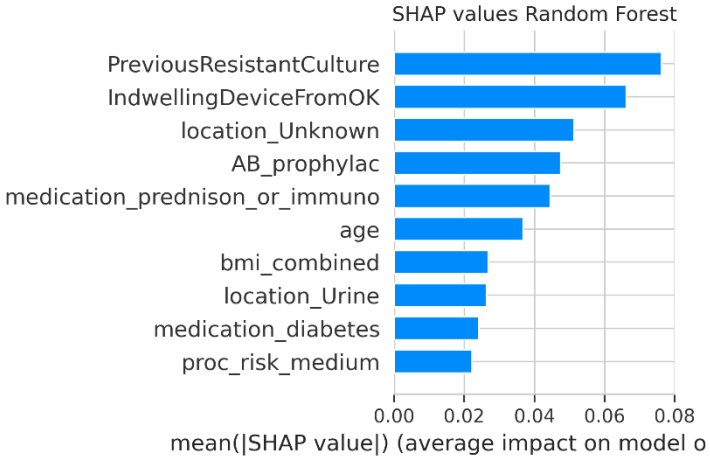
Logistic Regression



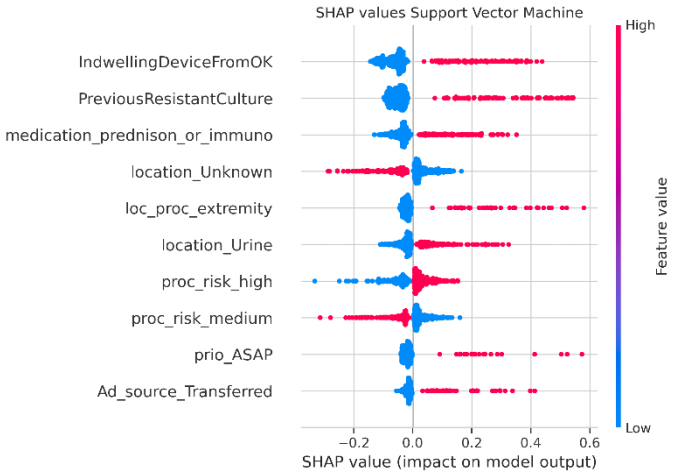


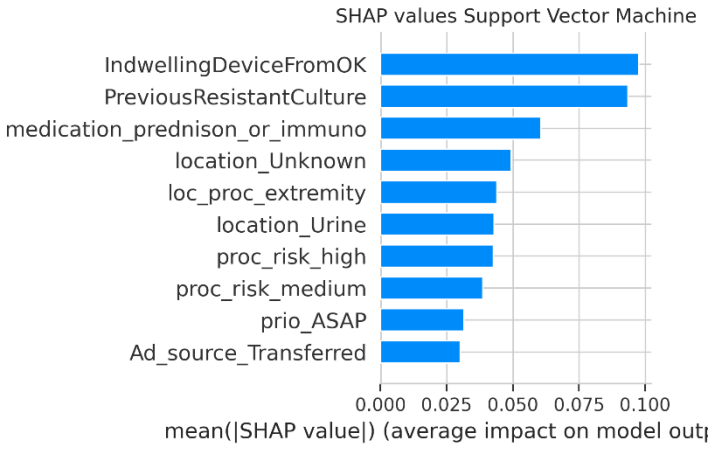
Random Forest





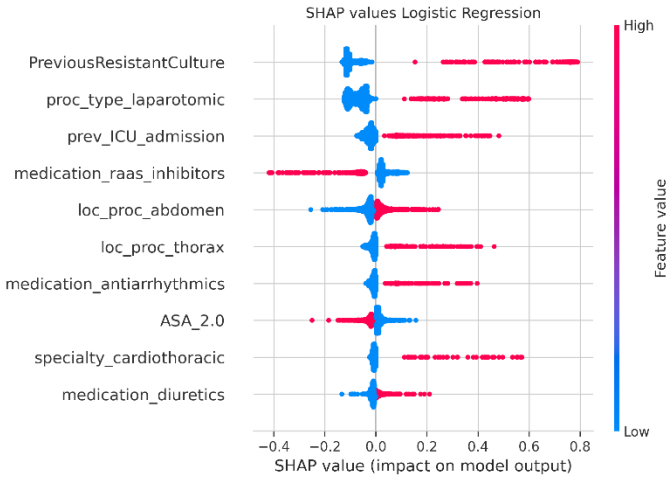
Support Vector Machine

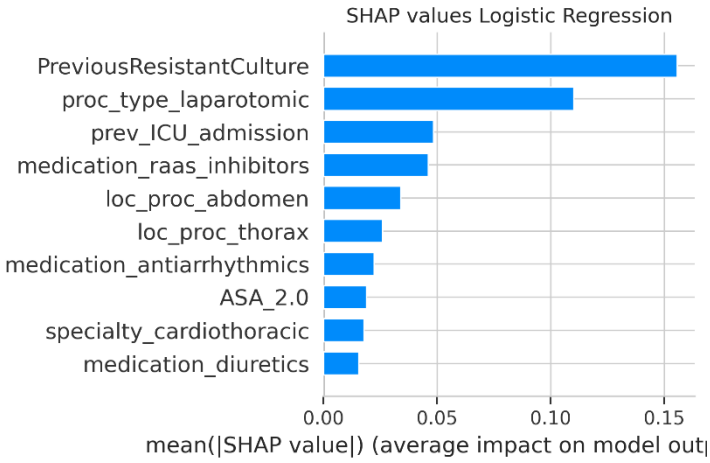




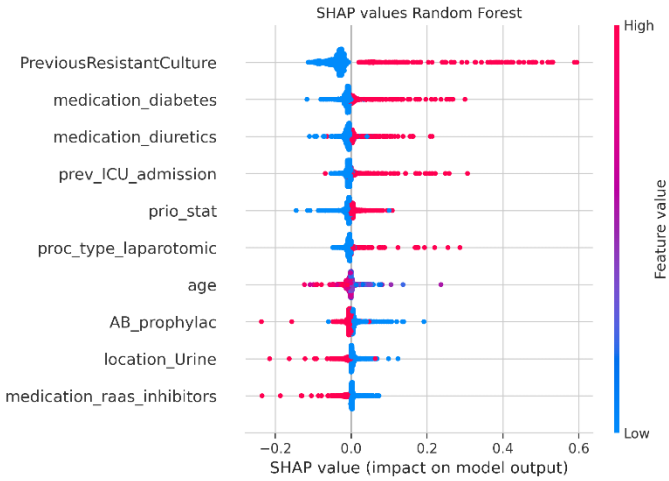
Pseudomonas aeruginosa

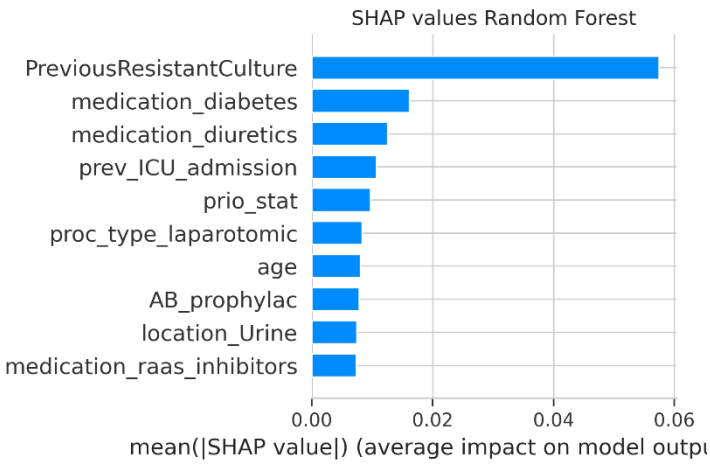
Logistic Regression



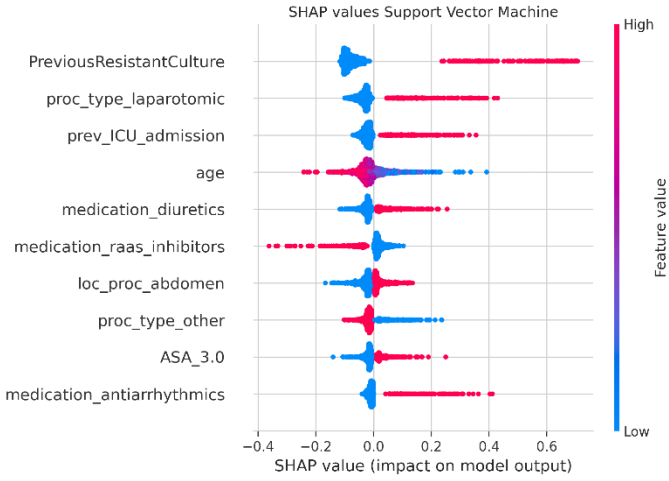


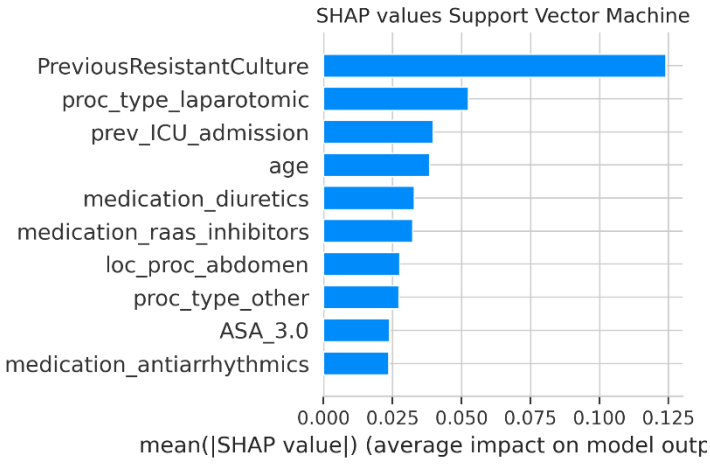
Random Forest





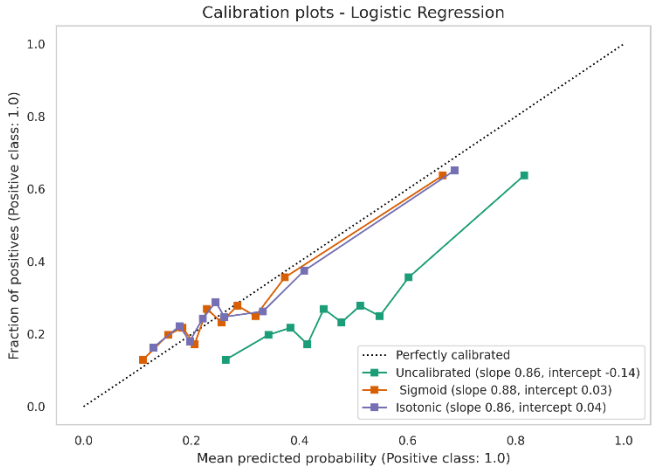
Support Vector Machine

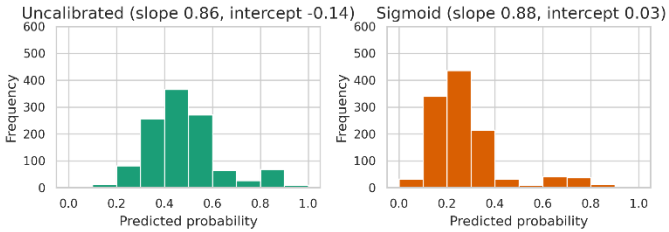


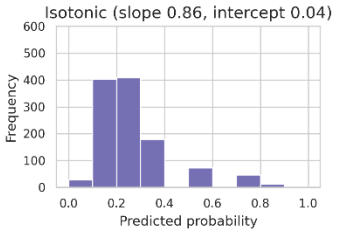


