Research Article

Hakan Ayyıldız* and Seda Arslan Tuncer

Is it possible to determine antibiotic resistance of *E. coli* by analyzing laboratory data with machine learning?

[Makine öğrenimi ile laboratuvar verilerini analiz ederek *E. coli*'nin antibiyotik direncini belirlemek mümkün müdür?]

<https://doi.org/10.1515/tjb-2021-0040>

Received February 22, 2021; accepted August 11, 2021;

published online August 30, 2021

Abstract

Objectives: Microbial antibiotic resistance remains a serious public health problem worldwide. Conventional culture-based techniques are time-taking procedures; therefore, there is need for new approaches for detecting bacterial resistance. The aim of this study was to assess antibiotic resistance of *Escherichia coli* by analyzing biochemical parameters with machine learning systems without using antibiogram.

Material and methods: In this article, machine learning systems such as K-Nearest Neighbors, Artificial Neural Networks (ANN), Support Vector Machine and Decision Tree Learning were used to investigate whether *E. coli* is sensitive or resistant to antibiotics. The study was conducted based on the clinical records of 103 patients who were previously diagnosed with *E. coli* infection, including CBC and complete UA results, and CRP values.

Results: The accuracy rates of antibiotic resistance/susceptibility detected by ANN were as follows: Amikacin (96.0%), Ampicillin (77%), Cefotaxime (62%), Cefixime (63%), Cefotaxime (68%), Colistin (95%), Ciprofloxacin

(76%), Cefepime (70%), Ertapenem (96%), Nitrofurantoin (90%), Phosphomycin (98%), Gentamicin (84%), Levofloxacin (98%), Piperacillin-Tazobactam (92%), and Trimethoprim-Sulfadiazine (79%).

Conclusions: The study determined the antibiotic resistance of *E. coli* with less time and cost compared to conventional culture-based methods machine learning based model contributes positively to artificial intelligence (AI) supported decision-making processes in laboratory medicine.

Keywords: antibiotic resistance; diagnostic decision making; laboratory medicine; machine learning; urinary tract infection.

Özet

Amaç: Mikrobiyal antibiyotik direnci dünya çapında ciddi bir halk sağlığı sorunu olmaya devam etmektedir. Geleneksel kültüre dayalı teknikler zaman alan prosedürlerdir; bu nedenle bakteri direncini tespit etmek için yeni yaklaşımlara ihtiyaç vardır. Bu çalışmanın amacı, antibiyogram kullanmadan makine öğrenme sistemleri ile biyokimyasal parametreleri analiz ederek *E. coli*'nin antibiyotik direncini değerlendirmektir.

Gereç ve Yöntemler: Bu makalede, *E. coli*'nin antibiyotiklere duyarlı mı yoksa dirençli mi olduğunu araştırmak için K-En Yakın Komşu, Yapay Sinir Ağları, Destek Vektör Makinesi ve Karar Ağacı Öğrenme gibi makine öğrenme sistemleri kullanılmıştır. Çalışma, daha önce *E. coli* enfeksiyonu tanısı almış 103 hastanın CBC, TİT sonuçları ve CRP değerleri dahil klinik kayıtlarına dayanılarak yapılmıştır.

Bulgular: ANN ile tespit edilen antibiyotik direnç/duyarlılık doğruluk oranları şu şekildedir: Amikasin (%96),

*Corresponding author: Hakan Ayyıldız, Department of Biochemistry, Fethi Sekin City Hospital, Elazığ, Turkey, E-mail: hknayildiz@hotmail.com. <https://orcid.org/0000-0002-3133-9862>

Seda Arslan Tuncer, Department of Software Engineering, Firat University, Elazığ, Turkey, E-mail: satuncer@firat.edu.tr. <https://orcid.org/0000-0001-6472-8306>

Ampisilin (%77), Seftazidim (%62), Sefksim (%63), Sefotaksim (%68), Kolistin (%95), Siprofloksasin (%76), Sefepim (%70), Ertapenem (%96), Nitrofurantoin (%90), Fosfomisin (%98), Gentamisin (%84), Levofloksasin (%98), Piperasilin-Tazobaktam (%92) ve Trimetoprim-Sulfadiazin (%79).

