



Diagnosis of urinary tract infection based on artificial intelligence methods

Ilker Ali Ozkan^{a,*}, Murat Koklu^a, Ibrahim Unal Sert^b

^a Department of Computer Engineering, Faculty of Technology, Selcuk University, Konya, Turkey

^b Department of Urology, Necmettin Erbakan University, Konya, Turkey

ARTICLE INFO

Article history:

Received 22 March 2018

Revised 5 September 2018

Accepted 1 October 2018

Keywords:

Artificial intelligence methods
Medical decision support systems
Urinary tract infection
Artificial neural networks
Decision tree
Support vector machine
Random forest

ABSTRACT

Background and Objective: Urinary tract infection (UTI) is a common disease affecting the vast majority of people. UTI involves a simple infection caused by urinary tract inflammation as well as a complicated infection that may be caused by an inflammation of other urinary tract organs. Since all of these infections have similar symptoms, it is difficult to identify the cause of primary infection. Therefore, it is not easy to diagnose a UTI with routine examination procedures. Invasive methods that require surgery could be necessary. This study aims to develop an artificial intelligence model to support the diagnosis of UTI with complex symptoms.

Methods: Firstly, routine examination data and definitive diagnosis results for 59 UTI patients gathered and composed as a UTI dataset. Three classification models namely; decision tree (DT), support vector machine (SVM), random forest (RF) and artificial neural network (ANN), which are widely used in medical diagnosis systems, were created to model the definitive diagnosis results using the composed UTI dataset. Accuracy, specificity and sensitivity statistical measurements were used to determine the performance of created models.

Results: DT, SVM, RF and ANN models have 93.22%, 96.61%, 96.61%, 98.30% accuracy, 95.55%, 97.77%, 95.55%, 97.77% sensitivity and 85.71%, 92.85%, 100%, 100% specificity results, respectively.

Conclusions: ANN has the highest accuracy result of 98.3% for UTI diagnosis within the proposed models. Although several symptoms, laboratory findings, and ultrasound results are needed for clinical UTI diagnosis, this ANN model only needs pollacuria, suprapubic pain symptoms and erythrocyte to get the same diagnosis with such accuracy. This proposed model is a successful medical decision support system for UTI with complex symptoms. Usage of this artificial intelligence method has its advantages of lower diagnosis cost, lower diagnosis time and there is no need for invasive methods.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Urinary tract infection (UTI) is a common disease (approximately 95% of the total urinary tract infections) generally caused by *E. coli* bacteria reaching the bladder from the urinary tract and affecting at least half of the people throughout their lives [1,2]. It may be observed as a simple inflammation of the bladder or a more severe inflammation of other urinary system organs. Inflammations involving only bladders that are not based on another underlying disease are called “simple infections” (cystitis) and all other urinary tract infections are called “complicated infections”. Cystitis is a bladder inflammation. This term is used as a patho-

logical, bacteriological and cystoscopy term. Suddenly starting dysuria, pollacuria, urinary urgency and suprapubic pain are the syndromes of cystitis [3]. Cystitis is diagnosed with pollacuria, suprapubic pain, and burning sensation during urination, if the infection is detected only in the bladder. If the infection has retained the ureter, it is diagnosed as a nonspecific urethritis with diuresis, pollacuria, leukocyte urine findings. This diagnosis is also complicated since the symptoms are intertwined with one of the symptoms. The infection results in septicemia, death or at least the chronicity of the disease, if the organ in which the urinary tract infection is located is not treated properly. Under the top name of UTI, exist all kinds of inflammation on kidney, bladder, and urethra [1,2].

150 million patients are reported to be diagnosed with UTI per year and \$16 billion was spent on these patients [4]. In addition, the lifetime incidence of UTI counts 14,000 from 100,000 men [5]. In women, the three entities, the anus, the vagina mouth and the

* Corresponding author.

E-mail addresses: ilkerozkan@selcuk.edu.tr (I.A. Ozkan), mkoklu@selcuk.edu.tr (M. Koklu).

urethra, are very close to one another. In women, the three entities, the anus, the vagina mouth and the urethra, are very close to one another. Anal and vaginal area bacteria can easily become cystitis caused by passing from the urethra to the bladder. This rate is even higher due to the women urethral anatomy [6].

Routine urinalysis and urine cultures are utilized for the diagnosis of UTI. However, it is not easy to identify the source of the infection and the organ is infected. Depending on the characteristics of patients, it may be required to apply invasive methods such as urineletting separately from the bladder and kidney by placing a catheter in the ureters with the help of the cystoscopy. These invasive methods may cause more complications for the patient. For this reason, developing an AI model based on symptom and laboratory findings will protect the patient from possible complications.

UTI causes 20–30% of all systemic infections. In order to start treatment of the patient, a definite diagnosis is required whether cystitis or nonspecific urethritis is present. Since cystitis and nonspecific urethritis need to be treated with different drugs.

During the diagnosis, cystitis is suspected primarily because the main organ of the lower urinary tract is the bladder. The main symptoms of the acute cystitis are diarrhea, pollacuria, hematuria, leucocyturia, urinary urgency. Sometimes fever rarely participates in this history of acute urinary tract infections in children [7]. It is also accompanied by suprapubic pain. However, fever may also be observed in infections involving the bladder muscle (detrusor). Diabetes and recurrent UTI are also a risk factor for cystitis.

Patients with urethritis are applying to polyclinic with the symptoms of defluxion in urethra and dysuria. Among the urethritis agents, neisseria gonorrhoeae, chlamydia trachomatis, ureaplasma urealyticum and trichomonas vaginalis are the most common factors. However, except for gonorrhea infections, it is difficult to detect other factors in the laboratory environment [8]. In the clinic, it is both impractical and difficult to get results from the urethral drainage culture to diagnose the infections separately except for gonorrhea infections.