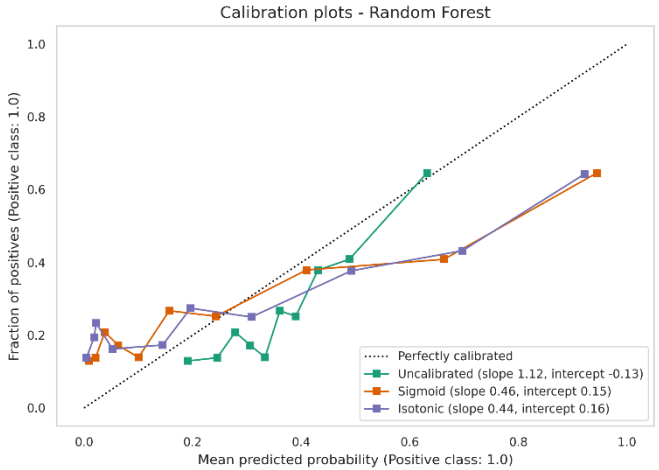
Results for model calibration

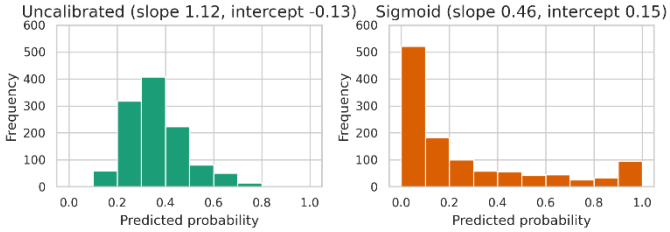
Total population

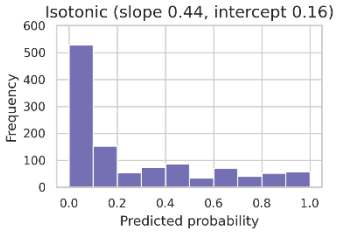


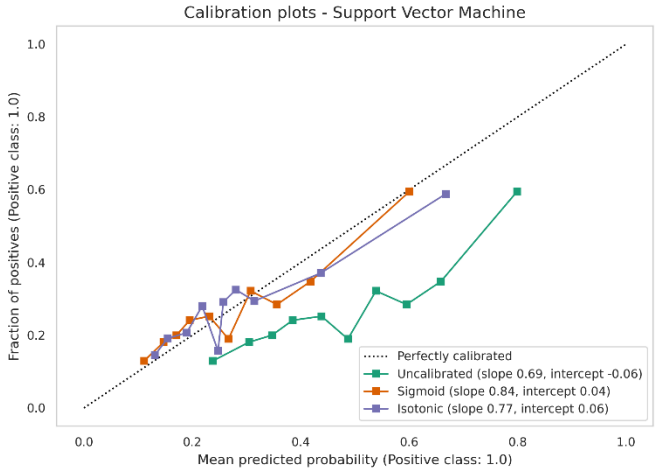


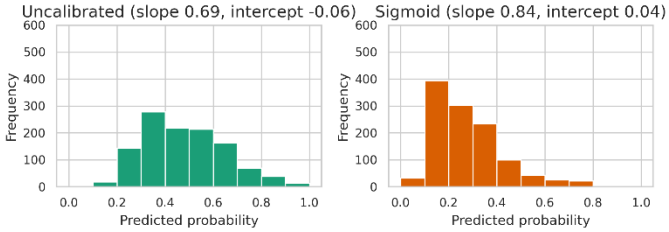


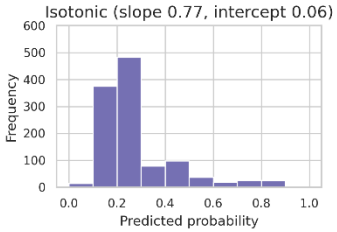




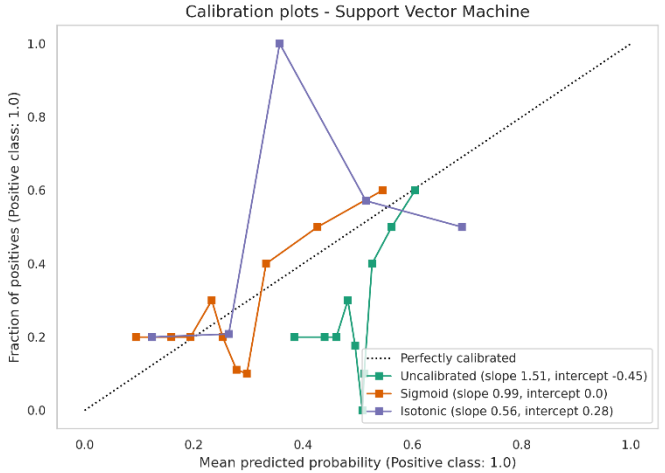


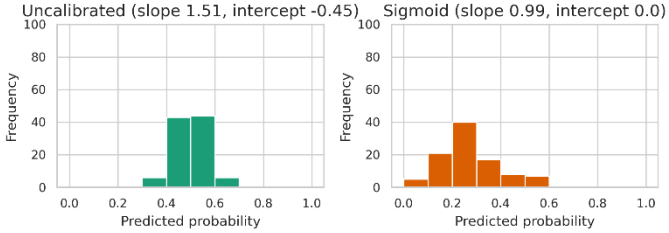


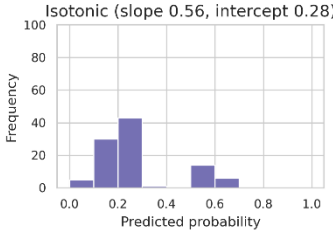


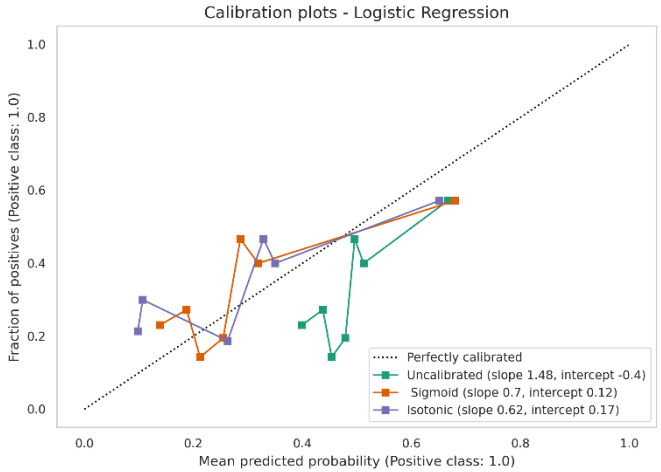


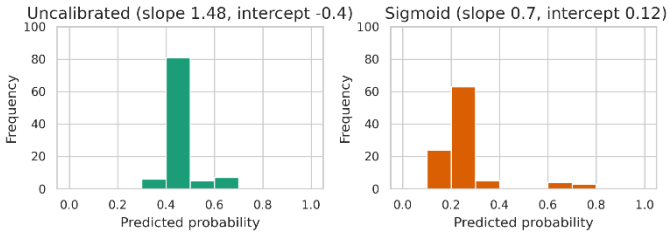
Ascites

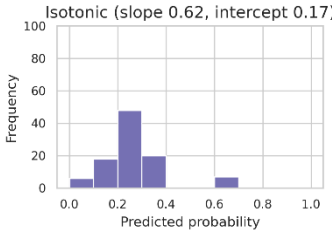

Calibration plots - Support Vector Machine
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 1.51, intercept -0.45)
