A new proposal for early stage diagnosis of urinary tract infection using computers aid systems

Duygu Busra Aydın [1], Orhan Er [2]

Duygu Busra AYDIN; duyguaydin26@gmail.com
Bozok University, Department of Electrical and Electronics, 66200 Yozgat, Turkey.

Orhan ER; Corresponding Author; orhaner2009@gmail.com; +90 354 2421001
Bozok University, Department of Electrical and Electronics, 66200 Yozgat, Turkey.

Abstract

Hundreds of newborns everyday are affected by urinary tract infection worldwide. Urinary tract infection can cause serious illness over the long term. Early diagnosis is crucial for the treatment of the disease and the health of the newborn baby. In this study, a decision support system was established for the preliminary diagnosis of whether the newborn who was infected with urinary tract. For this purpose, artificial neural network methods and other bioinformatics techniques for comparison were used. Tests conducted with artificial neural networks resulted in: The probabilistic neural network method gave the best result for the test with 91.4251 ratio, whereas the multilayer neural network method showed the best result with 98.9130 ratio in training. Thus, it has been shown that the classification process with accuracy rate that can be considered important in the use of flexible computation and bioinformatics techniques in diagnosing urinary tract infections in newborn infants is successful.

1. Introduction

Urinary tract infection usually occurs as a result of abnormalities in the congenital urinary tract. It is quite difficult to understand the symptoms since they are still newborn. It should be investigated that the urinary tract infection may be the cause of babies being uncomfortable, having fever, having odor and color change in their urine, vomiting, infant abstinence, constant crying, and failure to thrive.

In the world, an average of 360,000 babies a day are born, some of which are born with urinary tract infection. The incidence is reported as 3-28 / 1,000 in girls and 1,5-7 / 1,000 in boys. Urinary tract infection is one of the most important and common infectious diseases of childhood [Kandur and Kupeli 2003].

Urinary tract infection, especially if not treated early, may cause damage to kidneys of newborns and children under 6 years of age. Urinary tract infections can progress to hypertension, pregnancy complications and chronic renal failure in later periods. Therefore, it is very important for the children born with this disorder to start treatment immediately [Erdoğan and Öner 2002].

Studies to develop new approaches to early diagnosis problems of diseases and studies inspired by biological systems to solve these medical early diagnosis problems are increasing every year. For this purpose, studies made with artificial neural networks provided important successes. In addition, software of decision support systems has been developed using these methods and helped in decision making process [Er 2009; Er et al. 2012; Er 2014; Koksal et al. 2016].

This study aims to develop a decision support system for pre-diagnosis of urinary tract infections in newborns, as opposed to specialist studies of medicine. In the application developed for this purpose, for the first time, learning vector quantization (LVQ), probabilistic (PNN), multilayer (MLNN-1) with one hidden layer and multilayer (MLNN-II) with two hidden layers neural networks were used for diagnosis.

Also DVM, KNN, LDA, QDA, Naive Bayes methods have been tried to test the success of the system and to show the performance of the system for comparison purposes. The data set used in these studies was taken from the clinical studies of a pediatrician who is working at Dışkapı Training and Research Hospital, a local hospital in Turkey [Aydin 2014].

2. Related work

Artificial neural networks or bioinformatics techniques have never been used before to diagnose this disease, and as understood from similar studies, diagnosis can only be made by medical doctors after physical and laboratory tests. These studies can be briefly explained as follows:

In the first study [Duran 2014], the authors investigated the activity of colorimetrically measured urine Leukocyte Esterase and nitrite levels in UTI diagnosis based on urine microscopy and urine culture. Leukocyte Esterase, pyuria, and nitrite tests are not sufficient to diagnose alone, and leukocyte and nitrite tests should be evaluated together for the patients who are clinically matching with UTI, and no absolute correlation with the culture results should be expected. Incorrect positives or negatives in Eliza tests, possible causes of sterile diarrhea, and very low colony counts (≤ 102 cfu / ml) that can be excluded from culture can be shown to be the reason.

