

Supplementary material

1. Variable description:

Following are the descriptions of the variables that were concluded most important for the resistance prediction according to our analysis:

Previous resistance - specific – the proportion of previous same-species cultures resistant to the examined antibiotic, obtained previously from the patient.

Previous resistance - general – the proportion of previous same-species cultures resistant to any antibiotic, obtained previously from the patient.

Previous any bacterial resistance - specific – the proportion of previous any-species cultures resistant to the examined antibiotic, obtained previously from the patient.

Previous any bacterial resistance - general – the proportion of previous any-species cultures resistant to any antibiotic, obtained previously from the patient.

Previous use of same antibiotic - 1 if the patient previously used the examined antibiotic, 0 otherwise

Previous use of same-family antibiotic – 1 if the patient previously used a drug of same family as the examined antibiotics, 0 otherwise. Antibiotics were clustered into broad families: beta-lactams, fluoroquinolones, aminoglycosides, sulfonamides.

Nosocomial – 1 if the bacterial infection occurred >48 hours after hospitalization, 0 otherwise.

Sample location – source of bacterial sample.

Referral location – the department from which the patient arrived.

Functioning before admission – the functioning of the patient before hospital admission [independent, nursing, complicated nursing, etc.].

Functioning at admission – the functioning of the patient at admission [independent, nursing, complicated nursing, etc.].

Last hospitalization duration – the length (in days) of the last hospitalization.

Location from which the patient arrived the hospital – his home, nursing facility etc.

sex – 1 if the patient is a male, 0 if female

D_i –previously administered antibiotic drugs. The 65 most commonly prescribed drugs, as well as an additional category containing all drugs taken but not found in other categories.

All variables were scaled and centered with respect to the training set before application of the machine learning algorithms.

Two distinct samples were defined as obtained from different patients, or from the same patient but from different locations (e.g. blood, urine, etc.), or from the same patient and location but containing different bacteria; or from the same patient, at the same location, with the same bacteria, but at least 7 days apart.

2. Ensemble training and hyperparameter tuning:

Hyperparameter tuning was performed solely on the training set. Models differing in their hyperparameter values were trained and evaluated using cross validation (CV), and the model that achieved the best performance, based on the area under receiver operating characteristic curve (auROC), was chosen. Hyperparameter tuning was performed separately for Gradient-boosted trees, neural networks and L1 regularized logistic regression (LASSO):

- Gradient-Boosted Trees (**GBT**) – 2,000 models were trained. For each model, a random hyperparameter set was chosen for the following hyperparameters:
 - Number of gradient boosted trees
 - Maximum depth of each tree
 - Learning rate
 - Subsample ratio of columns when constructing each tree

- Subsample ratio of the training data prior to growing trees

The chosen hyperparameters were those achieving the highest auROC, based on a 3-fold CV on the training set.

- **Neural network (NN)** – we trained fully-connected neural networks with two hidden layers, and a random node dropout for the purpose of regularization of each hidden layer. The networks were trained for 100 epochs with batch sizes of 32. We used the ‘adam’ learning rate optimizer and a binary cross entropy loss function. The final layer contained one node with a sigmoid activation function. We examined several combinations of the sizes of both hidden layers (in the range of 10-200), and for each combination we examined different dropout probabilities (in the range of 0.3-0.9). Each NN configuration was trained and evaluated with 3-fold CV on the training data. The configuration achieving the highest auROC score was chosen for the next stages of the pipeline.

To avoid overfitting, we aimed to perform variable selection in the trained sub-models. As LASSO performs a built-in variable selection process through the regularization parameter, and GBT is considered robust to unimportant variables through the splitting and subsampling processes it contains, this procedure was only applied to the NN. We used the hyperparameters selected in the 3-fold CV evaluation and performed permutation tests. We trained the model independently 50 times, each time using a randomly selected 85% of the training set. A reference auROC score was measured by evaluating the NN’s performance on the remaining 15% of the training set (validation set). We then iterated over the variables, performing 20 random permutations of each of the variables in the validation set, while keeping the rest of the data intact. For each permutation, we examined the performance of the model on the data with one permuted variable, and counted the number of times the auROC score was lower than the reference score received under the un-permuted data. This procedure produced, for each variable, a score in the range 0-1000 (since we perform 50 model-

training sessions, and in each we perform 20 permutation rounds followed by prediction and evaluation).

We then repeated the hyperparameter selection procedure, now also adding variable selection. We examined models that considered all the variables, top 50% of the variables (according to the permutations test score), top 25%, top 10% and top 5%. For each variable composition, we performed a hyperparameter search (as described above, by iterating over the different layer sizes and dropout values). Finally, we chose the model parameters and variables that yielded the best auROC score based on 3-fold CV on the training set.

- **L1 Regularized Logistic Regression (LASSO)** – We used a logistic regression model with an $L1$ (LASSO) regularization. We trained 200 models, each with a different regularization parameter in the loglinear range of $(10^{-3} - 10^3)$. We selected the regularization parameter that yielded the highest auROC score based on 3-fold CV on the training-validation set.

At the end of the model selection stage, we obtained a GBT model with optimized hyperparameters, an NN model with optimized hyperparameters and selected variables, and a LASSO model with an optimized regularization parameter.

