

1 **Predicting antibiotic resistance in hospitalized patients by applying**  
2 **machine learning to electronic medical records**

3  
4

5 Ohad Lewin-Epstein<sup>1</sup>, Shoham Baruch<sup>2</sup>, Lilach Hadany<sup>1</sup>, Gideon Y Stein<sup>3,4</sup>, Uri Obolski<sup>2, 5, \*</sup>

6  
7

8 **Affiliations:**

9 <sup>1</sup> Department of Molecular Biology and Ecology of Plants, Tel-Aviv University, Tel-Aviv

10 <sup>2</sup> School of Public Health, Tel-Aviv University, Tel-Aviv

11 <sup>3</sup> Internal Medicine “A”, Meir Medical Center, Kfar Saba

12 <sup>4</sup> Sackler School of Medicine, Tel-Aviv University, Tel-Aviv

13 <sup>5</sup> Porter School of Environmental and Earth Sciences, Tel-Aviv University, Tel-Aviv

14  
15

16 **A running title:** Machine learning predicts AMR

17  
18

19 **Key Words (up to 5):** antibiotic resistance, machine learning, database research, decision support  
20 systems, prediction

21

22 **Corresponding Author:** Uri Obolski, Tel-Aviv University, Tel-Aviv 6997801, Israel

23 ([uriobols@tauex.tau.ac.il](mailto:uriobols@tauex.tau.ac.il))

24  
25

26 **40-word summary:** Machine learning models were applied to large and diverse datasets of medical  
27 records of hospitalized patients, to predict antibiotic resistance profiles of bacterial infections. The  
28 models achieved high accuracy predictions and interpretable results regarding the drivers of antibiotic  
29 resistance.

30

31 **Abstract**

32 **Background:**

33 Computerized decision support systems are becoming increasingly prevalent with advances in data  
34 collection and machine learning algorithms. However, they are scarcely used for empiric antibiotic  
35 therapy. Here we accurately predict the antibiotic resistance profiles of bacterial infections of  
36 hospitalized patients using machine learning algorithms applied to patients' electronic medical records.

37 **Methods:**

38 The data included antibiotic resistance results of bacterial cultures from hospitalized patients, alongside  
39 their electronic medical records. Five antibiotics were examined: Ceftazidime (n=2942), Gentamicin  
40 (n=4360), Imipenem (n=2235), Ofloxacin (n=3117) and Sulfamethoxazole-Trimethoprim (n=3544). We  
41 applied lasso logistic regression, neural networks, gradient boosted trees, and an ensemble combining  
42 all three algorithms, to predict antibiotic resistance. Variable influence was gauged by permutation tests  
43 and Shapely Additive Explanations analysis.

44 **Results:**

45 The ensemble model outperformed the separate models and produced accurate predictions on a test  
46 set data. When no knowledge regarding the infecting bacterial species was assumed, the ensemble  
47 model yielded area under the receiver-operating-characteristic (auROC) scores of 0.73-0.79, for different  
48 antibiotics. Including information regarding the bacterial species improved the auROCs to 0.8-0.88. The  
49 effects of different variables on the predictions were assessed and found consistent with previously  
50 identified risk factors for antibiotic resistance.

51 **Conclusions:**

52 Our study demonstrates the potential of machine learning models to accurately predict antibiotic  
53 resistance of bacterial infections of hospitalized patients. Moreover, we show that rapid information  
54 regarding the infecting bacterial species can improve predictions substantially. The implementation of  
55 such systems should be seriously considered by clinicians to aid correct empiric therapy and to  
56 potentially reduce antibiotic misuse.

57

58

59

60

## 61 Introduction

62 Antibiotic resistance is a major threat to public health. Substantial increases in antibiotic resistance rates  
63 have sprung concerns and bleak estimates as to the future of effective antibiotic treatment [1]. The  
64 emergence of antibiotic resistance is mainly shaped by the evolutionary forces of genetic variation (i.e.  
65 mutations and horizontal gene transfer) and selection exerted by antibiotic usage. Correspondingly,  
66 antibiotic consumption has been repeatedly correlated with increases in antibiotic resistance rates [2].  
67 However, decreases in antibiotic consumption can revert bacterial populations to antibiotic  
68 susceptibility, likely due to the fitness cost that antibiotic resistance incurs [3]. Hence, a straightforward  
69 intervention to reduce the burden of antibiotic resistance is to decrease antibiotic consumption, for  
70 example by reducing inappropriate antibiotic use during empiric therapy [4].

71 Empiric antibiotic therapy is the commencement of antibiotic therapy before a patient's precise  
72 etiology, source of infection, or antibiotic resistance profile of the infecting pathogen, are confirmed [5].  
73 It is both crucial, as immediate action might be necessary, and by definition based on educated guesses,  
74 as it is mostly derived from partial data available to doctors. Two main types of errors occur during  
75 empiric therapy – the prescription of inefficient antibiotics (i.e. the antibiotics prescribed do not clear  
76 the bacterial pathogen due to its resistance to them), or prescription of antibiotics with too broad of a  
77 coverage (i.e. antibiotics with lower coverage would suffice to treat the infection).

78 The first type of error has more immediate and obvious consequences: treatment with inefficient  
79 antibiotics will allow bacteria to keep infecting the patient, putting them at higher risk [6-8].  
80 Furthermore, patients with incorrectly treated infections may be able to keep spreading antibiotic  
81 resistance bacteria, causing even greater harm in the future [9]. The second type of error is perhaps not  
82 as immediately pronounced but could be detrimental to public health in the long-run. High frequency  
83 usage of broad-spectrum antibiotics is likely to increase the frequency of resistance to such antibiotics in  
84 the population, as has been observed repeatedly [10-13], rendering these antibiotics less efficient due  
85 to accumulated levels of resistance in the population. In turn, this can increase the rate of incorrect  
86 empiric therapy of the first kind [14, 15], and lead to increased broad-spectrum antibiotic usage, forming  
87 a positive feedback loop of frequent broad-spectrum antibiotic prescription and increased resistance  
88 [16, 17]. Moreover, patients treated with broad-spectrum antibiotics can have a substantial part of their  
89 microbiome eliminated, enabling subsequent colonization by dangerous and persistent pathogens such  
90 as *Clostridium difficile* [18, 19].