Artificial intelligence methods are successfully applied to situations like these where complexity exists. With the use of artificial intelligence in medicine, artificial intelligence programs can be created that can perform clinical diagnostic procedures and recommend treatment suggestions. There are a lot of successful applications of artificial intelligence in the field of medicine, such as classification and diagnosis of diseases, treatment recommendation, drug dosage determination, and so on [9–18]. These applications use artificial intelligence methods such as SVM, ANN, DT, KNN and RF. These artificial intelligence methods are powerful tools that physicians can use to analyze, model, and understand complex clinical data in a variety of medical practice areas.

Furthermore, the use of AI methods in medical practice has become widely accepted. These methods include advantages such as (a) economically and non-linear modeling of large data sets and ease of optimization; (b) the predictive accuracy that has a potential to support clinical decision process; (c) providing an explanation using rule inference or sensitivity analysis which ease to disseminate the information [19].

With artificial intelligence methods, inputs that are independent variables of the system can form complex, nonlinear models by associating them with dependent predicted variable outputs. Artificial intelligence is the most important tool for modeling and decision making. Thus, in this study, various AI classification methods have been compared for diagnose supporting whether UTI with complicated symptoms is cystitis or nonspecific urethritis. Four different AI methods, DT, SVM, RF and ANN, which are frequently used in medical classification problems, have been used in this study [20–23]. A comparative study of those will give us insight into the merits of the different AI methods when used on this problem.

2. Material and methods

2.1. Data set

In this study, the data set were used of the 59 patients 7–77 years old, 35 female and 24 male, who applied to the Necmettin Erbakan University Meram Medical Faculty Urology Polyclinic between the dates of 2016–2017. These patients were diagnosed and localized as urinary tract infection (cystitis) and nonspecific urethritis. All of these patients were diagnosed with the procedures such as anamnesis, clinical examination, urinalysis, and ultrasonography. A definitive diagnosis chart of the data set based on the gender of the patients is given in Fig. 1.

The dataset of this study has 17 different qualitative inputs obtained from anamnesis, clinical examination and ultrasound examination of the patients who applied to the urology clinic for definitive diagnosis and diagnosis of UTI. These data are given in Table 1. The renal and bladder ultrasound data given in Table 1 were used only for definitive diagnose purposes and were not used as an input in the machine learning applications.

With the decision number 2018/1180, Necmettin Erbakan University Meram Medical Faculty Ethics Board has permitted to use the data in this study for the engineering approaches.

2.2. Decision tree

Decision trees are a kind of tree-like decision structure that is learned by the inductive method of the known examples of each class. A decision tree is a structure that is used by dividing large amounts of records into very small groups by applying simple decision-making steps. Thus, with each successful dividing process, the members of the related result groups become much more similar to each other [24].

As seen in Fig. 6, decision tree nodes consist of branches and leaves. The decision node indicates the test to be performed. As a result of this test, the tree branches without losing any data. At each node, test and branching operations occur consecutively, and this branching operation is dependent on the upper-level branches. Each branch in the tree is prone to complete the classification process. If the classification operation can't be performed at the end of a branch, a decision node is formed at the end of that branch. However, if a particular class is formed at the end of a branch, there is leaf at the end of that branch. This leaf is one of the classes to be determined on the data. The decision tree process begins with the root node and follows the consecutive nodes until it reaches the leaf from top to bottom. The classification of the data using the decision tree technique is a two-step process. The first one is the learning step. In the learning step, a training data which is known beforehand is analyzed by a classification algorithm in order to create a model. The learned model is presented as classification rules or a decision tree. The second one is the classification step. The test data on the classification step is used to determine the correctness of the classification rules or decision tree. If accuracy is within an acceptable range, rules are reused to classify new data [25–28].

The decision tree is similar to the decision making the process of the doctor. Also, after the decision tree has been modeled, it can provide a good explanation about the problem. For this reason, the decision tree has been preferred in this study.

2.3. Support vector machines

Support Vector Machines (SVM) is a controlled classification algorithm based on statistical learning theory. The mathematical algorithms used by the SVM were originally designed for the problem of classification of two-class linear data, then generalized to

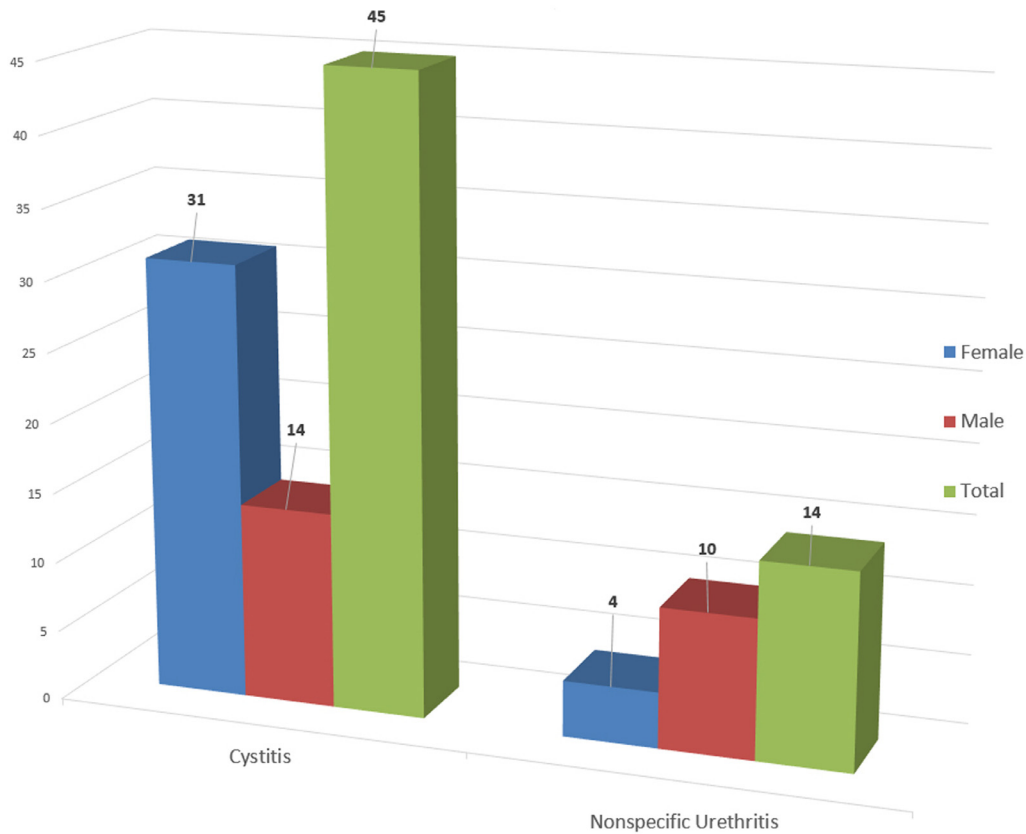


Fig. 1. Number range of UTI-patients according to gender.