In the second study [Benlidayı 2011] the authors state that urinary tract infections in children are one of the most common major health problems, and recurrent urinary tract infections can lead to chronic pyelonephritis, hypertension and chronic renal failure. One of the most important causes of chronic kidney failure in our country is reflux and chronic pyelonephritis which develop secondary to UTI. In the study, risk factors, etiology, clinical findings, etiologic factors, associated pathologies and the role of imaging modalities in UTI are investigated in UTI in children. As a result, because high incidence of renal scarring occurs in children with recurring UTI, it has been determined that early diagnosis, effective treatment and regular follow-up are very important to reduce long term complications significantly.

In the third study [Unal 2006], the authors conducted a study to determine the cause of fever in patients who complained of fever, to determine the incidence of urinary tract infections (UTIs) in patients in which the focus can't be determined, and to compare various urine analysis methods in detecting UTI. As a result, it is concluded that in the evaluation of little children with fever, history, physical examination and laboratory data can be used to determine whether high and low risk patients can be distinguished in terms of severe bacterial infection; UTI can also be a possible fever focus in patients in low risk group; and for diagnosis, urine culture must be taken with bladder catheterization.

In the fourth study [Itir 2016], the authors planned the study with the aim of assessing the validity of some symptoms and findings in urinary tract infections. Of these 4 complaints, hematuria (Sensitivity: 26.9 Specificity: 91.0), nocturia (Sensitivity: 46.7 Specificity: 73.1), turbidity (Sensitivity: 69.2 Specificity: 68.7), fever (Sensitivity: 64.1 Specificity: 59.5), high specificity (Specificity: 99.8) in serial use, high sensitivity (Sensitivity:91.0) in parallel use after evaluated together in parallel and serial use, are the values obtained.

3. Methodology

3.1 Data description

In this study, the parameters of 276 children report obtained from the clinical trial of a pediatrician who had expertise in Dışkapı Children's Hospital were used and 7 important measurement parameters (nitrite, leukocyte esterase, pyriform, blood culture, ultrasonography, bag culture, gender) considered to be helpful for diagnosis of disease by specialists were selected to designate as the system input. Artificial neural network methods (PNN, LVQ, MLNN -1 and MLNN -2) and bioinformatics techniques (DVM, KNN, LDA, QDA, Naive Bayes) were used with the aid of data set consisting of these parameters and it

was investigated whether the urinary tract infection disease could be diagnosed or not. For this purpose, suprapubic culture was identified as an exit parameter [Aydin 2016].

3.2 Diagnosis of the urinary tract infection disease using neural networks

Despite the existence of different biomedical preliminary diagnostic systems, this new study is presented in Figure 1 for the first time to diagnose urinary tract infection in newborn children. In this new model designed by us, the data set obtained with the guidance of the physician was divided firstly by applying the k-fold cross-validation techniques model to the training and test set. The set of data sets created for training was then trained with different artificial neural network models and other computational models determined for comparison purposes. Subsequently, classification with the applied test data set was performed. The classification result was applied as a decision support system for the physician [Aydin 2016].
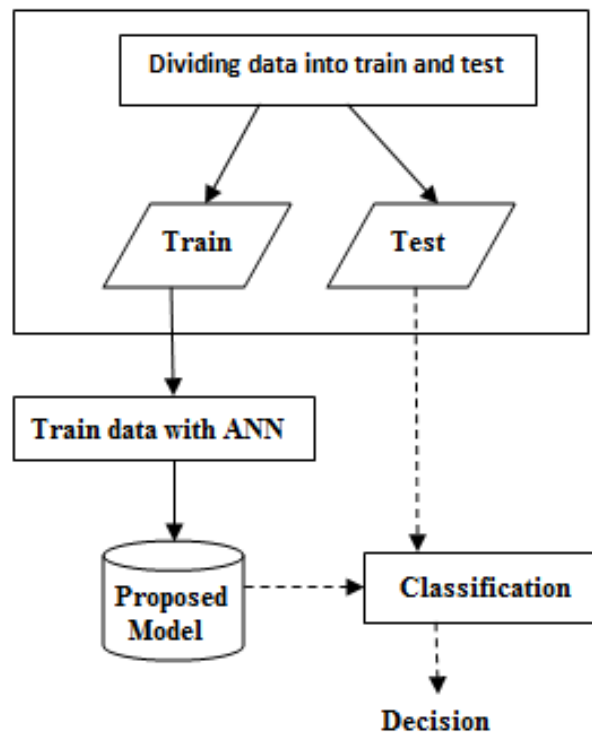

Figure 1 Proposed new system design for urinary tract infection (UTI) diagnosis system.