Sigmoid (slope 0.99, intercept 0.0)
Isotonic (slope 0.56, intercept 0.28)

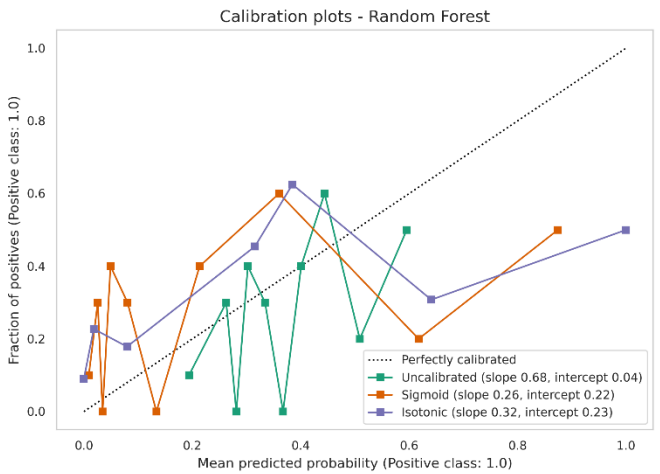

Uncalibrated (slope 1.51, intercept -0.45)
Sigmoid (slope 0.99, intercept 0.0)
Frequency
Predicted probability

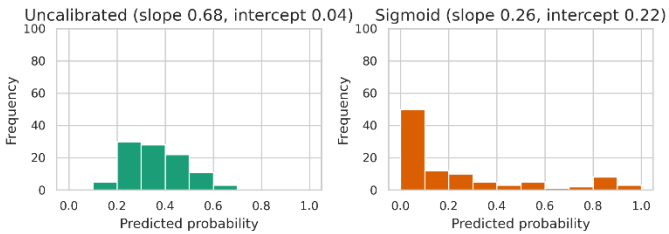

Isotonic (slope 0.56, intercept 0.28)
Frequency
Predicted probability

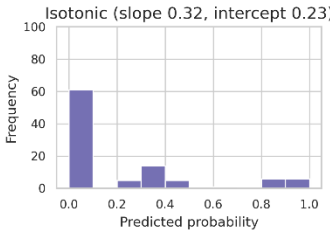

Calibration plots - Logistic Regression
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 1.48, intercept -0.4)
Sigmoid (slope 0.7, intercept 0.12)
Isotonic (slope 0.62, intercept 0.17)


Uncalibrated (slope 1.48, intercept -0.4)
Sigmoid (slope 0.7, intercept 0.12)
Frequency
Predicted probability

