Innovative Systems Design and Engineering
ISSN 2222-1727 (Paper) ISSN 2222-2871 (Online)
Vol.5, No.5, 2014

Case-Based-Reasoning System for Feature Selection and Diagnosing Disease; Case Study: Asthma

Somayeh Akhavan Darabi

Department of economic, management and accounting, Payame Noor University, Iran.

Email: akhavan.star@yahoo.com

Babak Teimourpour (Corresponding author)

Department of Information Technology, Tarbiat Modares University, Tehran, Iran.

E-mail: b.teimourpour@modares.ac.ir

Maryam Zolnoori
Mathematic and informatics group, Academic Center for Education, Culture and Research (ACECR), Tarbiat Modares University, Tehran, Iran
Email: M.Zolnoori@gmail.com

Dr.Hassan heydarnejad

Department of Medical, Shahid Beheshti University, Tehran, Iran.

Email: heidarnazhad@nritld.ac.ir

Abstract

Asthma is a chronic informatory disease of the respiratory canals in which it has not become obvious what is the reason for the reports argumentation on the ground of asthma prevalence. In the present research, the purpose would be to design a case-based-reasoning (CBR) model in order to assist a physician to diagnose the type of disease and also the needed therapy.

At first for designing this system, the disease variables were discriminated and were at the patients' disposal as a questionnaire, and after gathering the relevant data (CBR) algorithm was rendered on the data which led to the asthma diagnosis. The system was tested on 325 asthmatic and non asthmatic adult cases and was accessed with eighty percent accuracy. The consequences were promising. With regard to the fact that the factors of the disease are different in various countries, This study was performed in order to determine risk factors for asthma in Iranian society and the results of research showed that the most important variables of asthma disease in Iran are symptoms heperresponsivity, frequency of cough, cough.

Key words: data mining, case based reasoning, asthma, diagnosis.

Introduction

Asthma is a heterogeneous disease and the genetic, atopic and environmental factors such as viruses, occupational contacts and the allergens have function in its prevalence and its persistence. Atopy is the most predominant single risk factor for generating asthma. Allergic asthma is often accompanied with the individual or the family allergic diseases (such as rhinitis, with the rash and eczema (Callahan and et al, 2012). In accordance with the recent researches, 4.35 percent of Tehran 10 million populations reveal asthma symptoms that it is a catastrophe.

Asthma has also more morbidity in the developed countries: but it has more asthmatic mortalities which can be ascribed to the third world countries. The other reasons can be attributed to the nonexistence of sufficient specialists, the lack of necessary facilities for diagnosis and the ignorance of people in this field which is more significant than other reasons.

Asthma affects approximately on 300 million people worldwide, making it one of the most common chronic disease in the world (Gunnbjörnsdóttir, 2004). Despite the significant improvement in understanding the asthma mechanism and pathology there is a remarkable difference between prevalence of asthma related systems especially in developing countries. Stating that asthma usually is under diagnosed (Riedler, 2001). The asthma non diagnosed consequences include remarkable effects and create severe restrictions and the emotional aspects of the individual's lives, the effect and the vocational route and the children's learning.

Within the medical community, there has been significant research into preventing clinical deterioration among hospital patients. Data mining on electronic medical records has attracted a lot of attention but is still at an early stage in practice. Clinical study has found that 4–17% of patients undergo cardiopulmonary or respiratory arrest while in the hospital (Commission, 2008). Early detection and intervention are essential to preventing these serious, often life-threatening events. Indeed, early detection and treatment of patients with sepsis has already shown promising results, resulting in significantly lower mortality rates (Joans and Brown, 2008).

Case based reasoning (CBR) technique is one of the data mining tools which are applied for the medical diagnosis, anticipating the chronological order and so forth.

Literature Review

There are different kinds of studies for Data Mining techniques in medical databases. We identify the following categories:

1. Studies that summarize reviews and challenges in mining medical data in general (R.D. Canlas Jr, 2009), (Satyanandam and et al, 2012), (Hosseinkhah and et al, 2009), (Dakheel and et al, 2011), (Wasan and et al, 2006).