Table 1
List of variables with their descriptions and their value ranges.

	Variable description	Values (min –max)	Normalized	
Anamnesis	Gender	(0) Male (1) Female	(0,1)	
	Age	7–77	(0,1)	
	Fever	(0) No (1) Has	(0,1)	
	Sudation	(0) No (1) Has	(0,1)	
	Chill	(0) No (1) Has	(0,1)	
	Low back pain	(0) No (1) Has	(0,1)	
	Suprapubic Pain	(0) No (1) Has	(0,1)	
	Malaise	(0) No (1) Has	(0,1)	
	Dysuria	(0) No (1) Has	(0,1)	
	Pollacuria (more than 6 times per day)	(0) No (1) Has	(0,1)	
	Full urine analysis	Urine culture	(0) Positive (1) Negative	(0,1)
		Glucose	(0) Positive (1) Negative	(0,1)
		Leucocyte	(0) Positive (1) Negative	(0,1)
Erythrocyte		(0) Positive (1) Negative	(0,1)	
Protein		(0) Positive (1) Negative	(0,1)	
Ultrasound	Renal ultrasound			
	Bladder ultrasound			
Diagnose	Cystitis	(0) Positive (1) Negative	(0,1)	
	Urethritis	(0) Positive (1) Negative	(0,1)	

classify multi-class and non-linear data. The working principle of the SVM is based on the principle of predicting the most appropriate decision function that separates the two classes, in other words, on the definition of the hyperplane, which can distinguish the two classes from each other in a most appropriate way [29,30].

The basic idea in the SVM regression method is to find out a linear discriminant function that reflects the true nature of the training data in the true sense as closely as possible and fits the statistical learning theory. Similar to the classification, kernel functions are used to process nonlinear states in the regression [31]. The two situations that can be encountered in Support Vector Machines are that the data can be linearly separated or not separated.

In cases where the data cannot be separated linearly, non-linear classifiers can be used instead of linear classifiers.

Non-Linear Feature Space in this context: Linear classifiers may be obtained in this new space by transforming the $x \in R^n$ observation vector into a vector z in a higher order space. In this case, the expression ϕ is showed in the form of $z = \phi(x)$ to make the $R^n \rightarrow R^F$ mapping [29,32].

$$x \in R^n \rightarrow z(x) = [a_1, \phi_1(x), \dots, a_n, \phi_n(x)]^T \in R^F$$

When non-linear separability is considered, the training samples cannot be linearly separated in the original input space. In such cases, SVM transforms to the high dimensional feature space,

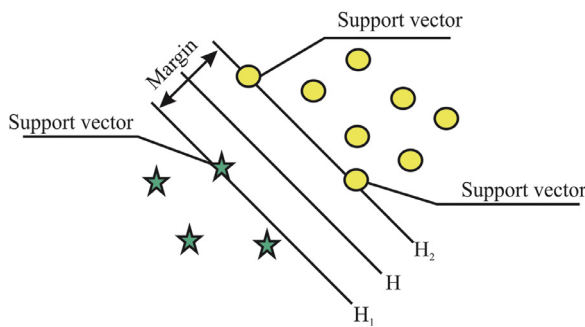


Fig. 2. Mapping of two class problems by SVM.

which can be easily classified linearly from the original input space by means of the nonlinear mapping function. Thus, instead of finding values by repeatedly multiplying them using kernel functions, it is substituted for that value directly in the kernel function and its counterpart found in the feature space. By this way, there is no need to deal with a very high-dimensional quality space. Another advantage of kernel functions is to make calculation of the kernel functions much easier so that after a function is established and values are found for a training example in the training phase, for other examples, the pattern values except the training example are completely ready to use [33–35]. SVM has four kernel functions which are widely used the linear function, polynomial function sigmoid function, radial basis function. A typical two-class problem given in Fig. 2 is similar to the problem of cystitis or non-specific prostatic urethritis in UTI patients.

2.4. Random forest (RF)

In ensemble classification methods, results produced by multiple classifiers are used instead of a classifier. One of the most widely used examples for ensemble classification is the RF method proposed by Breiman [36]. RF is a type of classifier that uses randomly generated samples from existing situations and consists of multiple trees. To classify a sample, each tree in the forest is given an input vector and a result is produced for each tree. RF algorithm chooses the class with the most votes as the result [37].

RF divides each node into branches using the best of the randomly selected variables on each node between all variables. In the RF method, trees are created with selected bootstrap samples and each node separation is created by randomly selected n estimators. It should be noted that the number of n units is quite small compared to the total number of estimators. Each decision tree is in its widest form and not truncated [36,38]. For classification trees; each leaf node is created to contain only members of a class. In Fig. 3, the general working principle of the random forest is given.

Since improved random sampling and other improved techniques in ensemble methods include in the RF method, it provides better generalizations and makes valid estimates [39]. Higher estimate certainty of the RF method are because of the less bias results and the low correlation between trees. The low amount of bias is obtained as a result of the creation of very large trees. Furthermore, RF does not overfitting [40].