91 A major possible improvement of empiric therapy can stem from using the large amounts of medical  
92 data, which are becoming more accessible, in conjunction with machine learning (ML) algorithms. This

93 approach of integrating ML models based on big medical datasets into medical decision making is  
94 gaining traction lately and is recognized likely being a part of future treatment in many medical fields  
95 [20]. Various studies have identified risk factors for antibiotic resistant infections based on patient  
96 comorbidities, demographics, previous received treatments and other patient characteristics [21].  
97 However, identification of risk factors is not necessarily equivalent to highly accurate prediction. Indeed,  
98 substantially fewer works have produced models trying to predict antibiotic resistance of infecting  
99 bacteria based on patient data. Despite the high quality of many of these studies, they often lacked  
100 large datasets [22-24], were limited to a specific types of infection [24-27], pertained only to few  
101 bacterial species [23], or only to outpatients [25].

102 In this work we used electronic medical records of patients hospitalized in Rabin Medical Center, Israel,  
103 to predict the antibiotic resistance of bacterial infections. The dataset contained over 16,000 antibiotic  
104 resistance tests of bacterial cultures of hospitalized patients with various types of infections, bacterial  
105 species, and examined antibiotics. We applied three ML models, and an ensemble combining their  
106 results, to predict antibiotic resistance of five antibiotics commonly tested for resistance – Ceftazidime,  
107 Gentamicin, Imipenem, Ofloxacin and Sulfamethoxazole-Trimethoprim. We show that accurate  
108 antibiotic resistance prediction is possible by using electronic medical records, and that a substantial  
109 increase in prediction accuracy occurs if information regarding the infecting bacterial species is  
110 available. Finally, we compare the different variables most influencing antibiotic resistance prediction,  
111 and explore their effects on resistance probability using two forms of variable influence analysis of the  
112 ML models.

113

114

## 115 Results

116 We retrieved medical records of patients which had positive bacterial culture results, from Rabin  
 117 Medical Center in Israel, from the period between May 2013 and December 2015. The dataset included  
 118 the bacterial species isolated from the patients and their resistance profiles to the antibiotics tested, as  
 119 well as the patients' demographics, comorbidities, hospitalization records and previous antibiotic usage  
 120 within the hospital (see Methods). In this study we focused on predicting resistance to the five  
 121 antibiotics most commonly tested for resistance in our dataset: Ceftazidime, Gentamicin, Imipenem,  
 122 Ofloxacin and Sulfamethoxazole-Trimethoprim (Sul-Trim). All data in the table are aggregated across  
 123 unique samples.  
 124

	<b>Ceftazidime</b>	<b>Gentamicin</b>	<b>Imipenem</b>	<b>Ofloxacin</b>	<b>Sulfamethoxazole-Trimethoprim</b>
Samples, n	2942	4360	2235	3117	3544
Resistance, %	42	32	16	47	50
Age, mean (sd), years	72 (16)	72 (16)	72 (16)	72 (17)	72 (16)
Female, %	42	41	40	43	42
Most common bacterial species	Escherichia coli (29%)	Escherichia coli (20%)	Escherichia coli (22%)	Escherichia coli (22%)	Escherichia coli (24%)
Second-most common bacterial species	Klebsiella pneumoniae (18%)	Klebsiella pneumoniae (12%)	Pseudomonas aeruginosa (18%)	Staphylococcus coag. neg. group (16%)	Klebsiella pneumoniae (15%)
Third-most common bacteria species	Pseudomonas aeruginosa (14%)	Staphylococcus coag. neg. group (12%)	Klebsiella pneumoniae (16%)	Klebsiella pneumoniae (13%)	Staphylococcus coag. neg. group (14%)
Latest hospitalization duration,	6.1 (10.4)	6.1 (10.2)	7.1 (11.4)	6.1 (10.4)	5.9 (10.1)

mean (std),					
days					

125

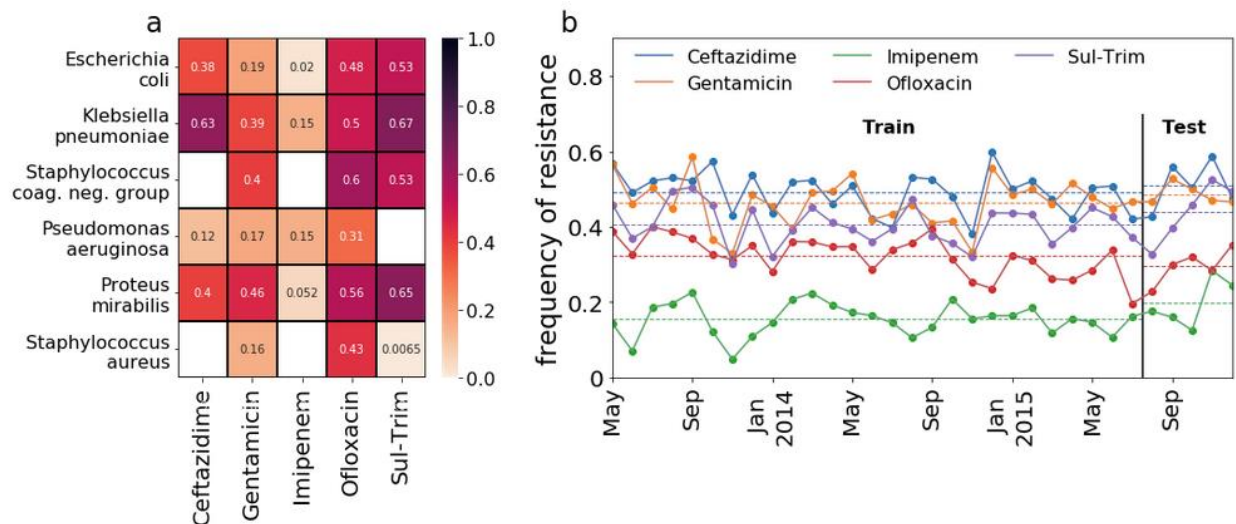
126 **Table 1. Summary statistics of the dataset.**