2. Studies of Data Mining techniques used for diagnosing and/or prognosing of specific diseases, which can be further classified into three other categories: those which use Data Mining techniques for disease diagnosis (Kumari and Godara, 2011), (Soni and et al, 2011), (Chi-Ming Chu and et al, 2009), (Prasad and et al, 2011), (Dangare and Apte, 2012), (Aftarczuk, 2007), for disease prognosis (Srinivas and et al, 2010), (Aljumah and et al, 2011), (Delen, 2009), (Osofisan and et al, 2011), (Floyd, 2007), (Kika and et al, 2010), or both diagnosis and prognosis (Gupta and et al, 2011), (Huang and et al, 2007).

3. Studies to investigate factors which have higher prevalence of the risk of a disease (Karaolis and et al, 2009), (Yang and et al, 2010).

4. Studies that present new technologies and algorithms (McCormick and et al, 2012- Ullah, 2012), (Habrard and et al, 2003), (Kusiak, 2001) and studies that present new techniques improving old ones, such as (Kavitha and et al, 2010), (Ha and Joo, 2010), (Parvathi and Palaniammalì, 2011), (Gao and et al, 2005).

5. Studies that present new frameworks, tool and applications in medicine and healthcare system (Shantakumar and Kumaraswamy, 2009), (Shukla and et al, 2009 – Duan and et al, 2011), (.Kumar and et al, 2011), (Palaniappan and Awang, 2008 – Jimenez and et al, 2008), (Sakthimurugan and Poonkuzhali, 2012).

Misra and Dehuri in their research paper Functional Link Artificial Neural Network for Classification Task in Data Mining created a Functional Link Artificial Neural network and compared its classification performance with other machine learning algorithms (Misra and Dehuri, 2007). Their FLANN has given 21.87% miss classification performance and MLP has given 24.8 % classification performance.

Lee and his colleagues used new data mining mechanism for the purpose of asthma attack disease prediction. These two methods are called the tree decision depended upon the law-based pattern. The accuracy of this system is 84.12 percent (Lee and et al, 2010). Nordquist and his colleagues studied about the rendering of the learning operation. As a matter of fact the aim was to determine the success level of performing the learning process on the basis of the sample (Nordquist and et al, 2012). In the article Chae and Hoo compared the two nervous lattices algorithms and case based reasoning (CBR) (Chae and Hoo, 1996).

Zolnoori and his colleagues created a phase proficient system for the asthma the age of 6-18 years old which system sensitivity is 88 percent and its specificity is l00 percent (Zolnoori and et al, 2010).

Chooi and his colleagues demonstrated that where there aren't necessary feasibilities for evaluating the respiration by using the computer program causes more ascension for the asthma diagnosis accuracy that in this essay attention has been paid to this subject (Chooi and et al, 2007). Bibi and his colleagues created a nervous lattice for the target of foretelling the asthma respiratory symptoms and COPD disease and bronchiectasis (Bibi and et al, 2001).

The input variable of this system is the air pollution. This system is able to predict with 12 percent of error. The designed system of asthma of the spirometry test facilities, so that this task causes saving in time and expense.

What which will come in continuity with this paper:

The system methodology will be surveyed in the methodology system. The (CBR) system will be described in the next section, the system performance test and evaluation in the "system performance evaluation section will be discussed and at the end, the discussion and adding up and can conclusion in the section of" "Discussion and can conclusion" will be reviewed.

Methodology

In the present research a case-based-reasoning system for the purpose of diagnosing the disease more truly and accurately followed by appropriation therapy has been designed. For the sake of creating and developing this system, at first the efficient variables in the asthma disease have been identified. The creation of this system includes determining the system inputs, weighting the variables, ascertain CBR algorithm structure, evaluation and performance test which will be discussed in this paper. The overall framework of the model which is presented in this section in Fig. 1 has been illustrated.