At the first step of the study, learning vector quantization (LVQ), probabilistic (PNN) and multilayer (MLNN) neural networks are used for UTI's disease diagnosis.

At the first stage of this study, a learning vector quantization neural network is used for the UTI's diagnosis. The network structure used for this purpose is shown in Figure 2. The LVQ structure used in this study has multilayer structure which consists of a hidden layer (competitive layer) and an output layer (linear layer). In this system, real valued input vector is feature's vector, and an output is index of two classes (Healthy and Urinary Tract Infection). The hidden layer consists of a set of competition functions [Kohonen 1990].
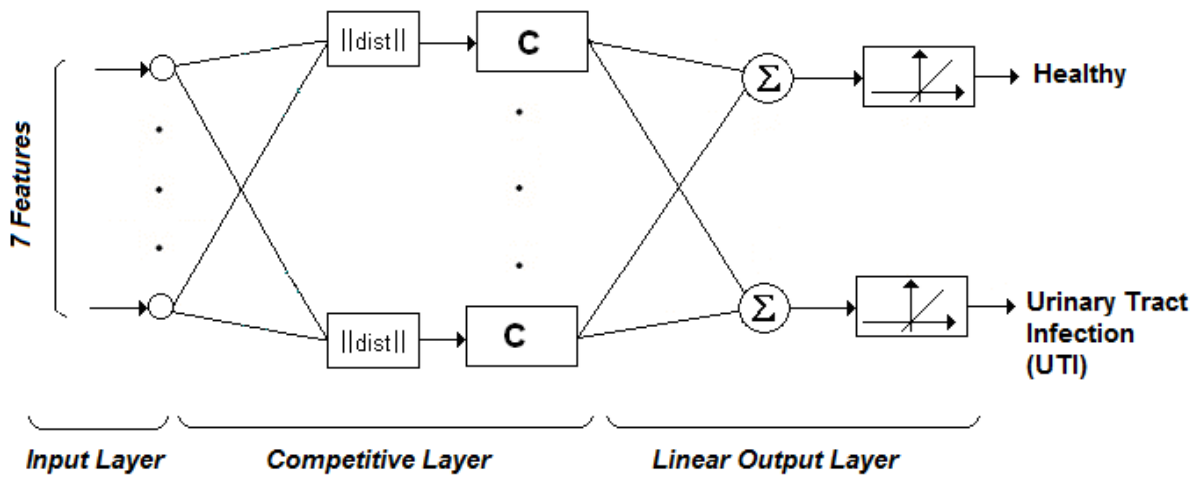
Figure 2 Implementation of LVQ NN structures urinary tract infection's disease diagnosis.

At the second stage of this study, a probabilistic neural network is used for the Urinary Tract Infection's diagnosis. The PNN structure used in this study has multilayer structure which consists of an input layer, a single hidden layer (radial basis layer), and an output layer (competitive layer) as shown in Figure 3. In this system, real valued input vector is feature's vector and two outputs are index of two classes (healthy and Urinary Tract Infection). All hidden units simultaneously received the 7-dimensional real valued input vector. The pattern layer consists of a set of radial basis functions. All of the radial basis functions are the same type (Gaussian) [Specht 1990].
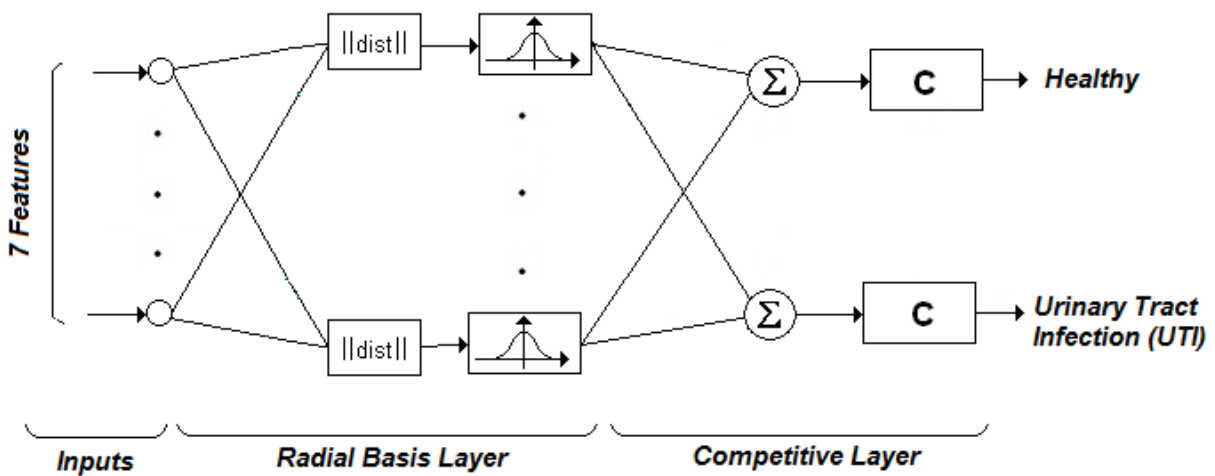