127

128 We split the data into a training set and a test set. The first 85% of each dataset (according to the  
 129 sample date) were used for training the models while the rest 15% were used for testing the models.

130 We found varying frequencies of antibiotic-resistance, between antibiotics and different bacterial  
 131 species. The frequencies of antibiotic-resistance also fluctuated through time, yet average resistance  
 132 frequencies remained similar in the training and test sets (see Figure 1).

133



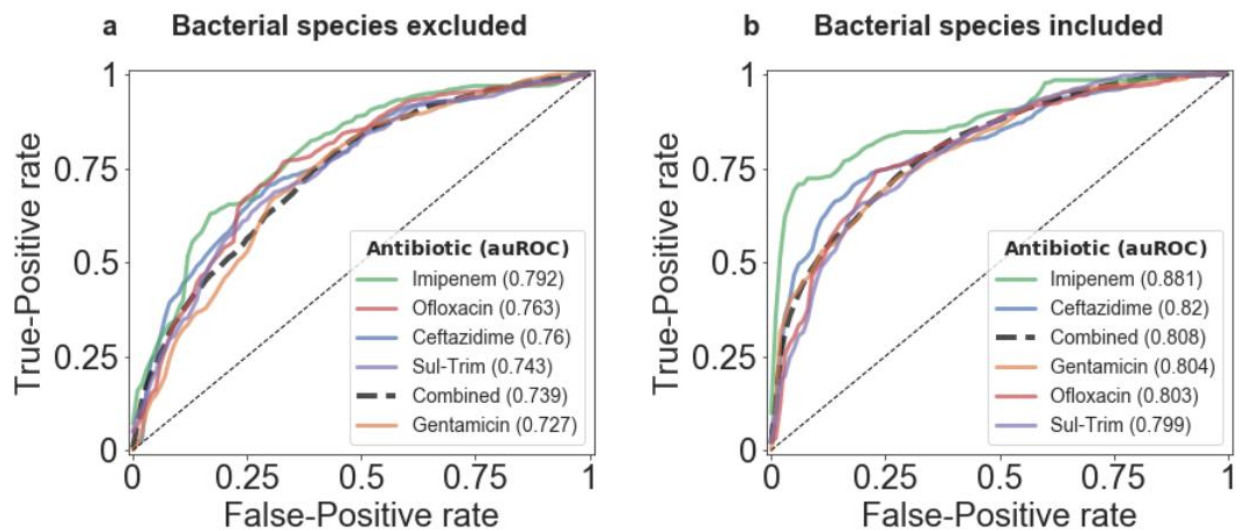
134

135 **Figure 1. Frequency of antibiotic resistance. (a)** A heatmap showing the frequencies of antibiotic  
 136 resistance for each antibiotic and bacterial species combination. Empty cells represent combinations for  
 137 which there were fewer than 100 data points. **(b)** A time series plot of the frequency of antibiotic  
 138 resistance observed in each month, for each antibiotic, across all bacterial species. Horizontal dashed  
 139 lines represent the average resistance frequencies of each antibiotic, separately for the training set and  
 140 the test set.