Sonuç: Çalışma, geleneksel kültür temelli yöntemlere göre daha az zaman ve maliyetle *E. coli*'nin antibiyotik direncini belirleyerek, makine öğrenmesi temelli modelin, laboratuvar tıbbında yapay zeka destekli karar verme süreçlerine olumlu katkı sağladığını belirlemiştir.

Anahtar Kelimeler: İdrar yolu enfeksiyonu; makine öğrenimi; laboratuvar tıbbı; tanısal karar verme; antibiyotik direnci.

Introduction

Antibiotic-resistant bacteria is a serious public health problem worldwide due to its potential in reducing the likelihood of treatment and even making it impossible. Infections caused by these bacteria often lead to higher rates of hospitalization and additional therapies as well as increased diagnostic and treatment costs. In the United States, almost 2.8 million people are diagnosed with antibiotic-resistant organisms each year and more than 35,000 of them are reported to die [1].

The growing incidence of antibiotic resistance has increased the variety of infections as well as the cost of additional treatments [2]. These processes are primarily affected by excessive and unnecessary use of antibiotics and easy access to these drugs [3].

Antibiotic resistance in bacteria differs between Gram negative and Gram positive bacteria [4, 5]. Complex mechanisms of antibiotic resistance include (i) natural (intrinsic) resistance (caused by non-target antibiotics), (ii) acquired resistance (caused by mutations, plasmids, or transposons), (iii) cross-resistance (i.e., cross-resistance of a drug-resistant organism to drugs with similar effects), and (iv) multidrug resistance (i.e., resistance caused by multiple imported genes or enzymatic inactivation and structural changes) [6].

Urinary tract infections (UTI) are the most common bacterial infections Gram negative bacteria are the leading cause of UTI in all age groups and in both sexes, with *Escherichia coli* (*E. coli*) being the most common UTI pathogen (65–75%) [7]. Assessment of antibiotic resistance by conventional methods can be a time-consuming process that comprises the following steps:

(1) Collection of urine specimens and pre-analytical processing of specimens in the laboratory (0–1 h),

- (2) Addition of the specimen to the medium (Eosin Methylene-blue [EMB] Lactose Sucrose Agar for *E. coli*),
- (3) Incubation of the medium (approximately 24 h),
- (4) Identification of the bacteria following culture growth (approximately 4–6 h for Gram negative bacteria),
- (5) Performing antibiogram to determine the antibiotic resistance of the identified bacteria (approximately 16 h for *E. coli*).

In total, these five steps take approximately two days to implement. For this reason, computer-aided machine learning algorithms are needed to reduce this period and to support the decision-making processes. Moreover, reducing the analytical process is highly important for taking prompt measures both in the diagnosis and treatment of the patients.

The present study was designed to assess antibiotic resistance of *E. coli* by only using urinalysis (UA) and complete blood count (CBC) parameters and C-reactive protein (CRP) value with four distinct machine learning systems including K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), Support Vector Machine (SVM), and Decision Tree Learning (DTL) without using antibiogram.

Materials and methods

The study was conducted based on the clinical records of 103 patients with *E. coli* infection aged 1–93 years (70 female and 33 male) who applied to Elazığ Fethi Sekin City Hospital Central Laboratory between 2019 and 2020. Clinical records including complete blood count (CBC) and complete urinalysis (UA) results, and CRP values were analyzed by machine learning systems including K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), Support Vector Machine (SVM), and Decision Tree Learning (DTL) without using antibiogram on the same day of antibiogram requests in order to investigate whether *E. coli* was susceptible or resistant to antibiotics. Clinical information from laboratory information system (LIS) data was extracted on each patient. Patients with resistant and/or recurrent urinary tract infections, who were requested for biochemistry laboratory tests at the same time as an antibiogram laboratory request were included in the study. Feature selection requires experimenting with many different possibilities and bringing together the intuition of the domain expert. In our study, we selected features that are known to be associated with urinary tract infection and that are made up of routine and easily accessible biochemistry laboratory parameters that are studied daily. Patients with incompatible clinical findings, those with suspected contamination, those who did not have any growth as a result of culture, patients with bacterial growth other than *E. coli*, and antibiogram evaluations that were not requested by the laboratory together with attribute selection (CRP, UA and CBC) were excluded from the study.