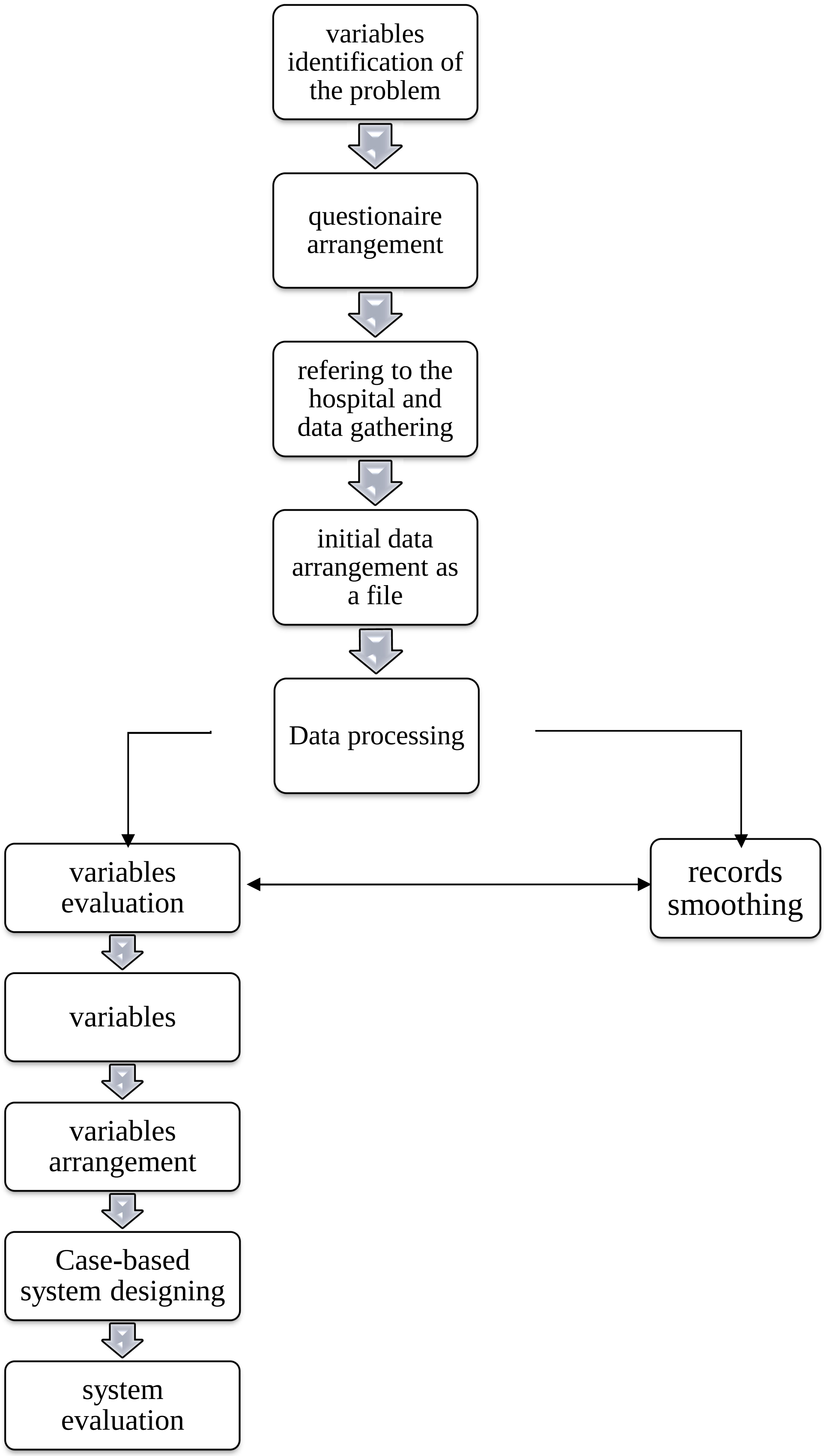


Figure1: Fundamental Method

After identifying the variables a questionnaire was arranged in conformity with it and after word by filling out the questionnaire through the patient's enquiry under the physician surveillance. The data gathering has been conducted. The data include asthmatic and non asthmatic patients that have been brought as a database in the excel table. Data processing: the data needed preprocessing in order to be applied in the (CBR) system. This section is very important and time consuming and it involves, following steps:

1. Deleting the data columns, file code, computer code, age, gender and the number of the patient's column.
2. We made the multi substandard variables ordinal in order to decrease the number of variables (ordinal in order to decrease cough, wheeze, …… column)
3. Improvement and repairing of noise and error data
4. Filling the misses by KnnImputation
5. Pert data identification and deletion
6. Editing the duplications
7. Mixing up the genetic factors which blend with each other for the reason of identical significance
8. Data normalization: the variables are classified into the classes of medical history (Ernst and et al, 2002), environmental factors (Bousquet and et al, 2003), allergic rhinitis (Shapiro and et al, 2006), and genetic factors.