2.5. Artificial neural network (ANN)

The human brain is a complex system with thinking, remembering and problem-solving abilities. The basic unit of this complex structure is neuron. Generally; obtaining the system model or identifying the system; based on the measurement of input, output and state variables means learning the structure of the physical system. One or more of the state variables can often be system

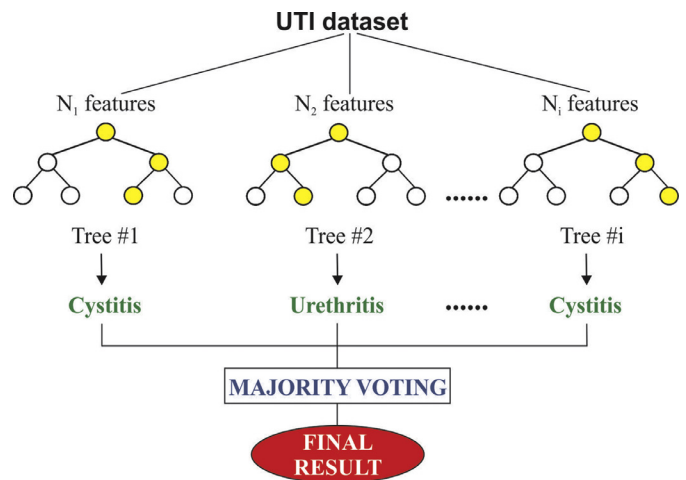


Fig. 3. Working principle of RF.

output. A system model can be specified by using the mathematical expression, input-output data set, or linguistic rules [41,42].

Artificial neural networks are an information processing system that is inspired by biological neural networks and includes some performance characteristics similar to biological neural networks. In a simple way, ANNs has many important features to imitate the way the human brain works such as the ability to learn from the data, generalize, work with an unlimited number of variables, etc.. The smallest unit that sets the basis for the work of ANN is called artificial neuron or process element. As shown in Fig. 4, the simplest artificial neuron consists of five main components: inputs, weights, transfer function, activation function, and output [43–46].

Inputs (x_1, x_2, \dots, x_n) are information that enters into the cell from other cells or from outside environments. These are determined by examples which desired to be learned by the network. Weights (w_1, w_2, \dots, w_n), are the values that represent the effect of the input set or another processed element on a previous layer on this processed element. Each input is multiplied by the weight value that connects it to the input processed element and combined through the sum function. The sum function is as given in Eq. (1) [43,44,47].

$$net = \sum_{i=1}^n w_i x_i + b \quad (1)$$

The value obtained as a result of the sum function is passed through a linear or non-linear differentiable transfer function to calculate the output of the processed element (Eq. (2)).

$$y = f(net) = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (2)$$

The artificial neural network consists of a series of neurons, shown in Fig. 4, interconnected by feedforward and backpropagation forms. At present, many artificial neural network models (Perceptron, Adaline, ANN, LVQ, Hopfield, Recurrent, SOM, ART etc.) suitable for use with specific purposes and in various fields have been developed. These are artificial neural networks that find the most common use in our network structures and are also used in our study [48].

In ANN, neurons are organized in layers. The first layer is the input layer. The input layer ensures that the information about the problem to be solved is input to ANN. The other one is called an output layer where information processed in the network is transmitted to the outside. The layer between the input and output layers is called the hidden layer. Multilayer Perceptron (MLP)

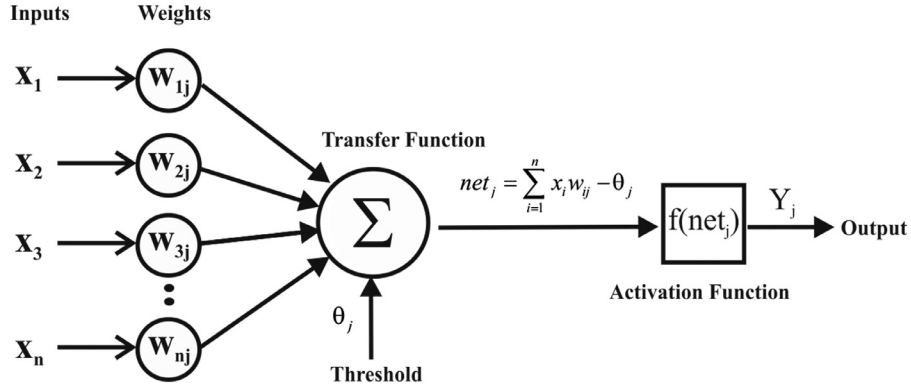


Fig. 4. Artificial Neuron.

networks may also have multiple hidden layers. Fig. 7 shows the structure of a typical ANN [48].

The network is regulated by minimizing the error function given in Eq. (3). In the equation, y_k denotes the output calculated by the network and t_k denotes the ground truth. The connection weights are re-calculated and updated to minimize the error. Thus, it is aimed to bring output values that are closest to the ground truth values of the network.

$$E = \frac{1}{2} \left(\sum_{k=1}^m (y_k - t_k) \right)^2 \quad (3)$$

2.6. Commonly-accepted performance evaluation measures

This is the case we focus on in this study. Classification performance without focusing on a class is the most general way of comparing algorithms. It does not favor any particular application. The introduction of a new learning problem inevitably concentrates on its domain but omits a detailed analysis. Thus, the most used empirical measure, accuracy, does not distinguish between the number of correct labels of different classes [49]:

- TP = true positives: number of examples predicted positive that are actually positive
- FP = false positives: number of examples predicted positive that are actually negative
- TN = true negatives: number of examples predicted negative that are actually negative
- FN = false negatives: number of examples predicted negative that are actually positive

Accuracy: It refers to the total number of records that are correctly classified by the classifier. The accuracy of a classifier is defined as the percentage of test set tuples that are correctly classified by the model [50].

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (4)$$

Sensitivity: Refers to the true positive rate that means the proportion of positive tuples that were correctly identified [50].

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

Specificity: Refers the rate at which a test or diagnostic method sets a correct (i.e., negative) diagnosis for a patient who is not ill.

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \quad (6)$$

Positive Predictive Value: The fraction of retrieved instances that are relevant [50].

$$\text{Positive Predictive Value} = \frac{TP}{TP + FP} \times 100\% \quad (7)$$

Negative Predictive Value: The result that the modeling makes a negative prediction [50].

$$\text{Negative Predictive Value} = \frac{TN}{FN + TN} \times 100\% \quad (8)$$

3. Experimental study

3.1. Preprocessing

The “normalization” of the data is important in terms of not corrupting the relationship between the variables, the accuracy of the analysis and the network performance. Normalization is a scaling process for each data in the dataset between the upper and lower bounds of the activation function used. The activation function used in the analysis is “sigmoid” function, during the implementation the data were normalized to the range of [0,1] by using the Eq. (9).

$$x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

Furthermore, Ultrasound values from the input parameters given in Table 1 were used only to determine the accuracy of the diagnosis. In the classification algorithms, values except ultrasonic input parameters were taken.