CBC parameters were measured using a Beckman Coulter DxH 800 hematology analyzer, urinalysis was performed using a Beckman Coulter IQ-2000 Elite analyzer, and CRP was measured using a

Table 1: Parameters used for the analysis and their descriptions, reference ranges and the laboratory characteristics of patients.

Parameter	Parameter description	Reference range	The laboratory characteristics of patients			
			Min	Max	Median	Interquartiles (25–75%)
CBC	WBC	3.6–11 (10^9 /L)	4.4	31.2	10	7.6–12.75
	Neu	1.7–7.6 (10^9 /L)	1.86	29.78	6.61	4.59–9.61
	Lym	1.0–3.2 (10^9 /L)	0.25	14.07	1.77	1.30–2.89
	Mon	0.3–1.1 (10^9 /L)	0.08	3.51	0.7	0.54–0.98
	Eos	0–0.5 (10^9 /L)	0.001	0.85	0.1	0.04–0.16
	Bas	0–0.1 (10^9 /L)	0.01	0.14	0.05	0.035–0.07
UA	Density	1.005–1.025	1.001	1.044	1.015	1.0105–1.019
	pH	4.05.2008	5	9	5.5	5–6.5
	Nitrite	Negative	Negative	Positive	None	None
	Erythrocyte	0–4	Negative	+++	None	None
	Leukocyte	0–4	Negative	+++	None	None
CRP	CRP	0–8 mg/L	1	398	18.6	9.20–73.1

CBC, Complete blood count; WBC, White blood cell; Neu, Neutrophil; Lym, Lymphocyte; Mon, Monocyte; Eos, Eosinophil; Bas, Basophil; UA, Urinalysis; CRP, C-Reactive Protein.

Beckman Coulter Image-800 Immunochemistry System. Bacterial culture was cultivated manually (From urine samples, 5% defibrinated sheep blood agar and eosin-methylene blue (EMB) agar plates were inoculated, single microorganism growth over 10^5 colonies (cfu/mL) was considered significant) and incubated for 24 h and then bacterial identification was achieved using conventional methods with MicroScan WalkAway 96 Plus ID/AST system (Beckman Coulter Inc., USA). *In vitro* antimicrobial sensitivity of isolates was assessed using MicroScan WalkAway 96 Plus ID/AST system with broth microdilution testing method, antibiotic susceptibility test according to the EUCAST (European Committee on Antimicrobial Susceptibility Testing) criteria [8]. Identification and antimicrobial susceptibility were testing was performed with B1017-165: Rapid Negative ID4, B1016-195: MIC EN52 combination panels. ESBL (Expanded Spectrum Beta-Lactamase) için bacterial suspension adjusted to McFarland 0.5 was cultivated in Mueller Hinton agar (RTA, Turkey) medium, then ceftazidime (Oxoid, UK) and ceftazidime/clavulanate (Oxoid, UK) discs were placed. After 24 h of incubation, a difference of ≥ 5 mm in the zone diameter was interpreted in favor of ESBL production. Additionally, antibiotic susceptibility of these isolates against 19 antibiotics were tested according to the EUCAST.

The antibiotics analyzed in the study included Amikacin (AK), Ampicillin (AMP), Amoxicillin-Clavulanate (AUG), Ceftazidime (CAZ), Cefixime (CFM), Cefotaxime (CTX), Colistin (CS), Ciprofloxacin (CIP), Cefepime (FEP), Cefuroxime (CXM), Ertapenem (ETP), Nitrofurantoin

(F), Phosphomycin (FOS), Gentamicin (CN), Imipenem (IMI), Levofloxacin (LEV), Meropenem (MRP), Piperacillin-Tazobactam (TZP), and Trimethoprim-Sulfadiazine (SUZ). Table 1 presents the parameters used for the analysis and their descriptions, reference ranges and the laboratory characteristics of patients.

Figure 1 illustrates the machine learning algorithm used for the assessment of antibiotic resistance of *E. coli*, which is the most common pathogen causing UTI. For each antibiotic drug, three parameters including (I) CBC (white blood cell count [WBC], neutrophil, lymphocyte, monocyte, eosinophil, basophil), (II) UA (density, pH, nitrite, erythrocyte, leukocyte), and (III) CRP were used as input parameters for the classifiers. As a result, a total of 12 parameters were obtained.