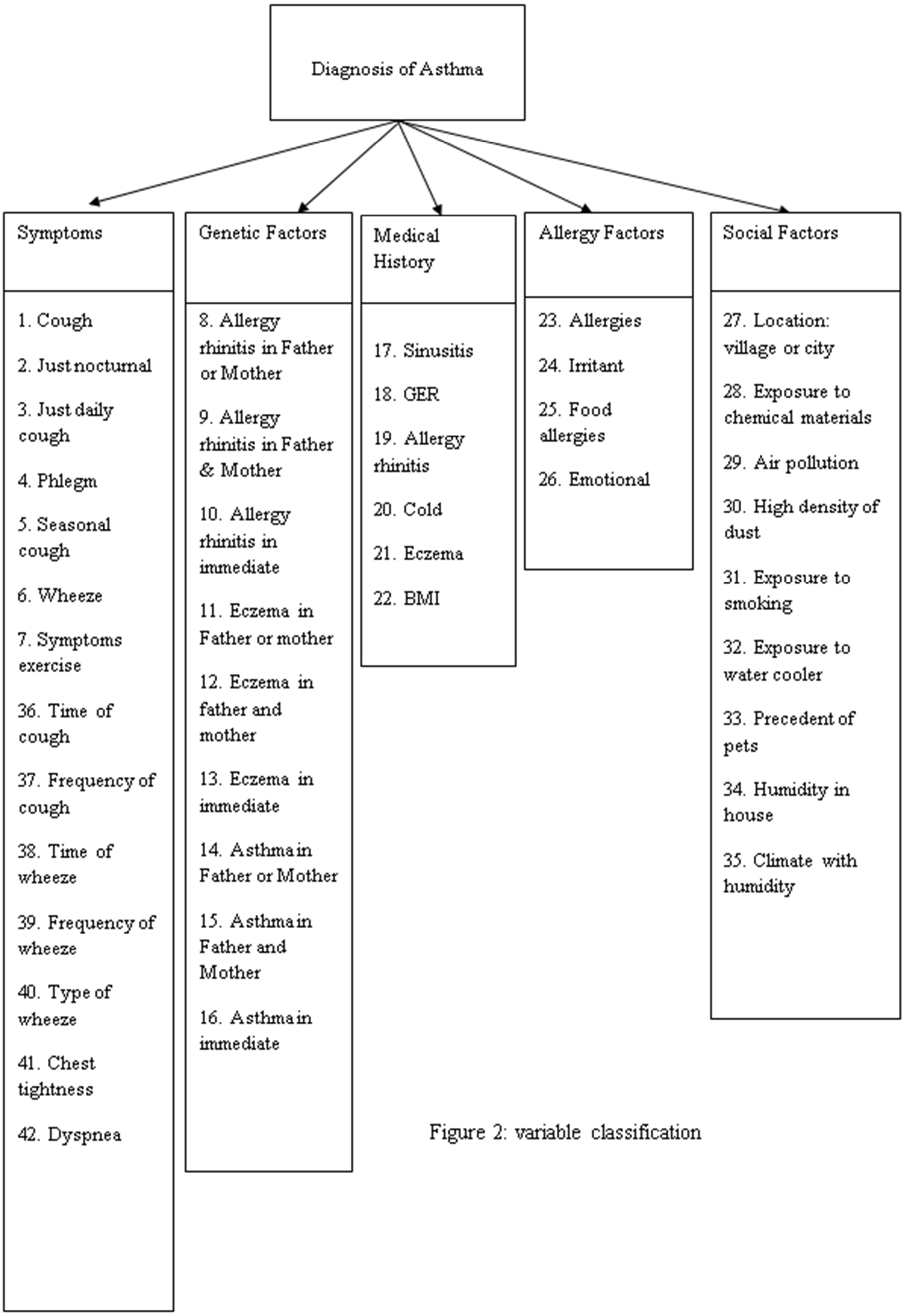


Figure 2: variable classification

Feature Evaluation

Variables evaluation is an important problem in statistical research and machine learning methods. Feature selection algorithms single out a more effective set of data matrices.