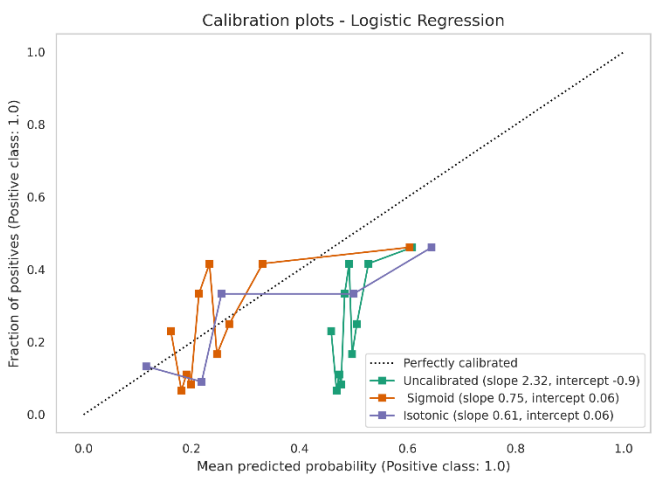

Isotonic (slope 0.62, intercept 0.17)
Frequency
Predicted probability

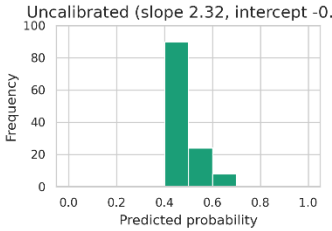

Calibration plots - Random Forest
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 0.68, intercept 0.04)
Sigmoid (slope 0.26, intercept 0.22)
Isotonic (slope 0.32, intercept 0.23)

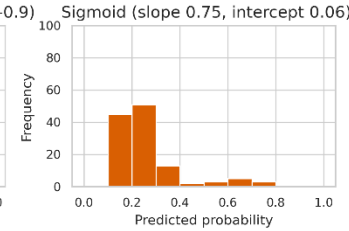

Uncalibrated (slope 0.68, intercept 0.04)
Sigmoid (slope 0.26, intercept 0.22)
Frequency
Predicted probability

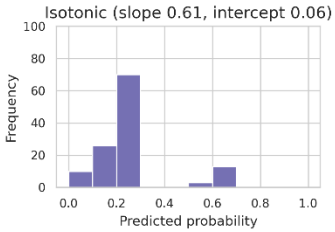

Isotonic (slope 0.32, intercept 0.23)
Frequency
Predicted probability

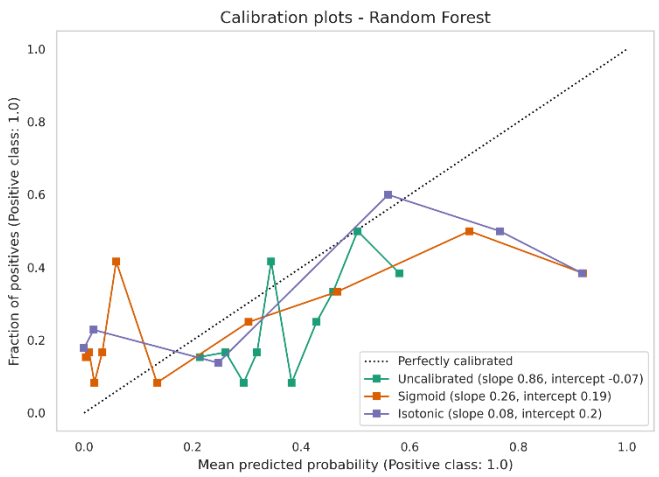
Blood

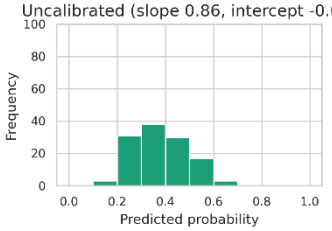

Calibration plots - Logistic Regression
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 2.32, intercept -0.9)
Sigmoid (slope 0.75, intercept 0.06)
Isotonic (slope 0.61, intercept 0.06)

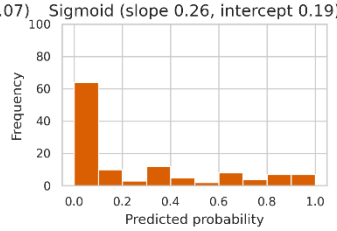

Uncalibrated (slope 2.32, intercept -0.9)
Frequency
Predicted probability

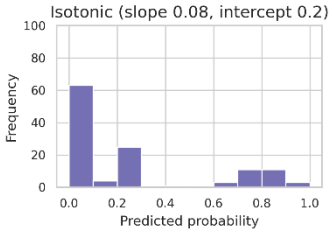

Sigmoid (slope 0.75, intercept 0.06)
Frequency
Predicted probability

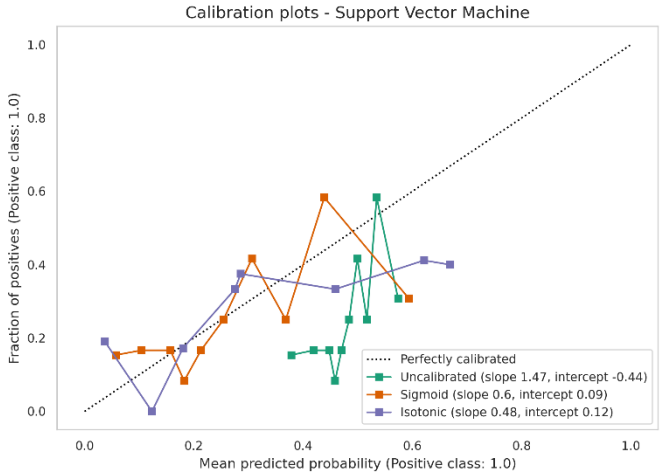

Isotonic (slope 0.61, intercept 0.06)
Frequency
Predicted probability

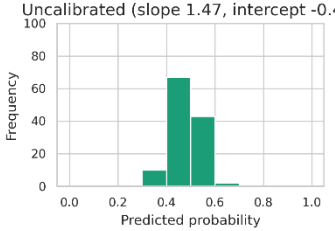

Calibration plots - Random Forest
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 0.86, intercept -0.07)
Sigmoid (slope 0.26, intercept 0.19)
Isotonic (slope 0.08, intercept 0.2)

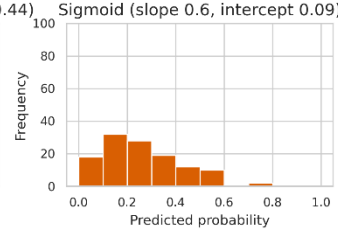

Uncalibrated (slope 0.86, intercept -0.07)
Frequency
Predicted probability

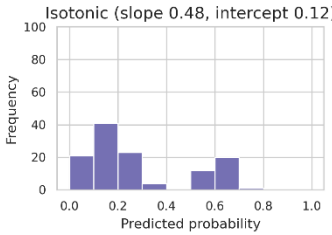

Sigmoid (slope 0.26, intercept 0.19)
Frequency
Predicted probability


Isotonic (slope 0.08, intercept 0.2)
Frequency
Predicted probability

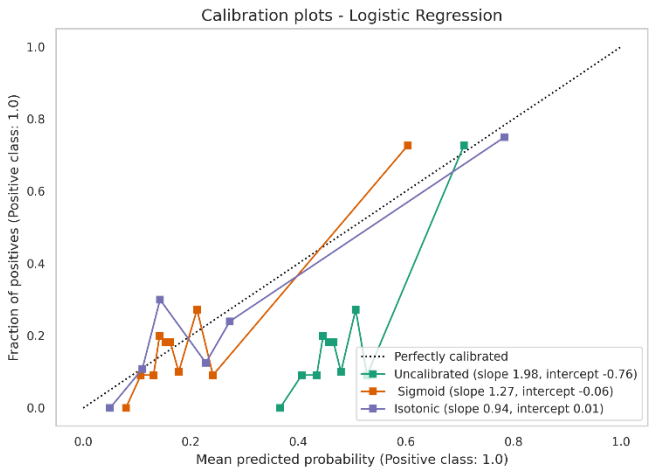

Calibration plots - Support Vector Machine
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 1.47, intercept -0.44)
Sigmoid (slope 0.6, intercept 0.09)
Isotonic (slope 0.48, intercept 0.12)