To facilitate data processing, raw data were normalized between -1 and $+1$. All the 12 parameters were used as an input vector for ANN, SVM, KNN, and DTL and the performances of these classifiers were compared for each parameter.

Data analysis

The models used in the study were tested in the Matlab R2018b (The MathWorks, Inc. Cambridge, United Kingdom) platform on a computer with an i7 9750 H CPU, 2.6 GHz, 16 GB RAM and Geforce GTX 1050 gCPU. Laboratory data analysis (laboratory characteristics of patients) was

Figure 1: Illustration of the algorithm used in the study.

performed on Jupyter Notebook using Python 3.0 (Python Software Foundation, Oregon, USA) program with Pandas library.

Classification

Artificial Neural Networks (ANN) is a machine learning technology that evolved from the idea of simulating the functioning of the human brain [9]. In ANN, back propagation is the most used algorithm for updating a neural network. The generalized delta rule is a mathematically derived formula used to determine how to update a neural network. In this technique, some portion of the difference (error) between the target and output values is back propagated to each training unit during a (back propagation) training step in order to update the weights according to the error and this procedure is iterated for a certain number of times to minimize this error [9].

The ANN model used in the present study consisted of 12 inputs (i.e., 12 parameters shown in Table 1) and 1 output (i.e., antibiotic susceptibility of the drugs). The error and learning rates of the training were set to 0.01 and 0.005, respectively.

Support vector machine (SVM) is another machine learning algorithm used for classification problems based on the structural risk minimization principle [10]. SVM tries to find the best hyperplane (also called decision boundary) to separate two classes. Equation (1) presents the decision boundary used for SVM:

$$f(x) : \text{sgn}\left(\sum_{i=1}^n y_i K(x, x_i) + b\right) \quad (1)$$

In this equation, α_i represents the Lagrange multipliers, x_i is the support vector, and b represents the bias term [11, 12]. Where linear separation is not possible, the following kernel functions can be used

$$\text{Radial Basis Function (RBF)} : K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2)$$

$$\text{Polynomial} : K(x_i, x_j) = (x_i x_j + 1)^d \quad (3)$$

In these equations, σ and d represent the kernel function parameters.

The k-near neighbor (KNN) classifier is a commonly used machine learning algorithm that measures the closeness of the new data to be classified to the k closest training examples in the feature space [13]. In the present study, decision tree learning (DTL), which is a tree-based learning algorithm, was also used to improve the classification performance [14].

On the other hand, k-fold cross validation was employed to minimize the distribution-related errors encountered during the training and testing of the proposed model [15]. In the study, the number of k was chosen as 10 in accordance with the data number. Figure 2 illustrates the implementation of k-fold cross validation.

Performance evaluation

The classification performance of the model was determined based on multiple criteria including Sensitivity (SN), Specificity (SP), Precision (PREC), Negative Predictive Value (NPV), False Positive Rate (FPR), False Discovery Rate (FDR), False Negative Rate (FNR), Accuracy (ACC), and F-Measure. Supplementary Material 1 presents the definitions of the parameters used for the Confusion Matrix and Supplementary Material 2 presents the formulas used for the calculation of the performance parameters [16].

Figure 2: Implementation k-fold cross validation.

Results

Seventy of our patients were female (68%), 33 were male (32%), and 81 of these patients (78%) were outpatient and 23 of them were inpatient. ESBL was detected in 31 of our patients (30%). Parameters used for the analysis and their descriptions, reference ranges and the laboratory characteristics of patients are given in Table 1.

The resistance of *E. coli* isolated from patients to antibiotics was found to be 69.9% to ampicillin, 51.4% to Cefixime and the least resistance to Imipenem and Levofloxacin with <1%. The resistance of *E. coli* isolated from patients to antibiotics Table 2, respectively; All four

Table 2: The resistance of *E. coli* isolated from patients to antibiotics.