Feature evaluation algorithms attribute one rank to each variable that shows the effect of that variable in the classification of the classes.

The more the rank of a variable is, the more share that rank has in the calculations. In fact some sort of weighting variables is performed.

Weighting variables: for weighing the variables two methods of weighting named RandomForest and p-value have been applied. The system was evaluated by the both methods. The random forest method yielded better results. In the test with assumption of $x^2=0$, there exist two independent variables and if p-value becomes less than the pre-determined amount of a = 0.05 or 0.1

Terminated to the zero test refusal and this means that the less the amount of p-values, the more dependent those both variables are. On one hand, the more a variable is closer to the target variable, the more influence it has on it. So the inverted values of p-value as a feature weight are used. (With obeying the rest of the weight vector provision like when total weights becomes one. RandomForest method samples the feature randomly. Each time sets a feature aside and form a tree with the remaining features and sums the error rate average up. The more the error rate becomes further by setting is it aside, the more importance of that feature is revealed.

Accordingly the feature is weighted.

Smoothing the records:

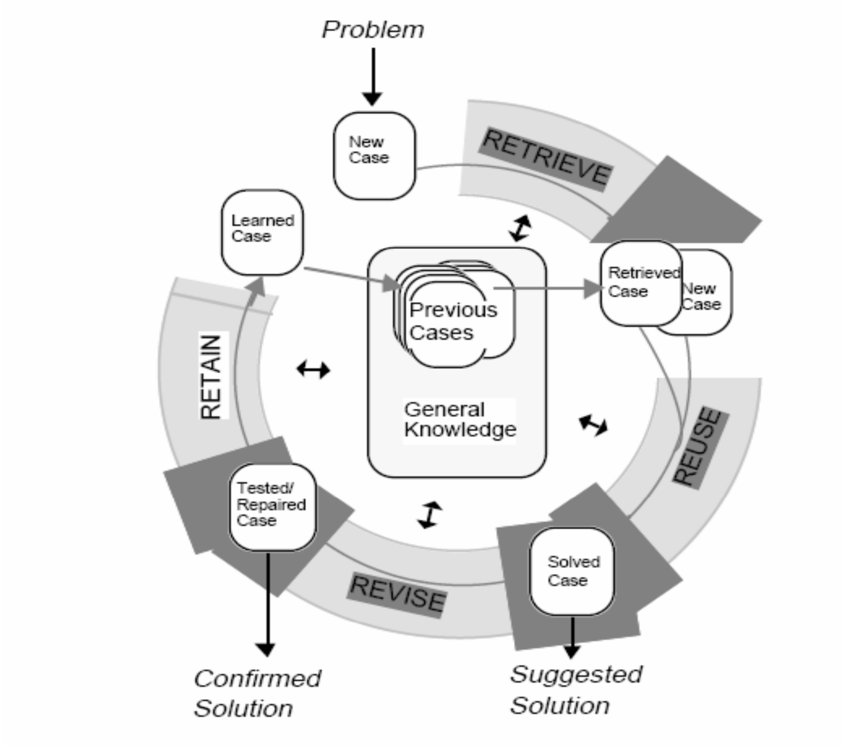
By applying the smoothing method (minimizing large classes and creating artificial data for smaller classes) some changes were made in the data number. Data arrangement: for multilevel variables changed into the ordinal ones.

(CBR) algorithm structure: Case-based-reasoning method has been formed on the basis of using the previous problem response for the solution of the new identical problems.

(CBR) is known as a method which has modeled the human behavior quality in conformity with the new problems; hence forth it uses the acquired experiences in solving the previous problems as a guide for the new problem solution (10).

A (CBR) system in order to find a solution for the new difficulty passes through a four step process. The steps of this process form a ring which has been illustrated in figure 2.

- Retrieving the identical previous cases.
- Reusing the cases for new problem solution.
- Revising the suggested solution in cases of necessity.
- Retaining the given solutions as a new case.



-

Figure 2: case based reasoning processing (Marir and Watson, 1994).

Case: It determines a description in series of questions which must be asked and the possible responses for them and also the operations that must be done for each of them. In this essay, the sort of database is smooth, In other words, it means that all the cases have a series of identical features.