Experimental studies have shown that the optimal value for k is between 5 and 10 in the k -fold cross-validation method [51,52]. In this study, the data set was divided into 5 parts as shown in Fig. 5 by using 5-layer cross-validation method. The system has been trained and tested with k different training and test sets and k success measures were obtained accordingly. For this reason, to determine the success of cross-validation, the arithmetic mean of the obtained k performance measures were taken.

Our data set consists of 59 samples. For the purpose of 5-fold cross-validation, the data set was divided into parts; the first one with 11 elements and other four with 12 elements for each fold. The elements of each part were randomly selected, not sequentially. After the parts were formed in this way, one part of each fold was separated for testing and the remaining parts were used for the training of the algorithm of the method. It was then tested on the separated part of the data. Since there are 5 parts, 5 different test results were formed. The average success of these test processes was calculated as the overall success, that is, the success of the method to be applied [53].

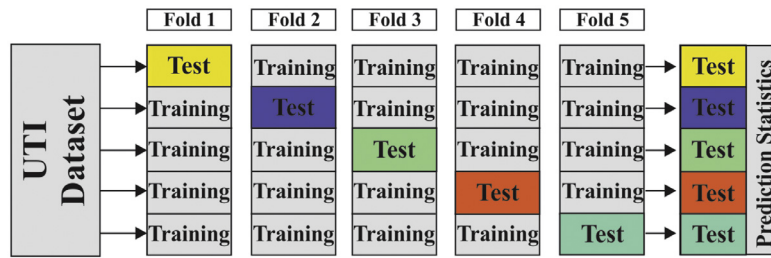


Fig. 5. 5-Fold cross-validation method.

Table 2
Decision tree model parameters.

Split criterion	Gini's diversity index
Number of folds	5
Score transformation	Identity (x)
Maximal number of decision splits	58

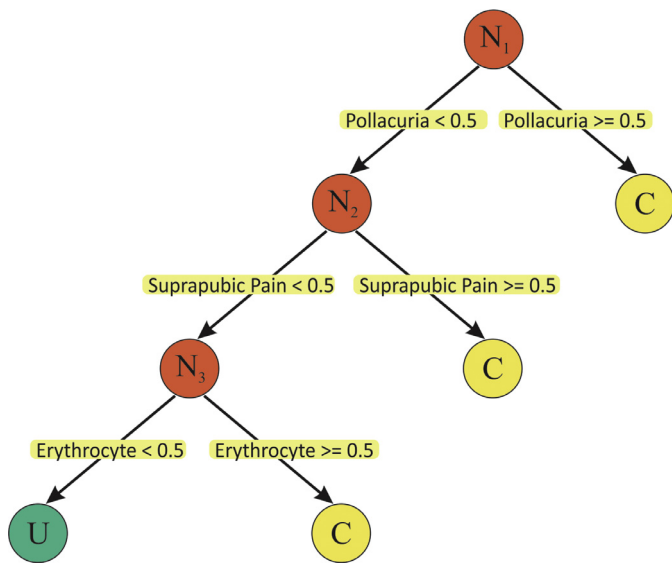


Fig. 6. Decision tree model method (N: Node, C: Cystitis, U: Urethritis).

3.2. Development of the AI methods

3.2.1. Decision tree

The decision tree algorithm was used first because this was similar to the general diagnostic procedure of the physicians. The parameters of the Decision Tree model are given in Table 2. The figure of the decision tree structure is seen. When the created decision tree model is examined, it is observed that the results can be obtained by using pollacuria, suprapubic pain and erythrocyte input features only. These input features given in Fig. 6 give us an idea of which input variables are important.

3.2.2. SVM

Support vector machines are a highly effective machine learning technique that is often used to classify data in the biomedical field. It is used for classification by specifying boundaries between two or more groups in a plane with the aim of separating these groups from each other. The place where the limits will be put should be the farthest place to the group members. With SVM, boundaries can be defined for both linear and nonlinear data sets. The Radial Basis Function (RBF) kernel SVM was used in this study because the SVM is successful in applications with high-dimensional but small amounts of data and also gives better results even though it

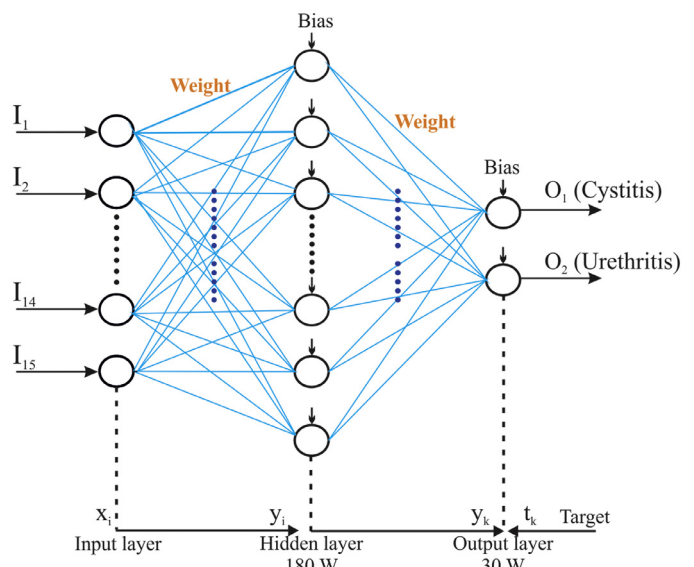


Fig. 7. Feedforward ANN.

is more complicated when compared to linear SVM. The RBF kernel maps the samples nonlinearly to a higher dimensional space. Thus, contrary to the linear kernel, the nonlinear relationship between classification outputs and features can be processed.