Figure 3 Implementation of PNN structures urinary tract infection's disease diagnosis.

At the third stage of this study, a multilayer neural network is used for the Urinary Tract Infection's diagnosis. The MLNN-1 (Figure 4-a) and MLNN-2 (Figure 4-b) structures used in this study has multilayer structures which consists of one and two hidden layers (competitive layer) and an output layer (linear layer). In this system, seven inputs are features, and two outputs are index of two classes (Healthy and Urinary Tract Infection). The network structure used for this purpose is shown in Figure 4.
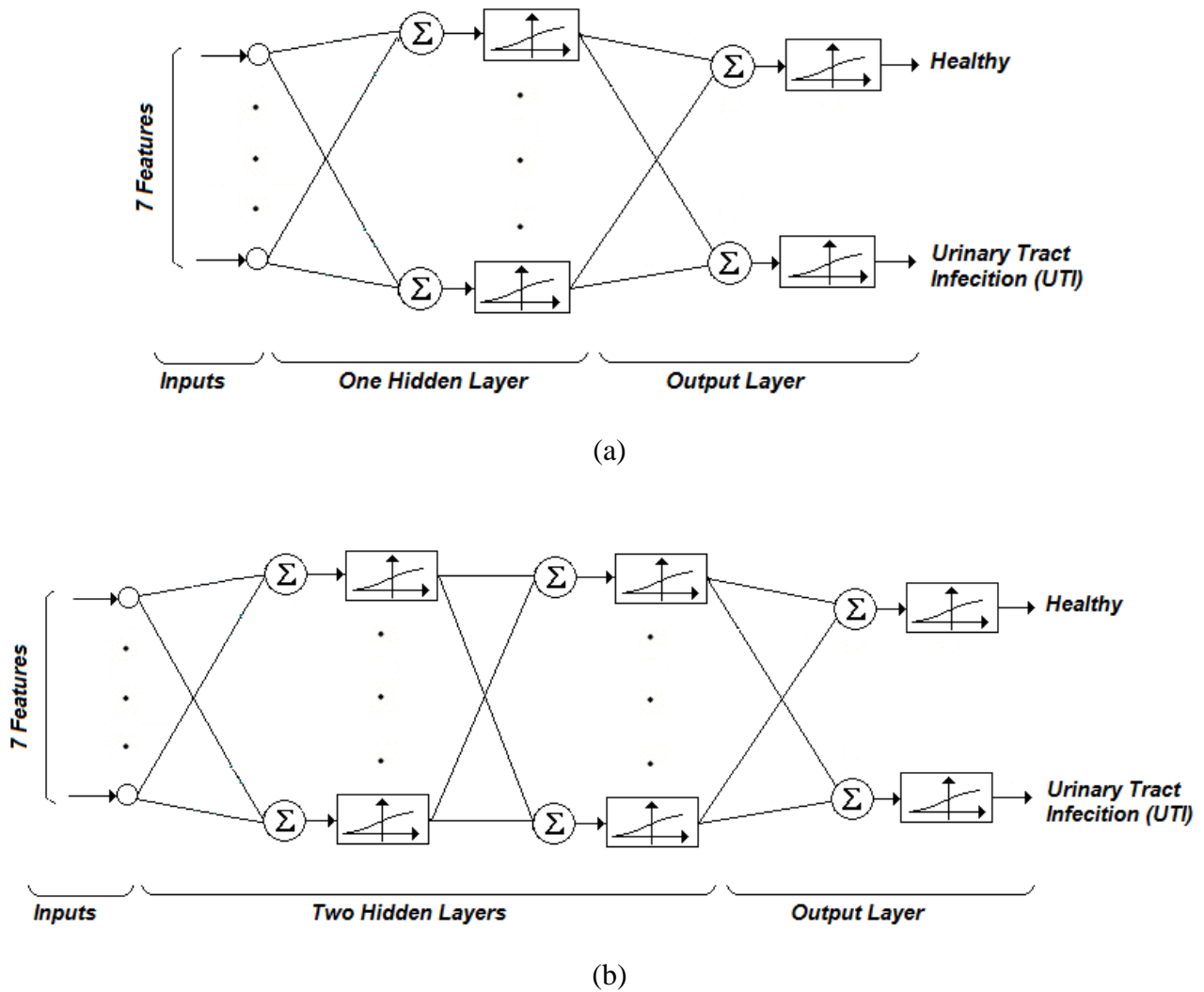
(a)