Retrieving step: this step includes retrieving the existed identical cases in the cases library in which the present similar difficulty has been confronted the closest neighborhood method is the most fundamental method of distance based learning method. After summing up the similarity among the various variables by using the algorithm (1), the similarity between the two records are calculates (Bagherjeiran and et al, 2006).

Algorithm (1) shows the similarity among the records with heterogeneous variables:

1. For K_{th} variable, calculate and normalize the similarity between the two records of S(X,Y).
2. For K_{th} variable define the k pointer as follows : $\delta_k = \begin{cases} 0 \\ 1 \end{cases}$
3. Calculate the overall similarity between the two records by applying the following phrase :

$$similarity(x,y) = \frac{\sum_{k=1}^{n} w_k \times \delta_k \times s_k(x,y)}{\sum_{k=1}^{n} \delta_k}$$

$$\sum w_k = 1 \text{ که در آن} \quad (1)$$

The phase 2 and 3 exhibit that at first every feature weight must be included in the distance calculation , secondly , when the K_{th} variable of one or both records are lost or for asymmetric variables when both records have the sum of zero, that K_{th} variable doesn't enter the calculations and thirdly the distances must eventually be normalized too.

Evaluating system performance and result

As a result, 325 samples of asthmatic and nine asthmatic patients were gathered in Masih Daneshvari hospital in Tehran city of Iran in 2012-2013.

These patients were adults and included men and women. For the purpose of increasing system accuracy, the non asthmatic patients sample had difficulty in their lungs. For performing the CBR system the R software was used.

The system performance was tested and evaluated by the specialist physicians. The system evaluation standards are specificity, sensitivity and accuracy.

After deleting the data, 30 percent of the database data was selected randomly for the system test and remaining 70 percent were used for making the system cases.

The consequences given in this section is the result of the system experiment on the test data (as shown in Table 1).

Sensitivity: Represents the ratio of the positive cases which mark their model as a plus or positive mark.

Specificity: Represents the ratio of the negative cases that the experiment marks them truly as a negative mark or minus.

Table 1: System evaluation results based on test data

Weighting with random forest	Weighting with p - value
K=5 accuracy = 80 percent sensitivity = 65 percent specificity = 84 percent	K=5 for each accuracy = 68 percent sensitivity = 17 percent specificity = 76 percent

As these results display, the RandomForest gives more favorite conclusions. In the feature selection step, the computerized and the statistical methods of p-value and the RandomForest have been used and a weight has been given to each feature (as shown in Table 2, 4). By applying the rFCV order and R software, the weights were gained from the RandomForest and where it gave the least error, the features numbers were selected. The model selected 32 variables with the least error (as shown in Table 3).

Table 2: Weighting with RandomForest

Features	Weights with RandomForest
cough	4.309436367
Frequency of cough	5.457837777
Just nocturnal cough	0.505595905
Just daily cough	0.688721699
Phlegm.1	0.586330282
phlegm.2	1.442354080
phlegm.3	0.606842163
phlegm.4	0.217611282
phlegm.5	0.775966010
seasonal cough	0.487514697
Wheeze	3.398243446
Frequency of wheeze	3.924343041
Type of wheeze	2.721408917
Dyspnea	3.380217743
Chest tightness	3.192712928
exercise.1	1.030655704
exercise.2	1.337704371
allergic rhinitis time	0.973531154
allergic rhinitis severity	1.475738453
cold.1	1.012688933
cold.2	0.895179984
cold.3	1.189704582
Sinusitis	1.087116772
GER	1.443680268
Genetic factors.1	0.432959880
Genetic factors.2	0.026558714
Genetic factors.3	0.195749195
Genetic factors.4	1.137898915
Genetic factors.5	0.005041667
Genetic factors.6	1.305793839
Symptoms heperresponsivity.1	1.600139042
symptoms heperresponsivity.2	4.206182003
symptoms heperresponsivity.3	6.550478915
symptoms heperresponsivity.4	2.093366198
BMI	1.007245004
eczema	0.799527860
social factores.1	0.674308672
social factores.2	0.217894946
social factores.3	1.029612446
social factores.4	0.509233286
social factores.5	1.077984397