3.2.3. Random forest (RF)

There are two important parameters in RF algorithm. The first is the number of trees used in the forest, and the second is the number of random nodes to select on a node. For many classification problems, the number of variables to be randomly selected on each node is calculated by \sqrt{p} equation [54]. Where p represents the number of prediction variables in the data set. RF parameters were determined by experimenting with 5-fold cross validation and grid search algorithm over training data. It is seen that experiments with larger and deeper forests have no effect on performance. The best RF values were obtained from fifteen trees with maximum tree depth of three and one sample value per leaf. These parameters were used in RF.

3.2.4. ANN

There are different ANN models that give diagnostic results using medical field symptoms, biochemical analyzes and other characteristics such as age, gender [15]. These models have been shown to be useful tools in diagnosing complex diseases that are at risk of misdiagnosis.

The steps for creating the ANN architecture are as follows:

- (a) Layer 1 corresponds to the input vector containing input variables. The 15 parameters specified in the dataset were selected as input parameters of ANN. Specified as $I_1 - I_{15}$.

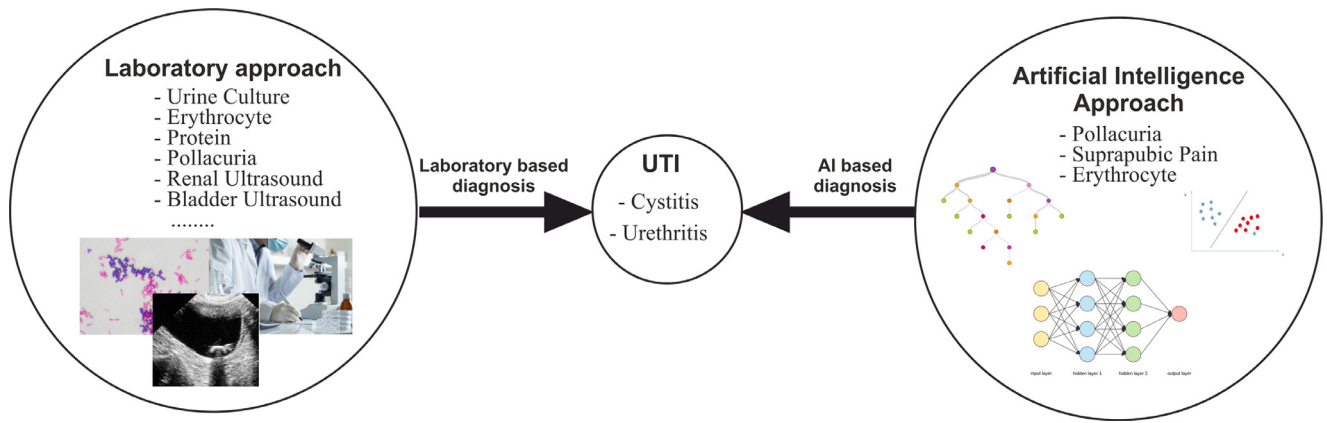


Fig. 8. Demonstration of the UTI classification with laboratory and artificial intelligence applications.

Table 3
ANN parameters.

Hidden layer number	1
Activation function, Hidden layer	Tansig
Activation function, Output layer	Softmax
Learning rate	0.01
Minimum performance gradient	1e-5
Performance goal	1e-3
Maximum number of epochs to train	5000

Table 4
Confusion matrix obtained for DT, SVM, RF and ANN classifiers.

Actual	DT predicted		SVM predicted		RF predicted		ANN predicted	
	P	N	P	N	P	N	P	N
P	43	2	44	1	43	2	44	1
N	2	12	1	13	0	14	0	14

- (b) In Layer 2, it is important to determine the number of neurons in the hidden layer within the ANN structure. In other words, the number of neurons in the hidden layer represents the network’s ability to generalize. As a result of the analysis realized, the number of neurons in the hidden layer that gave the optimal result was determined as 12.
- (c) Layer 3 is the output layer and has two outputs, cystitis, and nonspecific urethritis. These output parameters were also specified as O_1 and O_2 .

The ANN structure used in the study is given in Fig. 7.

In practice, a “multi-layer network structure” has been used, which is characterized as “forward feed”. In this study, a two-layer network structure as the multi-layer network was used. The network structure parameters in use are given in Table 3.

4. Results

In this study, it is aimed to develop a model that predicts cystitis and nonspecific urethritis diseases with similar symptoms from urinary tract infections. For this purpose, anamnesis, full urine and ultrasonic examinations of 59 patients who applied to the Urology Clinic for the Meram Medical Faculty Hospital were collected and a UTI dataset was created. Four different artificial intelligence methods, namely DT, SVM, RF and ANN, which are frequently utilized in medical diagnosis systems were used to create classification models on the UTI dataset and their classification performances were evaluated. All models have been tested with a 5-fold cross-validation method. Table 4 shows the confusion matrix ob-

tained from applying the DT, SVM, RF and ANN classifiers to the diagnostic problem of UTI.

The classification accuracy, sensitivity, specificity, positive predictive value and negative predictive value obtained using the confusion matrix are given in Table 5.

When Table 5 is examined, it is seen that SVM, RF and ANN classifiers have generally very high classification accuracy. If we evaluate the ANN classifier with the highest classification accuracy value of 98.30%, the positive value of the obtained model for cystitis finds the probability of really having cystitis as 100%. In addition, the probability of a patient with nonspecific urethritis diagnosed as truly nonspecific urethritis was determined to be 93.33%. When the patient is cystitis, the positive result of the model (i.e. the ability to recognize the cystitis disease) is 97.77%. The rate of recognizing non-specific urethritis was 100%.

Additionally, when the DT classification algorithm is examined, it is seen that it has a classification accuracy of 93.22% which can be considered as high due to pollacuria, suprapubic pain, and erythrocyte features. The DT classification algorithm shows that these three features have a great influence on distinguishing disease symptoms. The DT classification algorithm has helped us to reduce input features. A new ANN model with three inputs was created by using pollacuria, suprapubic pain and erythrocyte input features obtained from the DT method on the ANN artificial intelligence method, which has been successfully used in non-linear classification and has the highest degree of accuracy in this study. It is seen that the ANN model which uses parameters given in Table 3 and 5-fold cross-validation method has the same classification performance with accuracy of 98.30% as the ANN model given in Table 4. This shows that this feature selection process with DT may be combined with algorithms with high nonlinear classification performance.