(b)

Figure 4 Implementation of MLNN-I (a) and MLNN-II (b) structure for Urinary Tract Infection's disease diagnosis.

The MATLAB neural network tool (License No: 834260) is used for the construction, training and testing of the NN structures. Detailed calculations about the construction and training NN structures can be found in references [Hagan et al. 1994; Gulbag et al. 2007; MathWorks 2007].

3.3 Validation of the estimation results

In this study, 3-fold cross-validation techniques were performed to compute the accuracy of the PNN, LVQ, MLNN-I and MLNN-II neural networks for UTI's disease diagnosis. In k-fold cross-validation, whole data are randomly divided to k mutually exclusive and approximately equal size subsets. The classification algorithm trained and tested k times. In each case, one of the folds is taken as test data and the remaining folds are added to form training data. Thus k different test results exist for each training-test configuration [Delen et al. 2005]. The average of these results gives the test accuracy of the algorithm. If a neural network learns the training set of a problem, it makes generalization to that problem. So, this type trained neural network gives similar result for untrained test sets also. But, if a neural network starts to memorize the training set, its generalization starts to decrease and its performance may not be improved for untrained test sets. The k-fold cross-validation method shows how good generalization can be made using neural network structures [Ozyilmaz and Yildirim 2002].

3.4 Measures for performance evaluation:

Classification accuracy [Er 2009; Er et al. 2012; Er 2014; Koksal et al. 2016] is used for the comparison of studies reported in our study focusing on UTI's disease diagnosis and using the same raw data. So, the results obtained by neural networks are compared with the results of the tested other algorithms by using classification accuracy.

Equations which used in the classification accuracies are shown in Eq.1 and Eq.2:

$$classification\ accuracy(N) = \frac{\sum_{i=1}^{|N|} assess(n_i)}{|N|} \quad , \quad n_i \in N \tag{1}$$

$$assess(n) = \begin{cases} 1 & if\ classify(n) = nc \\ 0 & otherwise \end{cases} \tag{2}$$

where N is the set of data items to be classified (the test set), $n \in N$, nc is the class of the item n and $classify(n)$ is the output of neural network.

4. Experimental results

In this study, YSA model was investigated by using different algorithms benefiting training algorithms to determine whether newborn infants had urinary tract infections or not.

For this purpose, parameters of 276 children report obtained from the clinical trial of a pediatrician who has expertise in Ankara Children's Health and Diseases Hematology Oncology Training and Research Hospital were used and 7 input and 1 output parameters considered by expert physicians to be helpful to diagnose were selected. In the data set consisting of these parameters, flexible calculation and bioinformatics techniques were used and the results shown in Table 1 were obtained.

Table 1 Performance averages of training and test results obtained with artificial neural network models on UTI.