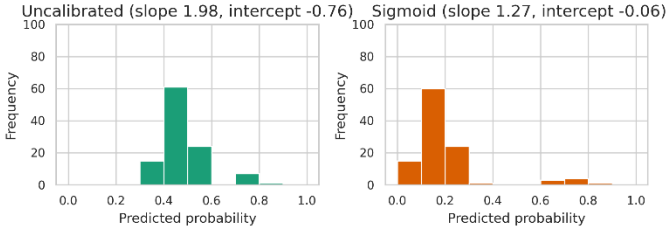

Uncalibrated (slope 1.47, intercept -0.44)
Frequency
Predicted probability

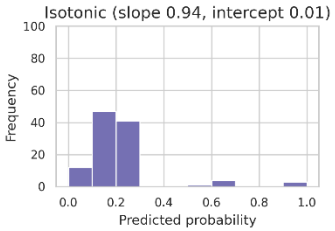

Sigmoid (slope 0.6, intercept 0.09)
Frequency
Predicted probability

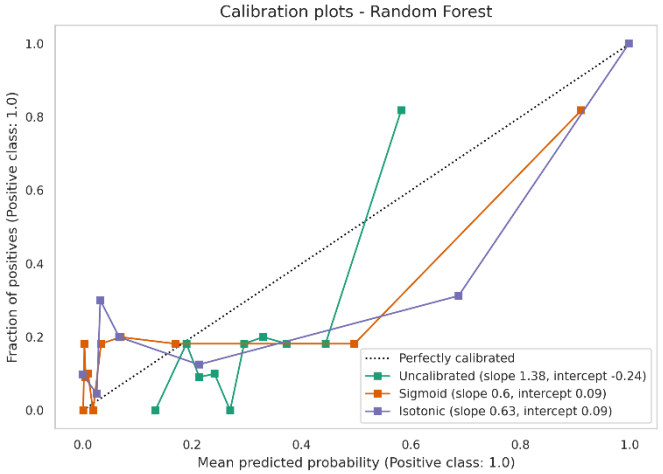

Isotonic (slope 0.48, intercept 0.12)
Frequency
Predicted probability

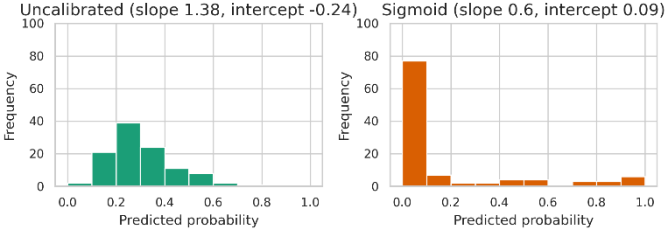
Respiratory

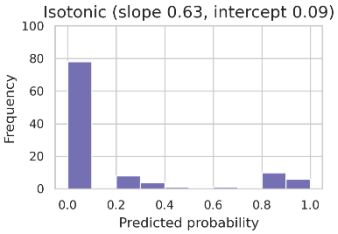


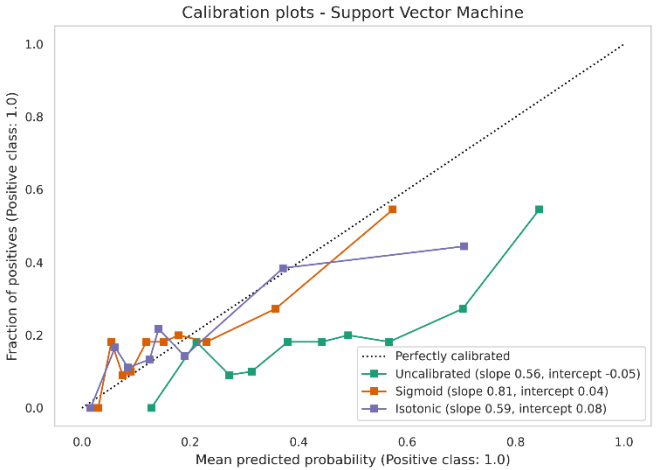


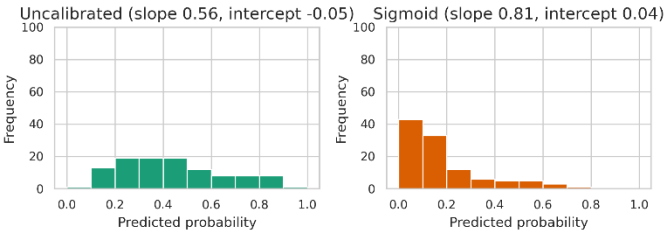


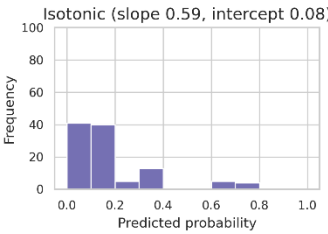




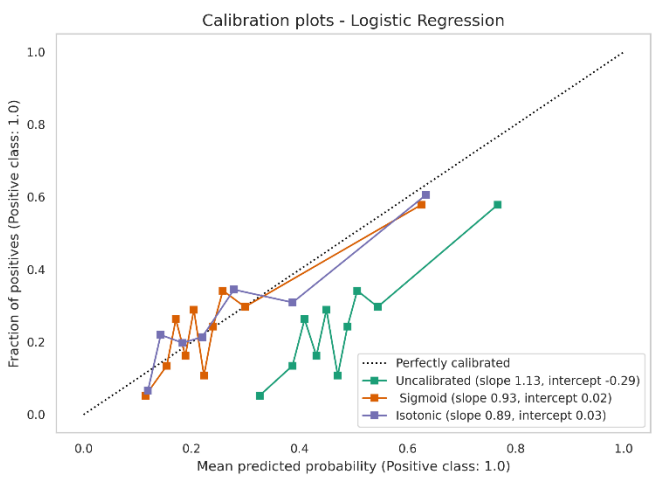


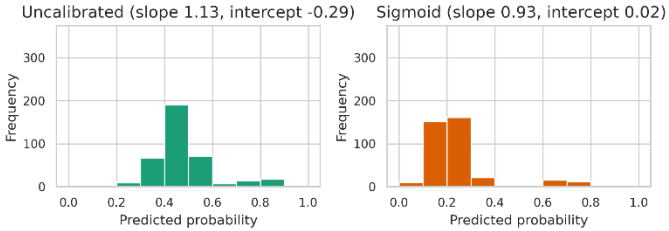




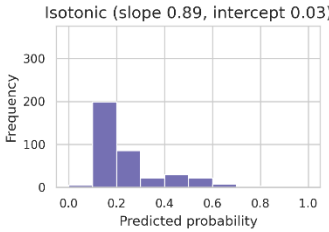


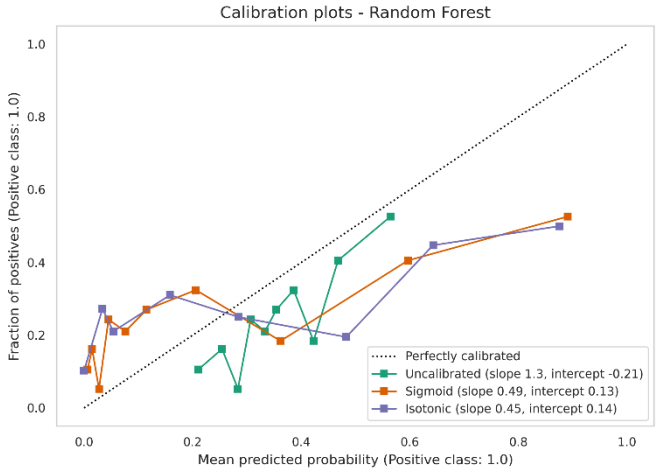
Urine

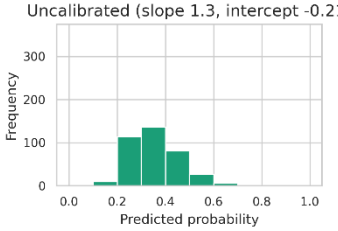

Calibration plots - Logistic Regression
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 1.13, intercept -0.29)
Sigmoid (slope 0.93, intercept 0.02)
Isotonic (slope 0.89, intercept 0.03)

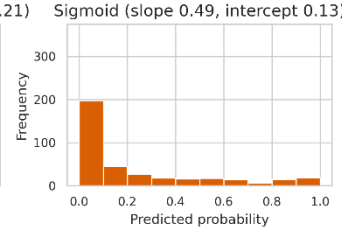