141

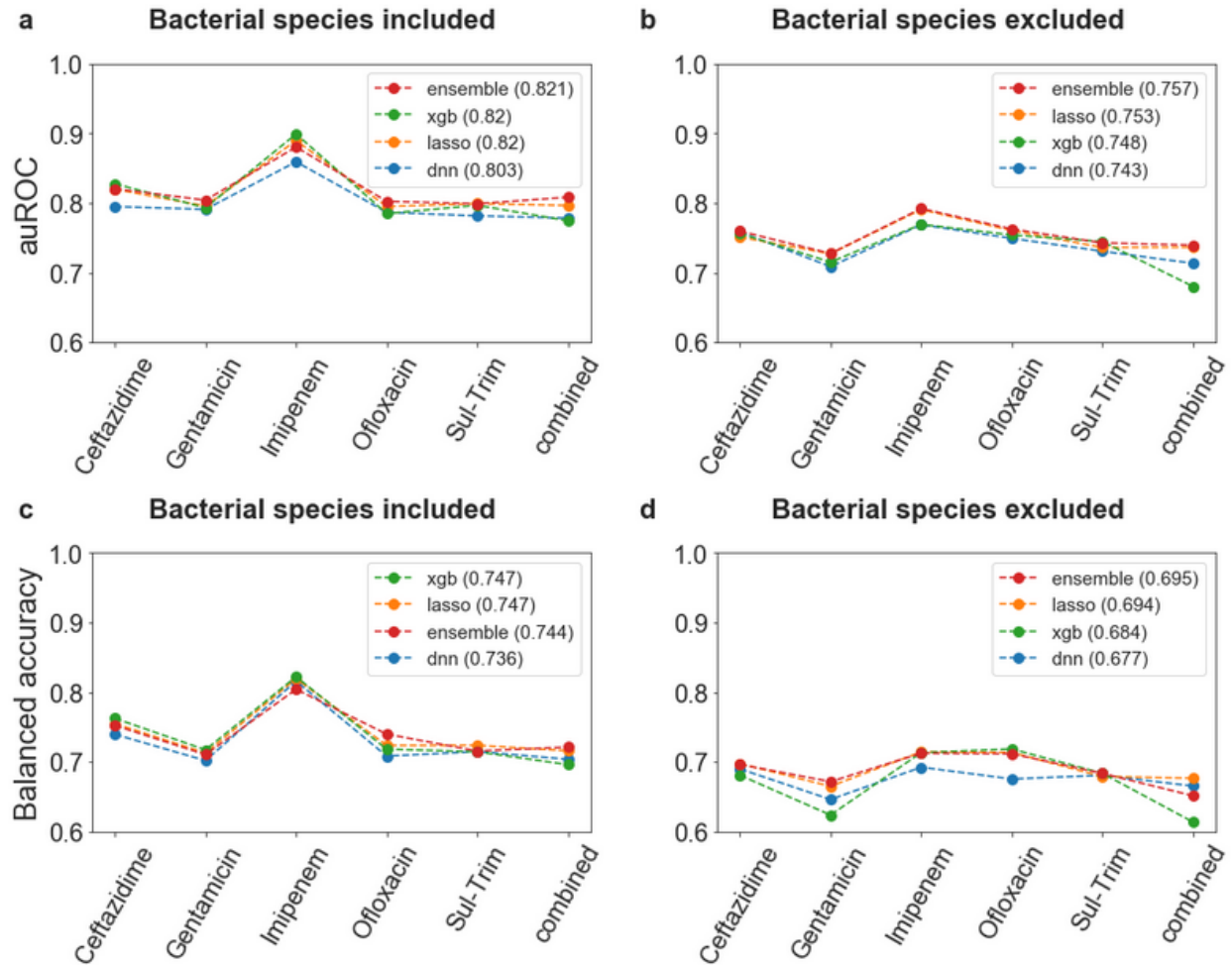
142 We applied ML algorithms to the data in order to generate models for predicting antibiotic resistance of  
 143 bacterial cultures. We used a supervised ML approach to classify each isolated bacterial culture as either  
 144 susceptible or resistant to each antibiotic (see Methods and Supplementary Material 2). The final model  
 145 chosen for predicting antibiotic resistance was an ensemble, composed of three sub-models: L1

146 regularized logistic regression (LASSO), gradient-boosted decision trees and neural network. Each sub-  
147 model was trained separately, and the ensemble provided a prediction based on the average predictions  
148 of the sub-models. We examined the success of the ensemble in predicting antibiotic resistance in two  
149 data conformations: one where the ensemble was trained and evaluated separately on each antibiotic,  
150 and another where the ensemble was trained and evaluated on data containing all five antibiotics,  
151 combined. In addition, training and testing of the ensemble was performed once on a dataset that  
152 included the identity of the isolated bacterial species, and once on the same data, barring the identity of  
153 the isolated bacterial species.  
154 The ensemble achieved high classification success both in terms of area under the receiver-operating-  
155 characteristic curve (auROC) and balanced accuracy (i.e., the unweighted average of the sensitivity and  
156 specificity rates; see Figures 2 and 3). In addition, the ensemble was found to slightly outperform the  
157 sub-models in most scenarios, especially when the identity of the isolated bacterial species was included  
158 in the data (see Figure 3).  
159



160  
161 **Figure 2. Receiver operating characteristic (ROC) curves of the ensemble.** ROC curves of the ensemble  
162 are presented separately for each antibiotic and for all antibiotics combined, for the datasets excluding  
163 (a) and including (b) the bacterial species' identity. The legends show the area under the ROC curve  
164 (auROC) for each antibiotic, ordered from highest to lowest.  
165





166  
167 **Figure 3. auROC and balanced accuracy scores of the ensemble and its sub-models.** The auROC (a, b)  
168 and the balanced accuracy (c, d) of the ensemble and its three sub-models, based on data including the  
169 identity of the bacterial species (a, c) and on data excluding it (b, d). The legends show the score of each  
170 model, averaged over the five antibiotics, ordered from highest to lowest.

171  
172 In contrast to classic statistical methods such as regression analysis, the influence of variables on model  
173 output is often difficult to gauge in ML models such as boosted trees and neural networks. We thus  
174 performed two types of analysis to determine the influence of variables on our ensemble model  
175 predictions. First, we performed a permutation-based variable importance analysis (see Methods).  
176 Briefly, each variable was randomly permuted to break its association with the outcome. Then,  
177 predictions were made using the new dataset with the permuted variable, and the change in the  
178 ensemble's auROC was recorded. Variables for which permutations resulted in substantial decreases in  
179 auROC were deemed important. This analysis revealed that the two variables with the highest average