Antibiotic	Sensitive	Resistant	%
AMP	31	72	69.9
CFM	50	53	51.4
CTX	54	49	47.5
CIP	55	48	46.6
FEB	58	45	43.6
SUZ	59	44	42.71
AUG	60	43	41.7
CAZ	61	42	40.7
CTX	67	36	34.9
CN	79	24	23.3
F	91	12	11.65
TZB	92	11	10.67
CS	95	8	7.76
ETP	97	6	5.82
AK	98	5	4.85
FOS	99	4	3.88
MRP	99	4	3.88
IMI	102	1	<1
LEV	102	1	<1

AK, Amikacin; AMP, Ampicillin; AUG, Amoxicillin-Clavulanate; CAZ, Ceftazidime; CFM, Cefixime; CTX, Cefotaxime; CS, Colistin; CIP, Ciprofloxacin; FEB, Cefepime; CXM, Cefuroxime; ETP, Ertapenem; F, Nitrofurantoin; FOS, Fosfomycin; CN, Gentamicin; IMI, Imipenem; LEV, Levofloxacin; MRP, Meropenem; TZP, Piperacillin/Tazobactam; SUZ, Trimethoprim-Sulfadiazine.

Table 3: Performance results of the antibiotics.

Antibiotic	Method	SN	SP	PREC	NPV	FPR	FDR	FNR	ACC	F ₁
AK	ANN	0.9636	0.9545	0.9138	0.9813	0.0455	0.0862	0.0364	0.9576	0.9381
AMP	SVM	0.3214	0.9688	0.8182	0.7654	0.0313	0.1818	0.6786	0.7717	0.4615
AUG	NED									
CAZ	SVM, KNN	0.750	0.9231	0.9231	0.7500	0.0769	0.0769	0.250	0.8276	0.8276
CFM	ANN	0.8333	0.9138	0.9091	0.8413	0.0862	0.0909	0.1667	0.8729	0.8696
CXM	SVM, KNN	0.7531	0.1818	0.8714	0.0909	0.8182	0.1286	0.2469	0.6848	0.8079
CS	SVM, KNN	0.9882	0.5714	0.9655	0.8	0.4286	0.0345	0.0118	0.9565	0.9767
CIP	SVM,	0.82	0.6905	0.7593	0.7632	0.3095	0.2407	0.18	0.7609	0.7885
FEP	ANN	0.7547	0.641	0.7407	0.6579	0.359	0.2593	0.2453	0.7065	0.7477
CXM	NED									
ETP	ANN	0.9885	0.6	0.9773	0.75	0.4	0.0227	0.0115	0.9674	0.9829
F	ANN	0.9756	0.3	0.9195	0.6	0.7	0.0805	0.0244	0.9022	0.9467
FOS	SVM, KNN	1	0.6667	0.9889	1	0.3333	0.0111	0	0.9891	0.9944
CN	SVM, KNN	0.9286	0.5909	0.8784	0.7222	0.4091	0.1216	0.0714	0.8478	0.9028
IMI	NED									
LEV	DTL, SVM KNN	1	0	0.989		1	0.011	0	0.989	0.9945
MRP	NED									
TZP	ANN	0.9759	0.4444	0.9419	0.6667	0.5556	0.0581	0.0241	0.9239	0.9586
SUZ	ANN	0.8269	0.75	0.8113	0.7692	0.25	0.1887	0.1731	0.7935	0.819

AK, Amikacin; AMP, Ampicillin; AUG, Amoxicillin-Clavulanate; CAZ, Ceftazidime; CFM, Cefixime; CTX, Cefotaxime; CS, Colistin; CIP, Ciprofloxacin; FEP, Cefepime; CXM, Cefuroxime; ETP, Ertapenem; F, Nitrofurantoin; FOS, Fosfomycin; CN, Gentamicin; IMI, Imipenem; LEV, Levofloxacin; MRP, Meropenem; TZP, Piperacillin/Tazobactam; SUZ, Trimethoprim-Sulfadiazine; ANN, Artificial Neural Networks; SVM, Support Vector Machine; KNN, K-Nearest Neighbors; DTL, Decision Tree Learning; SN, Sensitivity; SP, Specificity; PREC, Precision; NPV, Negative Predictive Value; FPR, False Positive Rate; FDR, False Discovery Rate; FNR, False Negative Rate; ACC, Accuracy; NED, Not Enough Data.

classifiers (ANN, SVM, KNN, and DTL) were used to determine whether *E. coli* isolates were resistant or susceptible to the 19 antibiotics administered in the patients. Classification results were then compared with antibiogram results. The results indicated that the performance of the classifiers varied across the antibiotics administered in the patients. Table 3 presents the performance results for each classifier based on the parameters used in the analysis.