RF algorithm has shown better performance than DT on UTI classification. In addition, the RF algorithm has very high specificity and positive predictive value along with the ANN algorithm.

5. Discussion

From a medical point of view, it is necessary to utilize a large number of symptom and laboratory findings in order to be able to diagnose UTI in the clinic. Even in some cases, subpubic ponction and urinary catheterization of both kidneys are used for localization. The two methods mentioned above are an invasive and complicated procedure. Especially, pollacuria and suprapubic symptoms are the most important indicators of cystitis. From this point of view, the results obtained from the DT method confirm the importance of these two input features.

Table 5
Performance values obtained for DT, SVM, RF and ANN classifiers.

		DT (%)	SVM (%)	RF (%)	ANN (%)
Classification Accuracy (%)	$\frac{TP+TN}{TP+FP+FN+TN} \times 100 =$	93.22	96.61	96.61	98.30
Sensitivity (%)	$\frac{TP}{TP+FN} \times 100 =$	95.55	97.77	95.55	97.77
Specificity (%)	$\frac{TN}{FP+TN} \times 100 =$	85.71	92.85	100	100
Positive Predictive Value (%)	$\frac{TP}{TP+FP} \times 100 =$	95.55	97.77	100	100
Negative Predictive Value (%)	$\frac{TN}{FN+TN} \times 100 =$	85.71	92.85	87.5	93.33

The end result of this study was that it would be possible to make a decision about the complicated urinary tract infection, by artificial intelligence methods using only erythrocyte values among suprapubic pain, pollacuria, and urinalysis results. As shown in Fig. 8, this ANN-based model structure can classify UTI without the need for different laboratory test results and ultrasound.

As a result, these classification algorithms, which we obtain high accuracy rates, appear to be an alternative to expensive laboratory and ultrasound tests. In addition, the acquired input parameters can give urologists an idea of the features to be used in distinguishing urinary tract infections. The ANN model obtained from this artificial intelligence method can be used as a medical decision support system for urologists conducting clinical examinations. The use of this method in the differential diagnosis of UTIs will make an important contribution to inexperienced physicians in this field, besides the advantages of avoiding time, cost and invasive methods.

This study can be improved by adding a diagnosis of pyelonephritis, which is a complication of untreated UTI. In addition, different hybrid artificial intelligence algorithms can be applied to the data set to increase the accuracy of the model.

Conflict of interest

The authors of this manuscript declare that they have no financial and personal relationships with other people or organizations that could inappropriately influence their work.

Acknowledgments

The authors would like to thank Necmettin Erbakan University Meram Medical Faculty Urology Polyclinic for its contributions to the study. The authors also would like to thank the reviewers and the editors for their valuable comments and contributions that helped to increase the readability and organization of the present paper significantly.