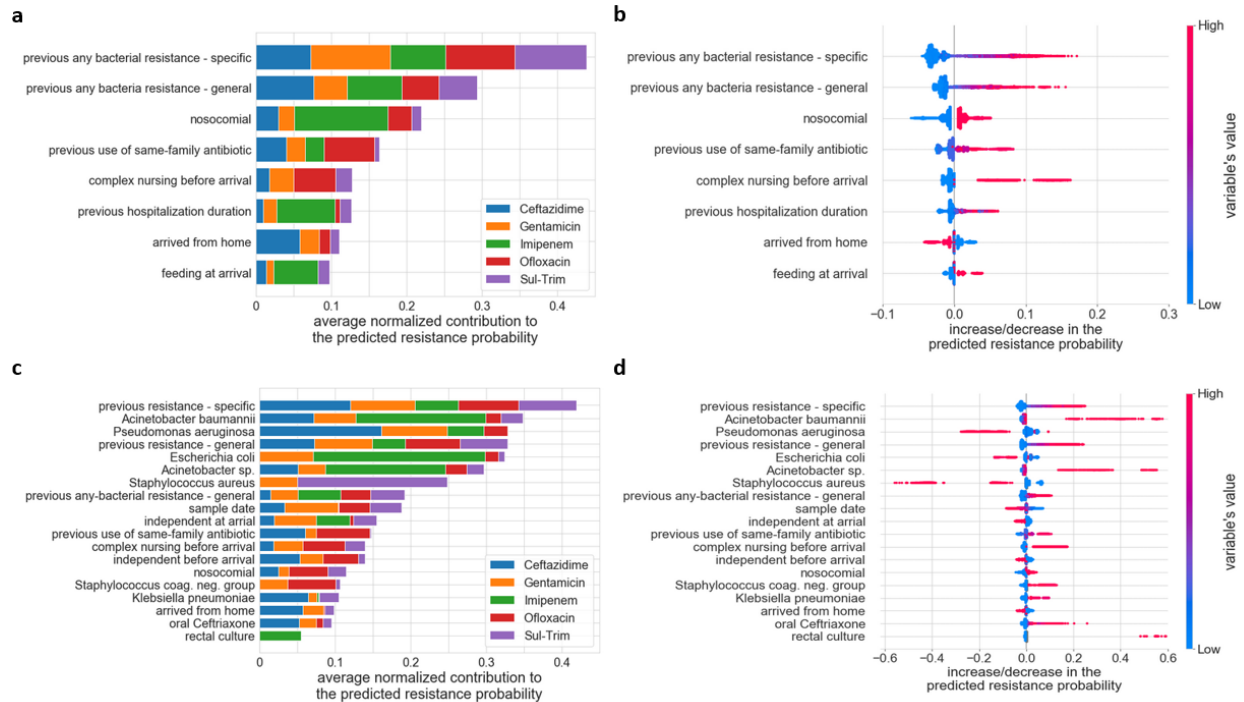


180 effect (across all five antibiotics) were the proportion of past antibiotic resistance infections: previous  
181 same-bacterial species resistance to the same antibiotic (previous resistance - specific), and to any  
182 antibiotic (previous resistance - general) when including information of the bacterial species; previous  
183 any-bacterial species resistance to the same antibiotic (previous any-bacteria resistance – specific), and  
184 to any antibiotic (previous any-bacteria resistance – general) when excluding information about the  
185 infecting bacterial species (see Supplementary Material 3 and Supplementary tables S1 and S2).

186 Furthermore, we performed a Shapley Additive Explanations (SHAP) analysis [28] (see Methods). The  
187 SHAP analysis allowed us to estimate the marginal contribution of each variable to the final prediction of  
188 the ensemble. We performed the SHAP analysis separately for each of the five antibiotics tested, both  
189 with and without the information regarding the infecting bacterial species. We present the variables  
190 having a substantial contribution to prediction of antibiotic resistance (as defined in the Methods) for all  
191 five antibiotics in Figure 4. When information regarding the bacterial species was excluded, the two top-  
192 contributing variables were consistent with the permutation-based importance analysis – previous any-  
193 bacteria resistance - specific and general. These were followed by variables indicating whether the  
194 infection was nosocomial or community acquired, and whether the patient was previously treated in the  
195 hospital with antibiotics of the same family (antibiotics were categorized into: beta-lactams,  
196 fluoroquinolones, aminoglycosides, sulfonamides). Other important variables were the patients’  
197 functioning and independence levels, and previous hospitalization duration. Similarly, when including  
198 data regarding the bacterial species, the average previous resistance of the same bacteria to the  
199 same/any antibiotic (previous resistance – specific/general, respectively) remained in the top-most  
200 affecting variables, alongside indicator-variables of the infecting bacterial species.

201 The SHAP analysis also allowed us to investigate whether the different variables in our model act to  
202 decrease or increase the probability of antibiotic resistance (Figure 4b,d). Reassuringly, the probability  
203 of resistance in our model increased in accordance to known risk factors of antibiotic resistance:  
204 previous antibiotic resistant infections, previous hospitalizations, nosocomial infections, previous  
205 antibiotic usage, location of sample derivation, and contraindications of patient independence (e.g.  
206 nursing home residence and dependence in feeding) [23, 25, 27, 29, 30]. When including information of  
207 the infecting bacterial species, additional patterns emerge. For example, while the presence of  
208 *Acinetobacter Baumannii* in cultures increases the probability of resistance, *Staphylococcus Aureus*  
209 decreases it. The patients’ sex was found to have only a minor effect over the resistance probability,  
210 with increased probability of resistance for males. The sample date, which was coded as a numeric

211 variable from the date of the earliest culture in the dataset, was also found to have some effect,  
 212 probably due to the fluctuating resistance frequencies through time captured by our model.  
 213



214  
 215 **Figure 4. Variable importance analysis using Shapley Additive Explanations (SHAP).** (a, c) The absolute  
 216 marginal contribution to predicted probabilities, normalized by the predicted population resistance  
 217 prevalence (x-axis) is plotted for each antibiotic (color-coded), both for data excluding information on  
 218 the bacterial species (a) and including it (c). The presented variables are those with an effect of at least  
 219 0.05 in any of the five antibiotics. (b, d) The marginal changes in predicted resistance probability derived  
 220 from the same variables shown in (a, c), respectively, are plotted for all antibiotics combined, both for  
 221 data excluding information on the infected bacterial species (b) and including it (d). Each row in panels  
 222 (b,d) shows the distribution of the data in two dimensions, and each dot represents one sample: The  
 223 color represents the value of the variable in a schematic scale from low value to high value (binary  
 224 variables are represented by the two colors on the edges of the color-bar); the position on the x-axis  
 225 represents the marginal change in probability of antibiotic resistance due to the variables' value.