3. Variable importance

SHAP analysis

Since the three sub-models yield predictions in the continuous range of 0-1, these results can be interpreted as the probability of antibiotic resistance. The SHAP analysis produces for each dataset a baseline score, which is the predicted population prevalence (according to the model's outputs) and a value for each variable in each observation. The sum of the SHAP values of each observation plus the baseline score, equals to the model's prediction. We normalize each SHAP value by the relevant baseline (predicted population prevalence) and calculate the average of the normalized absolute SHAP values to

evaluate the magnitude of the variables' effect over the resistance prediction. We also use the average of the raw, SHAP values (and not their absolute values) in order to deduce the direction of the effect – whether the probability of resistance increases or decreases as a function of a variable's value.

Permutations tests

In order to learn what are the most important variables to the model's prediction success we performed variable importance analysis using permutation tests (see methods). We identified the variables that were consistently identified as contributing the most for the performance (in terms of auROC score) of the ensemble for all five antibiotics. Two such variables had been found, both when excluding and including the identity of the bacterial species: the average previous resistance of the same bacteria (when bacterial species are included in the data) or any bacteria (when bacterial species are excluded) to the same antibiotic, and to any antibiotic. Below are tables showing the 20 most important variables, and their score on each antibiotic dataset.

Table S1. The 20 most important variables according to permutation tests, when bacterial species are excluded.

	Ceftazidime	Gentamicin	Imipenem	Ofloxacin	Sulfamethoxazole-Trimethoprim	mean
Previous any bacterial resistance – specific	0.0377	0.0602	0.021	0.0293	0.0312	0.0359
Previous any bacterial resistance - general	0.0294	0.0114	0.0129	0.0094	0.0095	0.0145
Rectal culture nosocomial	NaN	NaN	0.0505	NaN	NaN	0.0101
Previous use of same-family antibiotic	0.0131	0.005	0.0106	0.0065	0.0021	0.0075
Referral - general ER	0.0035	0.0003	0.0016	0.0233	0.0012	0.006
	0.009	0	0.0086	0	0.0054	0.0046

Independent functioning at arrival	-0.0017	0.0072	0.0093	0.0002	0.0056	0.0041
Urine culture	0.0079	0	0.0006	0.0016	0.0096	0.0039
Complex nursing before arrival	0.0029	0.0042	0	0.0082	0.0015	0.0034
Last hospitalization duration	0.0033	0.0029	0.0093	0.0009	0.0001	0.0033
Arrived from an institution	0.0065	0.0009	0.0008	0.0071	0.0004	0.0031
Oral Ciprofloxacin	0.0014	0.0011	0.0033	0.0005	0.0063	0.0025
Independent functioning before arrival	0.0031	0.002	0.0038	0.002	0.0009	0.0024
Previous use of same antibiotic	0	0	0	0	0.0117	0.0023
Arrived from home	0.0062	0.0045	0	0.001	0	0.0023
Nursing functioning before arrival	0	0	0.0019	0.0069	0.0008	0.0019
Referral - surgical unit B	0.0001	0	0.0068	0	-0.0001	0.0014
Oral Metronidazole	0	0.0001	0.0066	0	-0.0001	0.0013
sepsis	0.0003	-0.0003	0.0024	0.0016	0.0026	0.0013
sex	0.001	0.0035	0.002	0	0	0.0013

Table S2. The 20 most important variables according to permutation tests, when bacterial species are included.

	Ceftazidime	Gentamicin	Imipenem	Ofloxacin	Sulfamethoxazole-Trimethoprim	mean
Previous same bacterial resistance – specific	0.0385	0.0438	0.0133	0.0242	0.0166	0.02728
Previous same bacterial resistance - general	0.0275	0.0206	0.0069	0.0139	0.0112	0.01602
Acinetobacter baumannii	0.0085	0.0089	0.0236	-0.0004	0.0039	0.0089
Rectal culture	NaN	NaN	0.0393	NaN	NaN	0.00786

Staphylococcus aureus	NaN	0.0007	NaN	0	0.0295	0.00604
Previous use of same-family antibiotic	0.0021	-0.0005	0	0.0273	0.0002	0.00582
Acinetobacter sp.	0.0062	0.0074	0.0135	-0.0003	0.0022	0.0058
Pseudomonas aeruginosa	0.0218	0.0025	-0.0022	0	0	0.00442
Nosocomial	0.0059	0.0006	0	0.0082	0.0018	0.0033
Klebsiella pneumoniae	0.0119	-0.0004	0.0001	0.0003	0.0027	0.00292
Klebsiella oxytoca	0	0.0002	0.0121	0.0017	0.0006	0.00292
Previous use of same antibiotic	0	0	0	0	0.0145	0.0029
Admitted at internal medicine A department	0	0.0022	0	-0.0006	0.0128	0.00288
Escherichia coli	0	-0.0001	0.0117	0.001	0.0008	0.00268
Previous any bacterial resistance – general	0	0.001	-0.0001	0.0069	0.0051	0.00258
Arrived from an institution	0.0026	0.0022	0.0001	0.0051	0.0023	0.00246
Independent functioning before arrival	0.0042	0.0021	0	0.002	0.001	0.00186
Micrococcus sp.	NaN	0.0007	NaN	0.0059	0.0023	0.00178
Proteus mirabilis	0.0006	-0.0002	0.0077	0	0.0006	0.00174
Complex nursing functioning before arrival	0.0008	0.0025	0	0.0053	-0.0003	0.00166