Uncalibrated (slope 1.13, intercept -0.29)
Frequency
Predicted probability

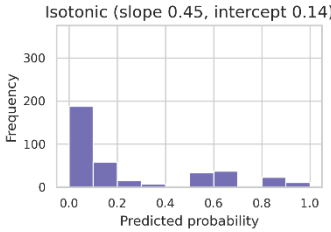
Sigmoid (slope 0.93, intercept 0.02)
Frequency
Predicted probability

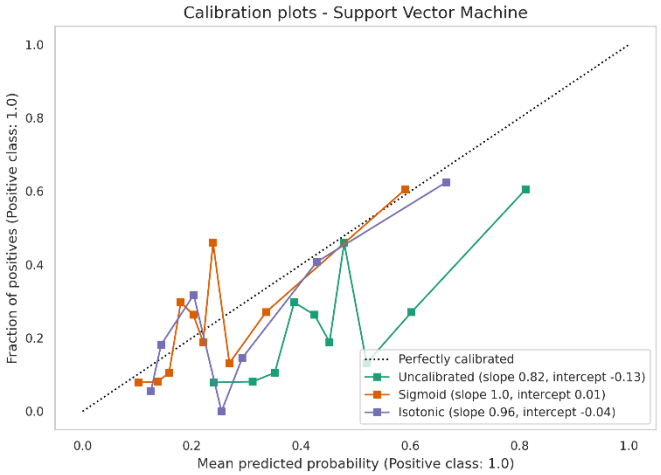

Isotonic (slope 0.89, intercept 0.03)
Frequency
Predicted probability

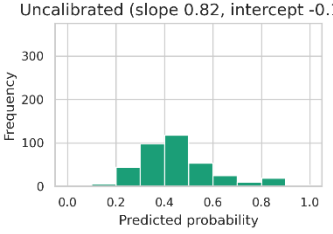

Calibration plots - Random Forest
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 1.3, intercept -0.21)
Sigmoid (slope 0.49, intercept 0.13)
Isotonic (slope 0.45, intercept 0.14)

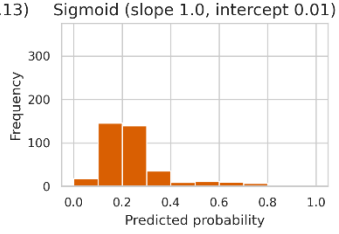

Uncalibrated (slope 1.3, intercept -0.21)
Frequency
Predicted probability

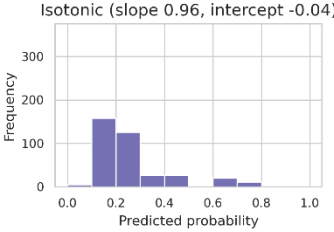

Sigmoid (slope 0.49, intercept 0.13)
Frequency
Predicted probability


Isotonic (slope 0.45, intercept 0.14)
Frequency
Predicted probability

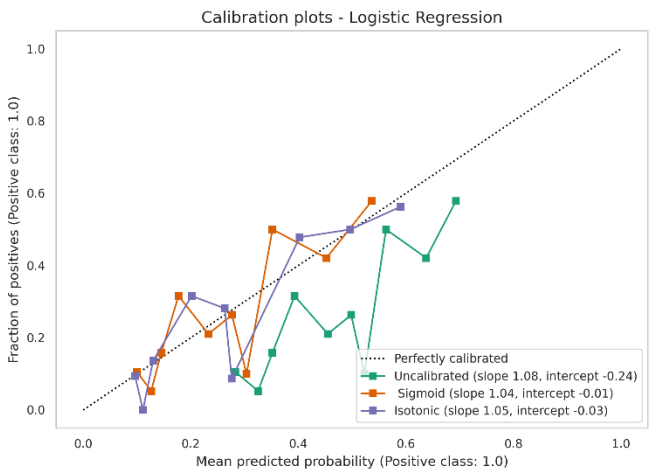

Calibration plots - Support Vector Machine
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 0.82, intercept -0.13)
Sigmoid (slope 1.0, intercept 0.01)
Isotonic (slope 0.96, intercept -0.04)