Table 3: Ranking Features with RandomForest

```
Symptoms heperresponsivity.3
Frequency of cough
cough
symptoms heperresponsivity.2
frequency of wheeze
Wheeze
Dyspnea
Chest tightness
Type of wheeze
Symptoms heperresponsivity.4
Symptoms heperresponsivity.1
Allergic rhinitis severity
GER
phlegm.2
exercise.2
genetic factors.6
cold.3
genetic factors.4
Sinusitis
Social factores.5
exercise.1
social factores.3
cold.1
BMI
Allergic rhinitis time
cold.2
eczema
phlegm.5
just daily cough
social factores.1
phlegm.3
Phlegm.1
Social factores.4
just nocturnal cough
seasonal cough
genetic factors.1
social factores.2
phlegm.4
genetic factors.3
genetic factors.2
genetic factors.5
```

With regard to the point that the system accuracy is better by using RandomForest method, according to the gather selection of the risk factors of the asthma disease in the Iranian society has been identified in accordance with this

Table 4: weighting with p-value

Features	weights with p-value
cough	3.214973e+06
Frequency of cough	3.079321e+07
Just nocturnal cough without any pattern	2.158449e+01
Just daily cough without any pattern	1.268884e+03
Phlegm.1	3.609570e+00
phlegm.2	5.888170e+01
phlegm.3	1.106078e+00
phlegm.4	1.603370e+00
phlegm.5	1.131160e+00
seasonal cough	8.849174e+00
Wheeze	3.687401e+10
Frequency of cough	2.453671e+09
Type of wheeze	4.268934e+08
Dyspnea	7.365768e+01
Chest tightness	4.702714e+04
exercise.1	1.434082e+00
exercise.2	5.396616e+02
allergic rhinitis time	5.503498e+01
allergic rhinitis severity	2.964147e+01
cold.1	1.980710e+01
cold.2	6.475056e+00
cold.3	3.773918e+01
Sinusitis	2.014447e+00
GER	2.061917e+00
Genetic factors.1	1.701684e+00
Genetic factors.2	1.672750e+00
Genetic factors.3	2.158449e+01
Genetic factors.4	2.461234e+02
Genetic factors.5	1.313355e+00
Genetic factors.6	2.461234e+02
Symptoms heperresponsivity.1	2.898393e+04
Symptoms heperresponsivity.2	2.109215e+07
Symptoms heperresponsivity.3	2.307461e+10
Symptoms heperresponsivity.4	6.695252e+05
BMI	4.713894e+00
eczema	5.929039e+00
social factores.1	1.006594e+00
social factores.2	1.672750e+00
social factores.3	1.086038e+00
social factores.4	2.244241e+00
social factores.5	3.909340e+00

Filling the lost amounts by using the closest neighborhood method, With K=10 for 209 series patient sample (as shown in table 5):

Table 5: Filling the missed data

1	2	3	0	1	0	1	0	0 1	0 0	0	0	0	0	01	0	0	0	0	1	0	0 0	1	0 1																					
209	4	3	0	0	0	1	0	0	0	0	1	2	3	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	NA	1	0	1	0	0	1	0	

For weighing by applying RandomForest method the amounts of all data must be completed so that by using KnnImpt the lost amount should be filled out.

Table 6: The data balancing results

	Before balancing	After balancing
Number of class one (asthma)	209	287
Number of class o (non asthma)	115	230

Discussion and Conclusion

This system in the first step diagnoses the adult asthmatic patients and identifies the risk factors in the Iranian society. The main target of using the case-based reasoning system is to assist the physician for the purpose of diagnosing the sort of disease and the therapy more truly, and the next step is creating a system for the sake of helping the public physicians and the medical students in the adults asthma diagnosis and the asthma diagnosis system development to assist in warning the potential asthmatic patients for the personal hygiene and health and also the disease prevention. The application of this system is in the health ministry medical learning and therapy and it is also used for physicians and the medical students. Through this system analysis in diagnosing the adults asthma disease the amount of 80% accuracy, 65% sensitivity and 84% specificity were obvious. For the sake of elevating the system accuracy through the data gathering step the physicians has been talked to and the patient's questionnaire which included the relevant variables have been filled out and for the sake of data correctness. The patient's files have been referenced. The data is without the laboratory information. The results of research showed that the most important variables of asthma disease in Iran country are symptoms heperresponsivity, frequency of cough, cough. The selected variables by the system have conformity with the physician's point of view. This system is also applicable for diagnosing the other diseases and depending upon the kind of the endemic variables in every country one can add new variables and delete the old ones.