226

227

## 228 Discussion

229 ML is widely applied in various fields of medicine and is likely to become an invaluable part of medical  
230 decision making and treatment [20]. However, ML is scarcely used in aiding the decision of empiric  
231 antibiotic therapy. Only a handful of studies have previously utilized the prediction abilities of machine  
232 learning models for the rapid detection of antibiotic resistance from patient medical records.

233 Our work demonstrates the ability to predict antibiotic resistance from patient medical records with  
234 high accuracy, and extends previous research in the field in several ways: Rather than relying on a single  
235 algorithm, we use an ensemble combining several algorithms substantially differing in their underlying  
236 prediction methods (logistic regression, boosted decision trees and neural networks) to produce high-  
237 accuracy, robust results. Importantly, we perform a controlled procedure of hyper parameter selection  
238 on a training subset of the data and then continue to test our predictions on a disjoint, previously  
239 unexplored subset of the data. Furthermore, we predict antibiotic resistance on a large and  
240 heterogeneous dataset. It comprises more than 16,000 antibiotic resistance tests of bacterial cultures of  
241 hospitalized patients, tested for various antibiotics, and containing multiple bacterial species and  
242 infection sites.

243 Despite the heterogeneity of our data, we were able to train models that achieved highly competitive  
244 results: if information regarding the infected bacterial species was excluded we obtained auROC scores  
245 in the range of 0.73-0.79, while including the bacterial species yielded an even higher auROC scores in  
246 the range of 0.8-0.88. Previous studies which included information regarding the infecting bacterial  
247 species obtained auROC scores in the ranges of 0.6-0.83 for antibiotics comparable to those examined in  
248 our dataset [27, 31]. Other studies, restricted to one bacterial species or to only one type of infection,  
249 had auROC scores in the range of 0.7-0.83 [23-25]. Even when previous auROC results were comparable  
250 to those achieved in our study, previous studies did not have such a heterogenic dataset, containing  
251 patients with different infections, bacterial species and antibiotics. This added a substantial challenge,  
252 which was successfully tackled by our models, and is likely to decrease predictive power of methods  
253 used in other studies.

254 Finally, despite the complex nature of the models used, further complicated by their combination into  
255 an ensemble model, we were able to provide interpretation of the influence of different variables on the  
256 ensemble's predictions. Reassuringly, most of the variables found influential in our analysis have been  
257 previously identified as increasing risk for antibiotic resistant infections. In addition to further validating  
258 our model against prior knowledge, understating which variables are influential can help indicate  
259 important drivers of antibiotic resistance. For example, the variables consistently highly ranked as

260 important in our models were those pertaining to previously resistant bacterial cultures. The importance  
261 of those variables might imply the persistence of resistant bacterial flora in patients, and may warrant  
262 further investigations into treatments restoring the normal bacterial flora after antibiotic treatments,  
263 especially in patients at risk for re-hospitalization [32]. However, such conclusions should be further  
264 investigated in other settings before concrete conclusions can be reached.

265 An especially important variable in our models was the identity of the bacterial species causing the  
266 infection. It is plausible that inherent biological differences, combined with different exposures to  
267 antibiotics, produce the substantial observed differences between resistance frequencies of different  
268 bacterial species (Figure 1), and hence rendered this information predictive in our models. Although not  
269 routinely performed in most hospitals, very rapid identification of the infecting bacterial species is  
270 possible, e.g. through PCR-based methods [33]. If our predictive model is to be implemented in real time  
271 in clinical settings, adding such rapid bacterial species identification tools might be cost-beneficial, given  
272 the improvement in our prediction results, and the major cost incurred by antibiotic resistant infections  
273 [34].

274 Additional potentially important variables are various community-derived risk factors for antibiotic  
275 resistant infections such as antibiotic use outside the hospital [25, 35, 36], patient location of residency  
276 [12], and other factors as microbiome composition, diet and exercise [37-40]. Unfortunately, these were  
277 not available to us and hence not included in our analysis. Our results can likely be improved by  
278 inclusion of such data, and future work should consider their inclusion when available.

279 To conclude, our results present an ML approach to predict antibiotic resistance of bacterial infections  
280 of hospitalized patients, using the patient's electronic medical record. Our method autonomously  
281 identified known risk factors of antibiotic resistance, and provided high-accuracy predictions based on  
282 the complex interactions between them and other patient information. Applying our approach, and  
283 further developing it by incorporating additional patient data is paramount for achieving highly  
284 informed, personalized empiric antibiotic therapy. Such therapy should result in less antibiotic misuse,  
285 and hopefully aid the fight against antibiotic resistance.

286

287

## 288 **Methods**

### 289 **Data processing and variable engineering**

290 We processed the datasets described in the Results section and in Table 1, and created variables that  
291 were fed into the machine learning models. All categorical variables were one-hot encoded as dummy  
292 variables. A total of 322-448 variables were included in the processed dataset (depending on the  
293 categorical variables in each dataset, and on whether bacterial species information was included). A  
294 description of the variables that were concluded most important is available at Supplementary material  
295 1. We considered intermediate resistance results as resistant for both simplicity of classification and  
296 under a conservative rationale, as is common practice [25].

297

### 298 **Machine learning models**

299 We used an ensemble model, composed of three sub-models: L1 regularized logistic regression (LASSO),  
300 gradient-boosted decision trees and neural network. Each sub-model was trained separately on the data  
301 and for each data point it provided a prediction: a number in the range 0-1. The ensemble prediction is  
302 the average of the predictions of these three models (Supplementary Material 2).