		Train Accuracy	Test Accuracy	Training Average	Test Average
PNN	1-Fold	98.0072	90.2174	98.0072	90.2174
	3-Fold	98.0072	89.4928	98.7922	**91.4251**
		98.3696	95.6522		
		100	89.1304		
LVQ	1-Fold	94.5652	91.3043	94.5652	91.3043
	3-Fold	95.6522	88.4058	95.6522	90.6135
		95.1087	93.4783		
		96.1957	89.9565		
MLNN - 1	1-Fold	98.3696	86.9565	98.3696	86.9565
	3-Fold	98.3696	86.9565	97.8261	89.1304
		95.1087	92.3913		
		100	88.0435		
MLNN - 2	1-Fold	98.3696	84.7826	98.3696	84.7826
	3-Fold	98.3696	85.8696	**98.9130**	86.2319
		98.9130	88.0435		
		99.4565	84.7826		

As shown in Table 1, when the best diagnostic accuracy rate is obtained by the two-layer MLNN structure in the training process, the best accuracy rate for the test operation is obtained with the PNN artificial neural network model. Secondly, the best accuracy is in the PNN method for training, and in the LVQ method for testing. Thirdly, the best accuracy rate is in the single layer MLNN method for training and in the single layer MLNN method for testing. For the fourth and last, the best accuracy rate is in the LVQ method for training and in the two-layer MLNN method for testing.

In addition, the number of exemplary cases that have been subjected to inaccurate training and inaccurate test procedures according to training and test results are given in Table 2 below.

Table 2 Incorrect training and test data counts of samples.

	N-Fold	Number of data used for training	Number of incorrectly trained data	Number of data used for testing	Number of incorrectly tested data
PNN	1 Fold	194	4	82	9
	3 Fold	184	3	92	7
LVQ	1 Fold	194	11	82	7
	3 Fold	184	8	92	9
MLNN - I	1 Fold	194	4	82	11
	3 Fold	184	4	92	11
MLNN - II	1 Fold	194	4	82	13
	3 Fold	184	3	92	13

Thus, it has been shown that the algorithms used can be used for the task of classification in diagnosing the disease. According to these results, it was concluded that the result of the study, which was developed as a decision support system for the pre-diagnosis of the disease and would help to diagnose the disease, was found to be successful.

In addition to these, performance studies were conducted with 5 methods for comparison. These are DVM, KNN, LDA, QDA and Naive Bayes. The results according to these methods are summarized in Table 3.

Table 3 Performance averages of training and test results obtained with other comparative methods on UTI.

		Train Accuracy	Test Accuracy	Training Average	Test Average
DVM	3-Fold	90,760	73,913	87,137	77,173
		83,695	82,608		
		86,956	75,000		
KNN	3-Fold	96,739	84,782	**94,383**	**82,608**
		92,391	88.043		
		94,021	75,000		
LDA	3-Fold	89,130	77.173	86,956	76,448
		84,239	81.521		
		87,500	70,652		
QDA	3-Fold	95,108	72.826	92,753	78,260
		90,760	83.695		
		92,391	78,260		
Naive Bayes	3-Fold	29,347	65,214	29,347	53,621
		29,347	41,304		
		29,347	54,347		

In general, both the literature studies and the results of this thesis show that YSA methods (PNN, LVQ, MLNN-I, MLNN-II) give better diagnostic accuracy than other methods used for comparison (DVM, KNN, LDA, QDA, Naive Bayes). The main reason for this is that the performance calculations of the algorithms are based on different algorithms. It has also been found that the data set has an extremely high impact on the performance of the algorithms. In this respect, the selection of the algorithm and the optimization of the input data are important to increase the diagnostic accuracy.

5. Conclusion

When the study is evaluated in general, it is seen that there is no study on the urinary tract infection of newborn infants in our country. Moreover, obtaining the data sets necessary for the development of this system is also very difficult. In the literature, classification problems about disease diagnosis have been found to be inadequate and out of date. It is concluded that the studies in this field is particularly focused on specific topics and newborn babies is neglected. Therefore, unlike the studies in the literature, it is aimed to decide a support system that measures whether the infants in the newborn intensive care unit are diseased. Thus, we have achieved a considerable result with our study.

It is thought that this study will contribute to early diagnosis and diagnosis studies with a new and different approach to other pediatric diseases in the newborn unit.

The algorithms used for practice of this decision support system, which can be used in daily clinical applications of physicians, have achieved successful results in many fields as well as in the medical field. Thus, the success of artificial neural networks in applications is once again seen with this thesis study.

In this project, an early diagnostic system is designed for the diagnosis of urinary tract infection. With this developed system structure, it also enables to serve people in different medical fields. It is thought that successful classification can be done on different types of diseases to be determined. Correct selection of the properties to be put into system and the correct sampling of the data that will characterize diseases will contribute to the success of the result. As the development of these systems takes place under the supervision of a specialist physician, it is thought that this may provide a basis for other works in this field.