References

- [1] A.L. Komaroff, T.M. Pass, J.D. McCue, A.B. Cohen, T.M. Hendricks, G. Friedland, Management strategies for urinary and vaginal infections, *Arch. Intern. Med.* 138 (1978) 1069–1073.
- [2] W.E. Stamm, G.W. Counts, K.R. Running, S. Fihn, M. Turck, K.K. Holmes, Diagnosis of coliform infection in acutely dysuric women, *N. Engl. J. Med.* 307 (1982) 463–468.
- [3] M.F. Campbell, P.C. Walsh, *Campbell's Urology*, W.B. Saunders Company, 1998.
- [4] W.E. Stamm, S.R. Norrby, Urinary tract infections: disease panorama and challenges, *J. Infect Dis.* 183 (Suppl 1) (2001) S1–S4.
- [5] T.L. Griebing, Urologic diseases in America project: trends in resource use for urinary tract infections in men, *J. Urol.* 173 (2005) 1288–1294.
- [6] D.R. Hickling, T.-T. Sun, X.-R. Wu, Anatomy and physiology of the urinary tract: relation to host defense and microbial infection, *Microbiol. Spectr.* 3 (2015), doi:10.1128/microbiolspec.UTI-0016-2012.
- [7] C.M. Ginsburg, G.H. McCracken, M. Petruska, Once-daily cefadroxil versus twice-daily cefaclor for treatment of acute urinary tract infections in children, *J. Antimicrob. Chemother.* 10 (1982) 53–56.
- [8] L. Dixon, S. Pearson, D.J. Clutterbuck, Chlamydia trachomatis infection and non-gonococcal urethritis in homosexual and heterosexual men in Edinburgh, *Int. J. STD AIDS* 13 (2002) 425–426.
- [9] W.G. Baxt, Application of artificial neural networks to clinical medicine, *Lancet* 346 (1995) 1135–1138.
- [10] M.F. Abbod, D.G. Von Keyserlingk, D.A. Linkens, M. Mahfouf, Survey of utilisation of fuzzy technology in medicine and healthcare, *Fuzzy Set Syst.* 120 (2001) 331–349.
- [11] A.N. Ramesh, C. Kambhampati, J.R.T. Monson, P.J. Drew, Artificial intelligence in medicine, *Ann. R. Coll. Surg. Engl.* 86 (2004) 334–338.
- [12] D. Delen, R. Sharda, Artificial neural networks in decision support systems, in: *Handbook on Decision Support Systems 1*, Springer, Berlin Heidelberg, 2008, pp. 557–580.
- [13] I. Saritas, I.A. Ozkan, I.U. Sert, Prognosis of prostate cancer by artificial neural networks, *Expert Syst. Appl.* 37 (2010) 6646–6650.
- [14] M. Dagli, I. Saritas, Using artificial neural network for the prediction of anemia seen in behcet disease, *Energy Educ. Sci. Technol. Part A Energy Sci. Res.* 28 (2012) 1085–1092.
- [15] I. Saritas, Prediction of breast cancer using artificial neural networks, *J. Med. Syst.* 36 (2012) 2901–2907.
- [16] F. Amato, A. López, E.M. Peña-Méndez, P. Vañhara, A. Hampl, J. Havel, Artificial neural networks in medical diagnosis, *J. Appl. Biomed.* 11 (2013) 47–58.
- [17] A. Arvin, Introduction to medical applications of artificial intelligence, in: *Medical Applications of Artificial Intelligence*, CRC Press, 2013, pp. 1–8.
- [18] E. Kaya, M.E. Aktan, A.T. Koru, E. Akdoğan, Diagnosis of anemia in children via artificial neural network, *Int. J. Intell. Syst. Appl. Eng.* 3 (2015) 24–27.
- [19] P.J. Lisboa, A.F.G. Taktak, The use of artificial neural networks in decision support in cancer: A systematic review, *Neural Netw.* 19 (2006) 408–415.
- [20] N.I.R. Yassin, S. Omran, E.M.F. El Houbby, H. Allam, Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review, *Comput. Methods Prog. Biomed.* 156 (2018) 25–45.
- [21] F. Karimi-Alavijeh, S. Jalili, M. Sadeghi, Predicting metabolic syndrome using decision tree and support vector machine methods, *ARYA Atheroscler.* 12 (2016) 146–152.
- [22] S. Dreiseitl, L. Ohno-Machado, Logistic regression and artificial neural network classification models: a methodology review, *J. Biomed. Inform.* 35 (2002) 352–359.
- [23] D. Tomar, S. Agarwal, A survey on data mining approaches for healthcare, *Int. J. Biosci. Biotechnol.* 5 (2013) 241–266.
- [24] J. Sun, H. Li, Data mining method for listed companies' financial distress prediction, *Knowl. Based Syst.* 21 (2008) 1–5.
- [25] J. Han, J. Pei, M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, 2011.
- [26] J. Han, M. Kamber, J. Pei, 8-Classification: Basic Concepts, in: J. Han, M. Kamber, J. Pei (Eds.), *Data Mining (Third Edition)*, Morgan Kaufmann, Boston, 2012, pp. 327–391.
- [27] J.R. Quinlan, Induction of decision trees, *Mach. Learn.* 1 (1986) 81–106.
- [28] J.R. Quinlan, Simplifying decision trees, *Int. J. Man Mach. Stud.* 27 (1987) 221–234.
- [29] J.A. Suykens, J. Vandewalle, Least squares support vector machine classifiers, *Neural Process. Lett.* 9 (1999) 293–300.
- [30] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer science & business media, 2013.
- [31] E. Comak, New Approaches for Effective Training of Support Vector Machines, Department of Electrical-Electronics Engineering Selcuk University Graduate School of Natural and Applied Sciences, Konya, 2008.
- [32] C. Cortes, V. Vapnik, Support vector machine, *Mach. Learn.* 20 (1995) 273–297.
- [33] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, K.-R. Mullers, Fisher discriminant analysis with kernels, neural networks for signal processing IX, 1999, in: *Proceedings of the IEEE Signal Processing Society Workshop*, IEEE, 1999, pp. 41–48.
- [34] J. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, *Adv. Larg. Margin Classif.* 10 (1999) 61–74.
- [35] V. Kecman, *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*, MIT press, 2001.
- [36] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
- [37] A. Liaw, M. Wiener, Classification and regression by randomForest, *The R Journal* 2 (2002) 18–22.
- [38] K.J. Archer, R.V. Kimes, Empirical characterization of random forest variable importance measures, *Comput. Stat. Data Anal.* 52 (2008) 2249–2260.
- [39] Y. Qi, Random forest for bioinformatics, in: C. Zhang, Y. Ma (Eds.), *Ensemble Machine Learning: Methods and Applications*, Springer US, Boston, MA, 2012, pp. 307–323.
- [40] L. Breiman, A. Cutler, RF tools for predicting and understanding data, 2004, pp. 62.
- [41] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, PTR, 1994.
- [42] R.J. Schalkoff, *Artificial Neural Networks*, McGraw-Hill, 1997.

- [43] S.K. Pal, S. Mitra, Multilayer perceptron, fuzzy sets, and classification, *IEEE Trans. Neural Netw.* 3 (1992) 683–697.
- [44] L. Fausett, L. Fausett, *Fundamentals of Neural networks: architectures, algorithms, and Applications*, Prentice-Hall, 1994.
- [45] S. Koçer, A.E. Tümer, Classifying neuromuscular diseases using artificial neural networks with applied autoregressive and cepstral analysis, *Neural Comput. Appl.* 28 (2017) 945–952.
- [46] X. Liu, H. Du, G. Wang, S. Zhou, H. Zhang, Automatic diagnosis of premature ventricular contraction based on Lyapunov exponents and LVQ neural network, *Comput. Methods Progr. Biomed.* 122 (2015) 47–55.
- [47] K. Sabanci, A. Kayabasi, A. Toktas, Computer vision-based method for classification of wheat grains using artificial neural network, *J. Sci. Food Agric.* 97 (2017) 2588–2593.
- [48] M.W. Gardner, S. Dorling, Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences, *Atmos. Environ.* 32 (1998) 2627–2636.
- [49] W. Dai, W. Ji, A mapreduce implementation of C4. 5 decision tree algorithm, *Int. J. Database Theory Appl.* 7 (2014) 49–60.
- [50] V.N. Vapnik, V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [51] L. Breiman, J. Friedman, R. Olshen, C. Stone, *Classification and Regression Trees*, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific, California, 1984.
- [52] R.J. Schalkoff, *Artificial Neural Networks*, McGraw-Hill, New York, 1997.
- [53] D. Thukaram, H. Khincha, H. Vijaynarasimha, Artificial neural network and support vector machine approach for locating faults in radial distribution systems, *IEEE Trans. Power Deliv.* 20 (2005) 710–721.
- [54] H. Yılmaz, *Studying the Missing Data Problem in Random Forests Method and an Application in Health Field*, Eskişehir Osmangazi University, 2014.