The performance of the proposed method in diagnosing antibiotic resistance was assessed by the individual use of CBC, UA parameters, and CRP and also by the use of all parameters (Figure 3).

Discussion

Inappropriate broad-spectrum antibiotics used in the treatment of community-acquired infections will cause the resistance of organisms to antibiotics to increase rapidly and the difficulties that will arise in this regard in the future will increase. For this reason, it is clear that much faster and inexpensive methods are needed to determine the resistance to antibiotics.

In the studies conducted in our region, Duman et al. found the highest antibiotic resistance against *E. coli* to Ampicillin and the lowest resistance to Amikacin and

Figure 3: Classification results based on the use of CBC, UA parameters, CRP, and all of these parameters.

Imipenem, while Denk et al. found the highest resistance to ampicillin and the lowest to Phosphomycin and Nitrofurantoin [17, 18]. In our study, the highest resistance was found against Ampicillin and the lowest resistance was against Imipenem and Levofloxacin, hence our dataset is compatible with the *E. coli* antibiotic resistances in our region (Table 2).

The results indicated that the analysis of routine biochemical laboratory parameters by machine learning systems can predict antibiotic resistance of *E. coli* infection, i.e., the most common cause of UTI. Similarly, numerous previous studies also used machine learning systems to investigate antibiotic resistance, as shown in Table 4.

As seen in Table 4, most of the studies investigated the prediction of antibiotic resistance based on the specific genotype of the pathogen [19–24]. In contrast, unlike other studies, Yelin et al. performed this prediction based on the 10-year clinical history of the patients [25]. The method proposed in the present study, however, is highly different from those reported in other studies in that it is based on the biochemical laboratory parameters that are routinely measured in almost all patients with suspected UTI. Accordingly, this method appears to be a reasonable option as it is highly cost-effective and employs a relatively lower number of datasets for the analysis.

In the present study, four classifiers (ANN, SVM, KNN, and DLT) were used to determine *E. coli* resistance to antibiotics. In machine learning systems such as ANN, the parameters used for the analysis are problem-bound; therefore, it is often not possible to determine which parameter (e.g., number of multilayer perceptron [MLPs], number of neurons in hidden layers, learning coefficient) will provide an optimal outcome. As a solution, the trial-and-error method is used to determine which classifier will provide the highest performance. For these reasons, there

is no valid comparison of these classifiers. Nonetheless, it could be asserted that an algorithm may have tend to a specific problem [26].

Accuracy (ACC) is calculated as the number of correctly classified instances divided by the total number of the dataset. Meaningfully, ACC alone is not sufficient for the evaluation of unbalanced datasets. In the literature, limited data is available regarding the antibiotic resistance of the input parameters used in our study [27, 28]. Therefore, a balanced distribution could not be achieved between our resistant and non-resistant parameters in terms of the number of cases. AUG, CXM, IMI and MRP are shown as NED in Table 3. Sensitivity (SN) refers to the ability of a test to estimate the number of correct positive predictions divided by the total number of positives. Specificity (SP) refers to the ability of a test to estimate the number of correct negative predictions divided by the total number of negatives. Meaningfully, SN and SP should be used together for the estimation of these predictions. A test with high SP helps in avoiding misunderstandings and unnecessary preventable interventions (True negative) while a test with high SN is needed particularly in cases of ambiguous diagnosis and in early disease conditions (True positive). The F_1 score was also used in the present study due to the fact that it employs the harmonic mean instead of the arithmetic mean to avoid overlooking extreme conditions.

In some of the antibiotics analyzed in the study, there was a remarkable difference between the SN and SP values of the classifiers, which indicated poor generalization and classification performance of classifiers. This difference was mostly seen in imbalanced datasets (Table 3).

For each classifier, 10-fold cross validation was performed to improve their validity and generalization performance (Figure 2). Low data exchange rates obtained

Table 4: Studies investigating antibiotic resistance by machine learning systems.