303 Each data point input into the models contained the abovementioned variables as well as the  
304 susceptible or resistant binary label for a single antibiotic as the predicted variable. For each antibiotic,  
305 the data were divided to distinct training and test sets based on dates, so that the early 85% of the data  
306 were assigned to the training set.

307

### 308 **Model evaluation**

309 After performing the hyperparameter tuning and variable selection on the training set (using cross  
310 validation), the chosen models were applied to the test set. The ensemble's predictions were compared  
311 to the actual resistance class of each data point to derive the auROC score.

312 When computing the balanced accuracy score, we used the first 85% of the training set in order to train  
313 the ensemble, and the remaining 15% of the training set (validation set) in order to find the optimized  
314 prediction threshold  $\rho \in (0,1)$ , under which each prediction was assigned to 1 if above  $\rho$  and to 0 if  
315 below  $\rho$ . The thresholds were chosen to maximize the balanced accuracy score on the validation set. We  
316 then trained the ensemble on the entire training set and used it to predict results on the test set. We  
317 then compared the ensemble's predictions to the actual class of each data point, and derived the  
318 balanced accuracy score.

319

320 **Variable importance analysis – permutation tests**

321 Each variable's values were randomly permuted in the test set, while other variables kept as they were.  
322 After a permutation was performed, the auROC of the ensemble's prediction on the test set was re-  
323 calculated. The absolute difference of the resulting auROC from that obtained on the original test set  
324 was recorded. This was repeated 100 times (for each variable) and the average result was deemed as  
325 each variable's importance score.

326

327 **Variable influence analysis – Shapley Additive Explanations (SHAP)**

328 We applied the SHAP analysis to the training sets, performing the analysis separately for each sub-  
329 model, and then averaging the results to obtain the ensemble's scores (supplementary material 3). The  
330 SHAP values can be interpreted as the marginal change in the probability of resistance of each  
331 observation, derived for each variable.

332

333 **Software used**

334 All analyses were performed using Python 3.6.

335

336

337

338 **Competing interests**

339 We declare we have no competing interests.

340

341 **Funding**

342 This project was supported by the Clore Foundation Scholars Programme (OLE).

343

344 **Acknowledgements**

345 We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU  
346 used for this research.

347

## 348 **References**

- 349 1. Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and*  
350 *therapeutics* **2015**; 40(4): 277.
- 351 2. Bell BG, Schellevis F, Stobberingh E, Goossens H, Pringle M. A systematic review and meta-  
352 analysis of the effects of antibiotic consumption on antibiotic resistance. *BMC infectious*  
353 *diseases* **2014**; 14(1): 13.
- 354 3. Melnyk AH, Wong A, Kassen R. The fitness costs of antibiotic resistance mutations. *Evolutionary*  
355 *applications* **2015**; 8(3): 273-83.
- 356 4. Laxminarayan R, Duse A, Wattal C, et al. Antibiotic resistance—the need for global solutions. *The*  
357 *Lancet infectious diseases* **2013**; 13(12): 1057-98.
- 358 5. Mandell G, Dolin R, Bennett J. *Mandell, Douglas, and Bennett's principles and practice of*  
359 *infectious diseases*: Elsevier, **2009**.
- 360 6. Paul M, Shani V, Muchtar E, Kariv G, Robenshtok E, Leibovici L. Systematic review and meta-  
361 analysis of the efficacy of appropriate empiric antibiotic therapy for sepsis. *Antimicrobial agents*  
362 *and chemotherapy* **2010**; 54(11): 4851-63.
- 363 7. Oshima T, Kodama Y, Takahashi W, et al. Empiric antibiotic therapy for severe sepsis and septic  
364 shock. *Surgical infections* **2016**; 17(2): 210-6.
- 365 8. Ibrahim EH, Sherman G, Ward S, Fraser VJ, Kollef MH. The influence of inadequate antimicrobial  
366 treatment of bloodstream infections on patient outcomes in the ICU setting. *Chest* **2000**; 118(1):  
367 146-55.
- 368 9. Obolski U, Stein GY, Hadany L. Antibiotic restriction might facilitate the emergence of multi-drug  
369 resistance. *PLoS computational biology* **2015**; 11(6).
- 370 10. Paterson DL. “Collateral damage” from cephalosporin or quinolone antibiotic therapy. *Clinical*  
371 *Infectious Diseases* **2004**; 38(Supplement\_4): S341-S5.
- 372 11. Vernaz N, Huttner B, Musciconico D, et al. Modelling the impact of antibiotic use on antibiotic-  
373 resistant *Escherichia coli* using population-based data from a large hospital and its surrounding  
374 community. *Journal of Antimicrobial Chemotherapy* **2011**; 66(4): 928-35.
- 375 12. Low M, Neuberger A, Hooton TM, et al. Association between urinary community-acquired  
376 fluoroquinolone-resistant *Escherichia coli* and neighbourhood antibiotic consumption: a  
377 population-based case-control study. *The Lancet Infectious Diseases* **2019**; 19(4): 419-28.
- 378 13. Pantosti A, Moro ML. *Antibiotic use: the crystal ball for predicting antibiotic resistance*. The  
379 University of Chicago Press, **2005**.