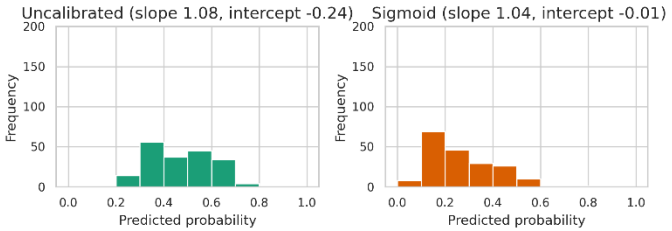

Uncalibrated (slope 0.82, intercept -0.13)
Frequency
Predicted probability

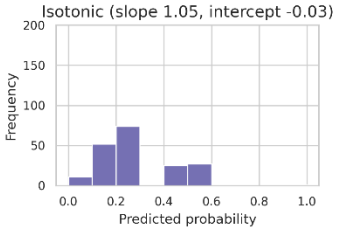

Sigmoid (slope 1.0, intercept 0.01)
Frequency
Predicted probability

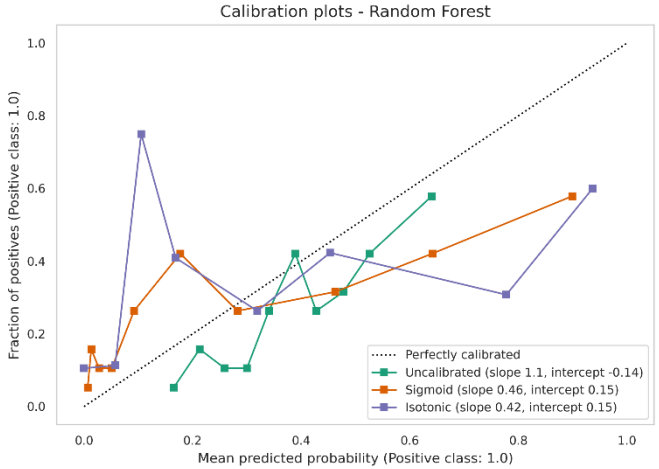

Isotonic (slope 0.96, intercept -0.04)
Frequency
Predicted probability

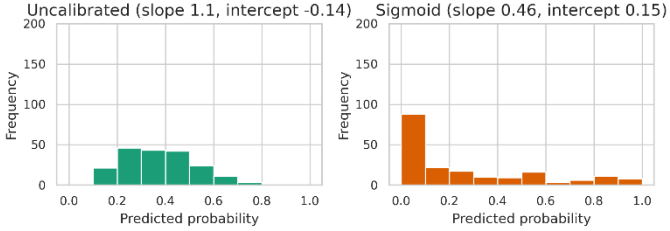
Wound

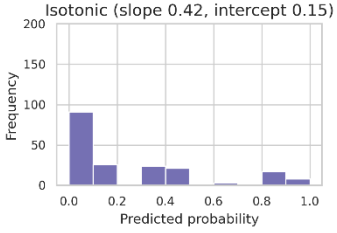

Calibration plots - Logistic Regression
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 1.08, intercept -0.24)
Sigmoid (slope 1.04, intercept -0.01)
Isotonic (slope 1.05, intercept -0.03)

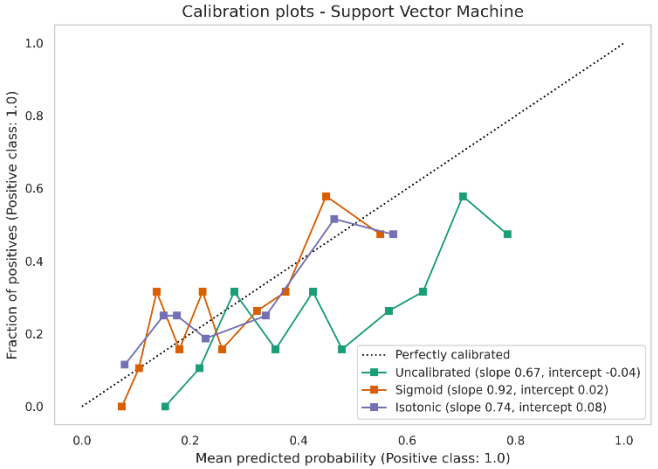

Uncalibrated (slope 1.08, intercept -0.24)
Sigmoid (slope 1.04, intercept -0.01)
Frequency
Predicted probability

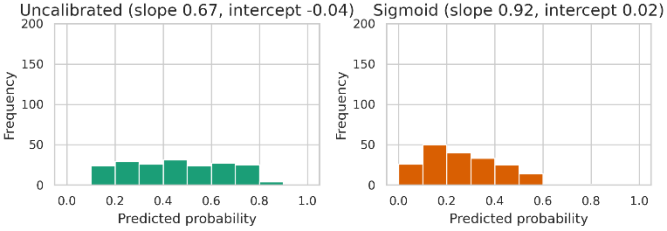

Isotonic (slope 1.05, intercept -0.03)
Frequency
Predicted probability

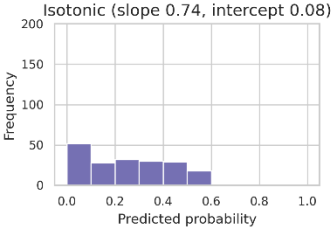

Calibration plots - Random Forest
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 1.1, intercept -0.14)
Sigmoid (slope 0.46, intercept 0.15)
Isotonic (slope 0.42, intercept 0.15)


Uncalibrated (slope 1.1, intercept -0.14)
Sigmoid (slope 0.46, intercept 0.15)
Frequency
Predicted probability