Study	Target bacteria	Data type	Method
Kavvas et al. [19]	<i>Mycobacterium tuberculosis</i>	Pan-genome	SVM
Moradigaravand et al. [20]	<i>E. coli</i>	Pangenome, population structure matrix	Gradient-boosted trees, random forest, deep neural networks
Yang et al. [21]	<i>M. tuberculosis</i>	DNA sequencing data	Random forest
Chen et al. [22]	<i>M. tuberculosis</i>	Whole-genome sequencing	Deep neural networks
Li et al. [23]	<i>S. pneumoniae</i>	Penicillin binding protein (PBP) sequences	Random forest
Nguyen et al. [24]	<i>K. pneumoniae</i>	All genomes	XGBoost
Yelin et al. [25]	<i>E. coli</i> , <i>K. pneumoniae</i> and <i>P. mirabilis</i>	Personal clinical history	Multivariate logistic regression, gradient boosting decision trees
Present study	<i>E. coli</i>	CBC, UA, CRP	ANN, SVM, KNN, DTL

SVM, Support Vector Machine; CBC, Complete Blood Count; UA, Urinalysis; CRP, C-Reactive Protein; ANN, Artificial Neural Networks; KNN, K-Nearest Neighbors; DTL, Decision Tree Learning.

during cross validation further support the linearity assumption of the dataset.

Given that the performance of a classifier depends on the problem to be analyzed, it is tempting to consider that using appropriate input parameters is equally important as using an appropriate classifier. In our study, although the use of all parameters (CBC, UA parameters and CRP) provided an acceptable classification performance, the individual effect of these parameters on this performance was also assessed and the results indicated that the use of all parameters showed the highest classification performance in most of the antibiotics analyzed in the study (Figure 3, Table 2).

Our results also indicated that the performance of the classifiers used for the diagnosis of *E. coli* infection varied across the antibiotics administered in the patients. For this reason, it cannot be mentioned that the same classifier is prone to problem solving for each antibiotic in the study. On the other hand, the study also found a relationship between the parameters analyzed in the study (CBC and UA parameters and CRP) and the resistance of different antibiotics used for the treatment of *E. coli* infection and also indicated that these parameters could be used in the diagnosis of this infection.

Limitations

The limitation of the study is that we cannot control how the performance parameters of our model will change in different datasets. Therefore, it is necessary to use different datasets to ensure the general accuracy of the study. It is aimed to use different datasets in future studies.

Conclusion

Antibiotic resistance is a potentially serious public health problem worldwide [29]. The present study provides a relatively more cost-effective and more practical method to algorithmic treatment modalities that require no use of antibiogram. Moreover, the study also showed that *E. coli* infection can be diagnosed by the analysis of CBC, UA parameters and CRP with machine learning systems and without the use of antibiogram, something that has never been documented in the literature. The present study also contributed to the variety of diagnostic and treatment modalities by combining biochemical and microbiological laboratory parameters, thereby providing a substantial solution for clinical problems.

Conflict of interest: All authors read and approved the final manuscript. None of the authors had a conflict of interest.

Çıkar çatışması: Bildirilmemiştir.