- 380 14. Merli M, Lucidi C, Di Gregorio V, et al. The spread of multi drug resistant infections is leading to  
381 an increase in the empirical antibiotic treatment failure in cirrhosis: a prospective survey. *PLoS*  
382 *One* **2015**; 10(5).
- 383 15. Carrara E, Pfeffer I, Zusman O, Leibovici L, Paul M. Determinants of inappropriate empirical  
384 antibiotic treatment: systematic review and meta-analysis. *International journal of antimicrobial*  
385 *agents* **2018**; 51(4): 548-53.
- 386 16. Kollef MH. Appropriate empirical antibacterial therapy for nosocomial infections. *Drugs* **2003**;  
387 63(20): 2157-68.
- 388 17. Murthy R. Implementation of strategies to control antimicrobial resistance. *Chest* **2001**; 119(2):  
389 405S-11S.
- 390 18. Crowther GS, Wilcox MH. Antibiotic therapy and *Clostridium difficile* infection—primum non  
391 nocere—first do no harm. *Infection and drug resistance* **2015**; 8: 333.
- 392 19. Fridkin S, Baggs J, Fagan R, et al. Vital signs: improving antibiotic use among hospitalized  
393 patients. *MMWR Morbidity and mortality weekly report* **2014**; 63(9): 194.
- 394 20. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *New England Journal of Medicine*  
395 **2019**; 380(14): 1347-58.
- 396 21. Control CfD, Prevention. Antibiotic resistance threats in the United States, 2013: Centres for  
397 Disease Control and Prevention, US Department of Health and ..., **2013**.
- 398 22. Oonsivilai M, Mo Y, Luangasanatip N, et al. Using machine learning to guide targeted and locally-  
399 tailored empiric antibiotic prescribing in a children's hospital in Cambodia. *Wellcome open*  
400 *research* **2018**; 3.
- 401 23. Sullivan T, Ichikawa O, Dudley J, Li L, Aberg J. The rapid prediction of carbapenem resistance in  
402 patients with *Klebsiella pneumoniae* bacteremia using electronic medical record data. In: *Open*  
403 *forum infectious diseases*: Oxford University Press US, 2018:ofy091.
- 404 24. Dan S, Shah A, Justo JA, et al. Prediction of fluoroquinolone resistance in Gram-negative bacteria  
405 causing bloodstream infections. *Antimicrobial agents and chemotherapy* **2016**; 60(4): 2265-72.
- 406 25. Yelin I, Snitser O, Novich G, et al. Personal clinical history predicts antibiotic resistance of urinary  
407 tract infections. *Nature medicine* **2019**; 25(7): 1143-52.
- 408 26. Dickstein Y, Geffen Y, Andreassen S, Leibovici L, Paul M. Predicting antibiotic resistance in  
409 urinary tract infection patients with prior urine cultures. *Antimicrobial agents and*  
410 *chemotherapy* **2016**; 60(8): 4717-21.

- 411 27. Vazquez-Guillamet MC, Vazquez R, Micek ST, Kollef MH. Predicting resistance to piperacillin-  
412 tazobactam, cefepime and meropenem in septic patients with bloodstream infection due to  
413 Gram-negative bacteria. *Clinical Infectious Diseases* **2017**; 65(10): 1607-14.
- 414 28. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: *Advances in*  
415 *neural information processing systems*, 2017:4765-74.
- 416 29. Chatterjee A, Modarai M, Naylor NR, et al. Quantifying drivers of antibiotic resistance in  
417 humans: a systematic review. *The Lancet Infectious Diseases* **2018**; 18(12): e368-e78.
- 418 30. MacFadden D, Coburn B, Shah N, et al. Utility of prior cultures in predicting antibiotic resistance  
419 of bloodstream infections due to Gram-negative pathogens: a multicentre observational cohort  
420 study. *Clinical Microbiology and Infection* **2018**; 24(5): 493-9.
- 421 31. Tandan M, Timilsina M, Cormican M, Vellinga A. Role of patient descriptors in predicting  
422 antimicrobial resistance in urinary tract infections using a decision tree approach: A  
423 retrospective cohort study. *International journal of medical informatics* **2019**; 127: 127-33.
- 424 32. Francino M. Antibiotics and the human gut microbiome: dysbioses and accumulation of  
425 resistances. *Frontiers in microbiology* **2016**; 6: 1543.
- 426 33. Järvinen A-K, Laakso S, Piiparinen P, et al. Rapid identification of bacterial pathogens using a  
427 PCR-and microarray-based assay. *BMC microbiology* **2009**; 9(1): 161.
- 428 34. Dadgostar P. Antimicrobial Resistance: Implications and Costs. *Infection and Drug Resistance*  
429 **2019**; 12: 3903.
- 430 35. Costelloe C, Metcalfe C, Lovering A, Mant D, Hay AD. Effect of antibiotic prescribing in primary  
431 care on antimicrobial resistance in individual patients: systematic review and meta-analysis. *Bmj*  
432 **2010**; 340: c2096.
- 433 36. Pouwels KB, Freeman R, Muller-Pebody B, et al. Association between use of different antibiotics  
434 and trimethoprim resistance: going beyond the obvious crude association. *Journal of*  
435 *Antimicrobial Chemotherapy* **2018**; 73(6): 1700-7.
- 436 37. Sommer MO, Church GM, Dantas G. The human microbiome harbors a diverse reservoir of  
437 antibiotic resistance genes. *Virulence* **2010**; 1(4): 299-303.
- 438 38. Baron SA, Diene SM, Rolain J-M. Human microbiomes and antibiotic resistance. *Human*  
439 *Microbiome Journal* **2018**; 10: 43-52.
- 440 39. Corpet DE. Antibiotic resistance from food. *The New England journal of medicine* **1988**; 318(18):  
441 1206.

- 442 40. Mascaro V, Capano MS, Iona T, Nobile CGA, Ammendolia A, Pavia M. Prevalence of  
443 Staphylococcus aureus carriage and pattern of antibiotic resistance, including methicillin  
444 resistance, among contact sport athletes in Italy. *Infection and drug resistance* **2019**; 12: 1161.  
445