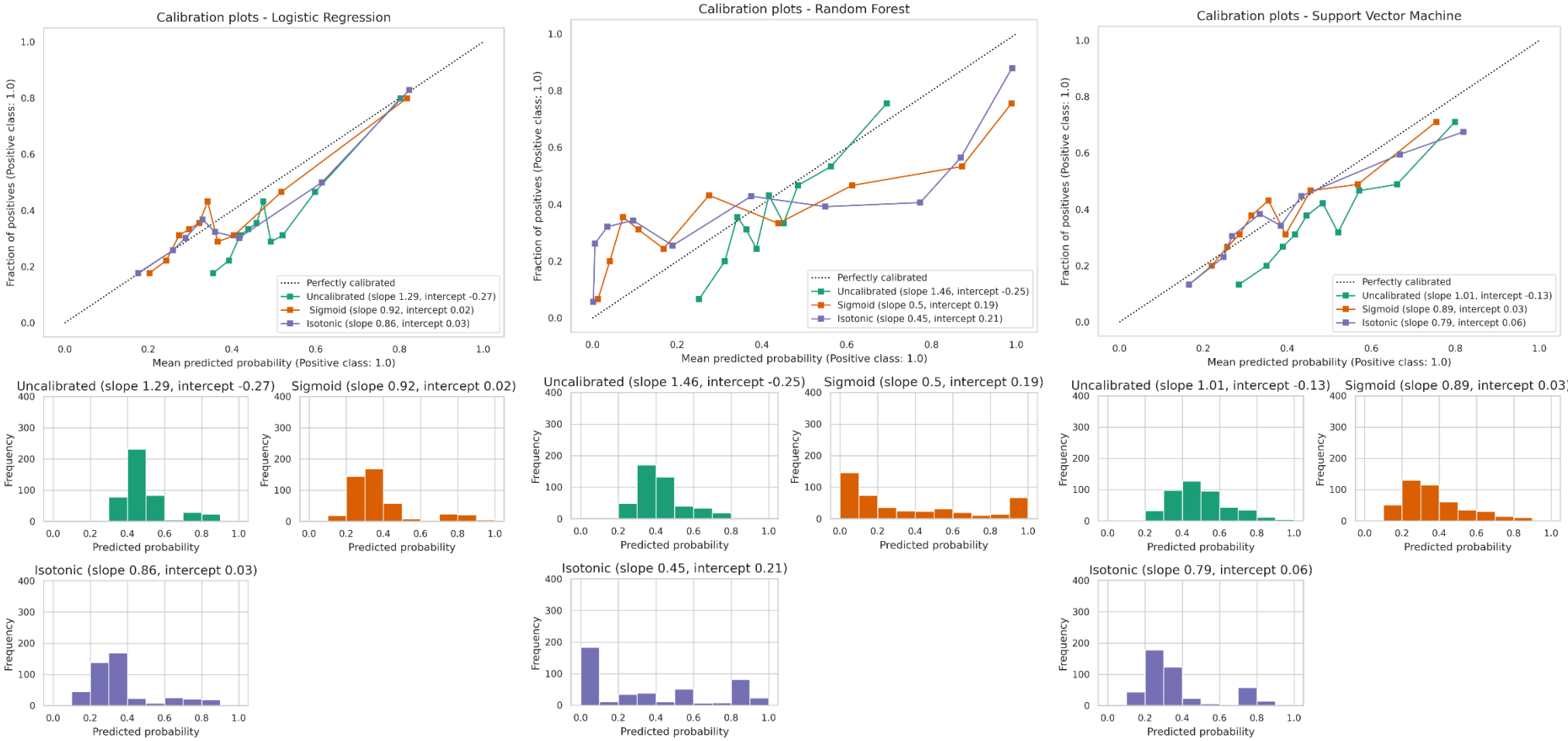

Isotonic (slope 0.42, intercept 0.15)
Frequency
Predicted probability


Calibration plots - Support Vector Machine
Fraction of positives (Positive class: 1.0)
Mean predicted probability (Positive class: 1.0)
Perfectly calibrated
Uncalibrated (slope 0.67, intercept -0.04)
Sigmoid (slope 0.92, intercept 0.02)
Isotonic (slope 0.74, intercept 0.08)

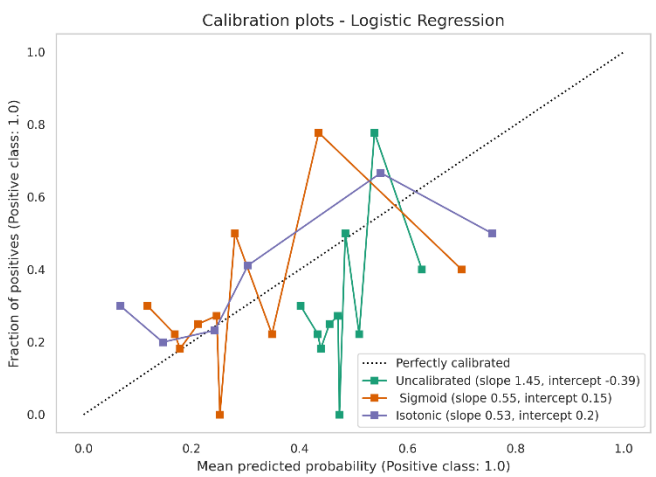

Uncalibrated (slope 0.67, intercept -0.04)
Sigmoid (slope 0.92, intercept 0.02)
Frequency
Predicted probability

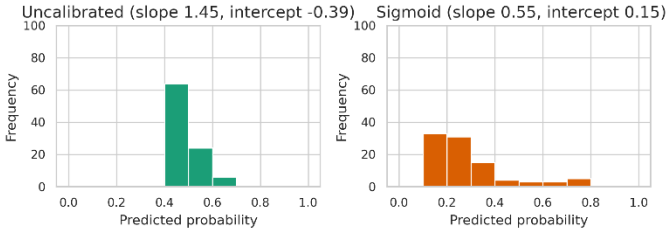

Isotonic (slope 0.74, intercept 0.08)
Frequency
Predicted probability

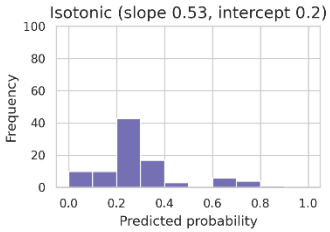
Escherichia coli

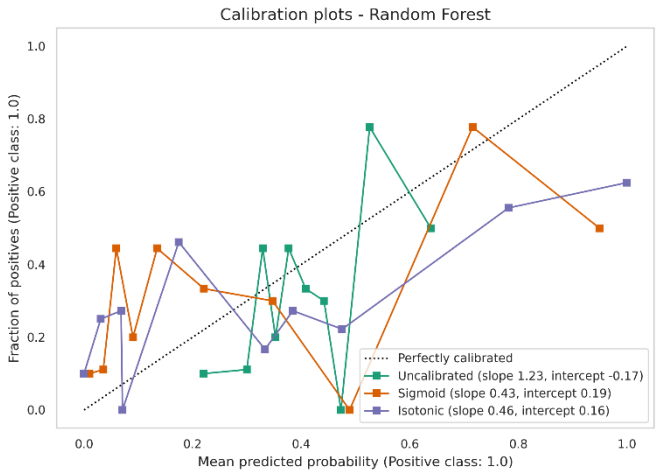


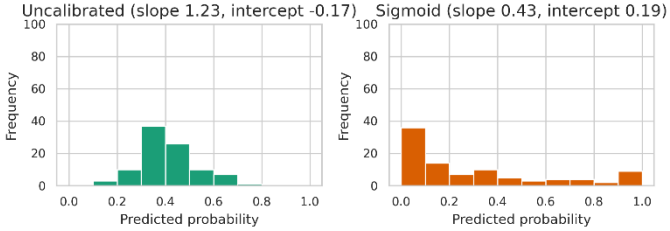
Klebsiella pneuomniae

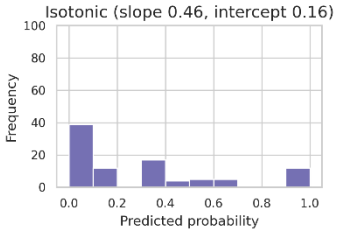


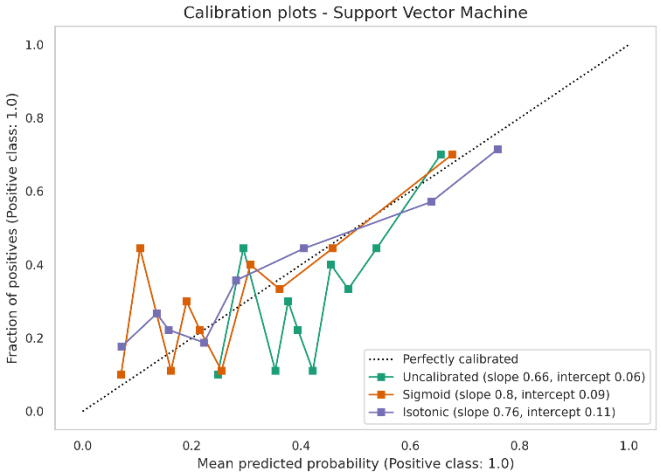


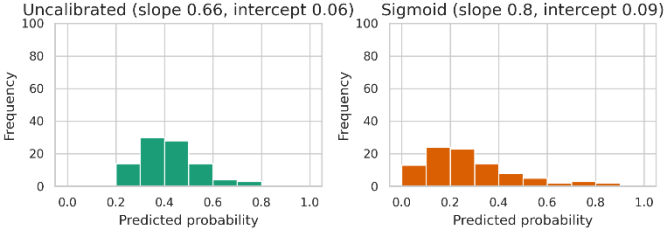


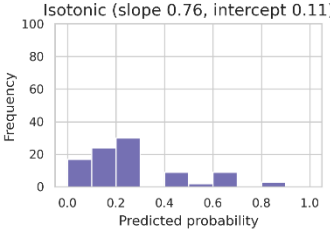












Pseudomonas aeruginosa