References

1. <https://www.cdc.gov/drugresistance/about.html> [Accessed 1 Aug 2020].
2. Reygaert WC. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol* 2018;4:482.
3. Yoshida M, Reyes SG, Tsuda S, Horinouchi T, Furusawa C, Cronin L. Time-programmable drug dosing allows the manipulation, suppression and reversal of antibiotic drug resistance in vitro. *Nat Commun* 2017;8:1–11.
4. Chancey ST, Zähler D, Stephens DS. Acquired inducible antimicrobial resistance in Gram-positive bacteria. *Future Microbiol* 2012;7:959–78.
5. Mahon CR, Lehman DC, Manuselis G. Antimicrobial agent mechanisms of action and resistance. In: *Textbook of diagnostic microbiology*. St. Louis: Saunders; 2014:254–73.
6. Cesur S, Demiröz AP. Antibiotics and the mechanisms of resistance to antibiotics. *Med J Islamic World Acad Sci* 2013;109:1–5.
7. Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat Rev Microbiol* 2015;13:269–84.
8. The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Version 5.0; 2015. <http://www.eucast.org>.
9. He C, Ma M, Wang P. Extract interpretability-accuracy balanced rules from artificial neural networks: a review. *Neurocomputing* 2020;387:346–58.
10. Khazaei A, Ebrahimzadeh A. Classification of electrocardiogram signals with support vector machines and genetic algorithms using power spectral features. *Biomed Signal Process Contr* 2010;5:252–63.
11. Arslan TS, Alkan A. A decision support system for detection of the renal cell cancer in the kidney. *Measurement* 2018;123:298–303.
12. Dursun ÖO, Toraman S, Türkoğlu İ. A comparison of the classification performances of criminal tendencies of schizophrenic patients by artificial neural networks and support vector machine. *EJT* 2017;7:177–85.
13. Ayyıldız H, Arslan TS. Determination of the effect of red blood cell parameters in the discrimination of iron deficiency anemia and beta thalassemia via neighborhood component analysis feature selection-based machine learning. *Chemometr Intell Lab Syst* 2020;196:103886.
14. Açıkoğlu M, Tuncer SA. Incorporating feature selection methods into a machine learning-based neonatal seizure diagnosis. *Med Hypotheses* 2020;135:109464.
15. Xu M, Papageorgiou DP, Abidi SZ, Dao M, Zhao H, Karniadakis GE. A deep convolutional neural network for classification of red blood cells in sickle cell anemia. *PLoS Comput Biol* 2017;13: e1005746.
16. Luque A, Gomez-Bellido J, Carrasco A, Barbancho J. Optimal representation of anuran call spectrum in environmental monitoring systems using wireless sensor networks. *Sensors* 2018;18:1803.

17. Duman Y, Güçlüer N, Serindağ A, Tekerekoğlu M. *Escherichia coli* Suşlarında Antimikrobiyal Duyarlılık Ve Genişlemiş Spektrumlu-Beta Laktamaz (Gsbl) Varlığı. *Fırat Tıp Dergisi* 2010; 15:197–200.
18. Denk A, Tartar A. İdrar Kültürlerinden İzole Edilen Toplum Kökenli *Escherichia coli* Suşlarında Antibiyotik Direnci. *Fırat Univ Sağlık Bilim* 2015;29:51–5.
19. Kavvas ES, Catoiu E, Mih N, Yurkovich JT, Seif Y, Dillon N, et al. Machine learning and structural analysis of *Mycobacterium tuberculosis* pan-genome identifies genetic signatures of antibiotic resistance. *Nat Commun* 2018;9:4306.
20. Moradigaravand D, Palm M, Farewell A, Mustonen V, Warringer J, Parts L. Precise prediction of antibiotic resistance in *Escherichia coli* from full genome sequences. *PLoS Comput Biol* 2018;14: e1006258.
21. Yang Y, Niehaus KE, Walker TM, Iqbal Z, Walker AS, Wilson DJ, et al. Machine learning for classifying tuberculosis drug-resistance from DNA sequencing data. *Bioinformatics* 2017;34: 1666–71.
22. Chen ML, Doddi A, Royer J, Freschi L, Schito M, Ezewudo M, et al. Deep learning predicts tuberculosis drug resistance status from whole-genome sequencing data. *BioRxiv* 2018:275628.
23. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Walker H, et al. Validation of β -lactam minimum inhibitory concentration predictions for pneumococcal isolates with newly encountered penicillin binding protein (PBP) sequences. *BMC Genom* 2017;18: 1–10.
24. Nguyen M, Brettin T, Long SW, Musser JM, Olsen RJ, Olson R, et al. Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. *Sci Rep* 2018;8:421.
25. Yelin I, Snitser O, Novich G, Katz R, Tal O, Parizade M, et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. *Nat Med* 2019;25:1143–52.
26. Altun H, Polat G. On the comparison of classifiers' performance in emotion classification: critiques and suggestions. In: 2008 IEEE 16th signal processing, communication and applications conference; 2008. <https://doi.org/10.1109/siu.2008.4632592>.
27. Sabir S, Anjum AA, Ijaz T, Ali MA, Khan MUR, Nawaz M. Isolation and antibiotic susceptibility of *E. coli* from urinary tract infections in a tertiary care hospital. *Pak J Med Sci* 2014;30:389–92.
28. Kaçmaz B, Aksoy A, Sultan N. The investigation of resistance to oral antibiotics in *Escherichia coli* isolates obtained from urine. *Türk Hij Den Biyol Derg* 2007;64:11–5.
29. <https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance>.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/tjb-2021-0040>).