References

Aydin DB. 2016. "Neural network based decision support system that detects for newborn urinary tract infection." *Master Thesis*, Bozok University.

Aydın S. 2014. "Yenidoğan Yoğun Bakım Ünitesinde İdrar Kültürü Alma Yöntemlerinin İdrar Yolu Enfeksiyonunu Tanımlamadaki Yeri." *Ankara Çocuk Sağlığı ve Hastalıkları Hematoloji Onkoloji Eğitim ve Araştırma Hastanesi*, Ankara.

Benlidayi YA. 2011. "Tekrarlayan idrar yolu enfeksiyonu geçiren çocukların retrospektif taranması ve renal skar üzerine etkisinin değerlendirilmesi." *Kahramanmaraş Sütçü İmam Üniversitesi Tıp Fakültesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı*, Tez no: 308184.

Delen D, Walker G. and Kadam A. 2005. "Predicting breast cancer survivability: A comparison of three data mining methods", Artificial Intelligence in Medicine 34(2) 113-127.

Duran, O. 2014. "İdrar yolu enfeksiyonu düşünülen hastalarda nitrit ve lökosit esteraz pozitifliğinin idrar mikroskopisi ve kültürü ile karşılaştırılması." *Cumhuriyet Üniversitesi Tıp Fakültesi Acil Tıp Anabilim Dalı*, Tez No: 386681.

Er O. 2009. "Thoracic Disease Diagnosis using Flexible Computing and Bioinformatics Computing System." *PhD Thesis*, Sakarya University.

Er O, Yumusak N, Temurtas F. 2012. "Diagnosis of chest diseases using artificial immune system." *Expert Systems with Applications* 39:1862.

Er O. 2014. "Network intrusion detection: a usability comparison study on neural networks." *Electronics World*, 120 (1943): 26-32.

Erdogan O, Oner A. 2002. "Sami Ulus Çocuk Sağlığı ve Hastalıkları Eğitim ve Araştırma Merkezi Pediatrik Nefroloji Departmanı" *Türkiye Klinikleri J Pediatr* 11(4):221-35, Ankara.

Gulbag A, Temurtas F. 2007. "A study on transient and steady state sensor data for identification of individual gas concentrations in their gas mixtures." *Sensors and Actuators B-Chemical*. 121(2): 590-599.

Hagan MT, Menhaj M. 1994. "Training feed forward networks with the Marquardt algorithm." *IEEE Trans. on Neural Networks.* 5: 989-993.

Itir EE. 2003. "Bazı semptom ve bulguların idrar yolu enfeksiyonu tanısındaki geçerliliğin değerlendirilmesi." *Erciyes Üniversitesi Sağlık Bilimleri Enstitüsü Halk Sağlığı Anabilimdalı,* Tez No: 129733.

Kandur Y, Kupeli S. 2003. "Vezikoüreteral Reflü ve İdrar Yolu Enfeksiyonu." *Klinik Pediatri*. 2(2): 69-73.

Kohonen T. 1990. "Improved versions of learning vector quantization." *In Proc. of the IEEE International Joint Conference on Neural Networks*. pp. 545–550, New York.

Koksal AS, Er O, Evirgen H, Yumusak N. 2016. "A secure communication using cascade chaotic computing systems on clinical decision support." *Biomedical Engineering-Biomedizinische Technik*. 61(3): 323-329.

MathWorks. 2007. "Matlab Documentation, The MathWorks." *Inc. 3 Apple Hill Drive.* Natick, MA 01760-2098, USA.

Ozyilmaz L and Yildirim T. 2002. "Diagnosis of thyroid disease using artificial neural network methods", *In Proceedings of ICONIP'02 9th International Conference on Neural Information Processing*, Singapore: Orchid Country Club, 2033-2036.

Specht DF. 1990. "Probabilistic neural networks", *Neural Networks*. 3: 109–118.

Unal B. 2006. "Odağı belli olmayan ateşli çocuklarda idrar yolu enfeksiyonunu saptamada çeşitli idrar inceleme yöntemlerinin karşılaştırılması." *Mersin Üniversitesi Tıp Fakültesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı,* Tez No: 192974.