Future Studies

Future studies and researches CBR can be with the fuzzy variables and can also concentrate on non adults asthmatic patients.

References

- Aamodt, A. and Plaza E., Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches., *Artificial Intelligence Communications*, Vol.7, No.1, pp. 39-59, 1994.
- Aftarczuk, K." Evaluation of selected data mining algorithms implemented in Medical Decision Support Systems", (2007).
- Aka, DW., Kilber, D., Albert, Mk., Instance-Based Learning Algorithms and Machine Learnings, 37-66, 1991.
- Aljumah, A.A., Ahamad, M. G., Siddiqui, M.K, "Predictive Analysis on Hypertension Treatment Using Data Mining Approach in Saudi Arabia, Intelligent Information Management, 3, pp. 252-261, (2011).
- Asenjo, A." Knowledge representation for diagnosis of care problems through an expert system: Model of the auto-care deficit situations, Expert Systems with Applications 34" pp.2847–2857, (2008).
- Association Rules", European Journal of Scientific Research, ISSN 1450-216X, Vol.61, No.1, pp. 49-59, (2011).
- Bagherjeiran, Abraham, et al., Distance Function Learning for Supervised Similarity Assessment, Department of Computer Science University of Houston, 2006.
- Bibi, M., Nutman, A.,et.al. Prediction of Emergency Department Visits for Respiratory Symptoms Using an Artificial Neural Network. *Chest Journal* ,Vol. 122 ,No.5,pp. 1627-132, 2001.
- Bousquet, J., Vignola, A. M., and Demoly, P., Links Between Rhinitis and Asthma. *Allergy 58*, pp. 691–706, 2003.
- Callahan, K. A., Matsui, E. C., Curtin-Brosnan, J., Eggleston, P A., Diette G. B., Infantile Eczema as a Predictor of Asthma and Asthma Severity in Preschoo Children. *Journal Allergy*. Clin. Immunol.117(2).S78, 2012.
- Canlas Jr R.D, "Data Mining in Healthcare:Current Applications and Issues ", (2009).
- Chae, Y. M., Ho, S. H.1996. Comparison of Alternative Knowledge Model for Diagnosis of asthma. *Expert System with Application*, Vol. 11, No. 4, pp. 423–429.
- Chi-Ming Chu, Wu-Chien Chien, Ching-Huang Lai, Hans-Bernd Bludau, Huei-Jane Tschai, LuPai, Shih-Ming Hsieh, Nian-Fong Chu, Angus Klar, Reinhold Haux, Thomas Wetter, " A Bayesian Expert System for Clinical Detecting Coronary Artery Disease", J Med Sci 2009; 29(4), pp. 187-194, (2009).
- Choi, B. W., Yoo, K. H.,et.al. Easy Diagnosis of Asthma: Computer-Assisted, Symptom-Based Diagnosis. *Journal of Korean Medical Society*, Vol. 22, No. 5, pp. 832–838, 2007.
- Commission, T. J, national patient safety goals, 2008.
- Dakheel, F.I., Smko, R., Negrat, K., Almarimi, A." Using Data Mining Techniques for Finding Cardiac Outlier Patients", (2011).
- Dangare, C. S., Apte, S.S. " Improved Study of Heart Disease Prediction System Using Data
- Mining Classification Techniques", (2012).
- Delen, D." Analysis of cancer data: a data mining approach", (2009).
- Dincer, E., Duru, N." Prototype of a tool for analysing laryngeal cancer operations Acute Coronary Syndrome Prediction Using Data Mining Techniques- An Application, World Academy of Science", Engineering and Technology 59, pp.474-478, (2009).
- Duan, L. Street, W. N., & Xu, E. " Healthcare information systems: data mining methods in
- the creation of a clinical recommender system, Enterprise Information Systems", 5:2, pp.169-181, (2011).

- Ernst, P., Ghezzo, H., and Becklake, M. R., Risk Factors for Bronchial. Hyper responsiveness in Late Childhood and Early adolescence. *European Respir Journal 20*. pp. 635–639, 2002.
- Floyd, S.” Data Mining Techniques for Prognosis in Pancreatic Cancer”, (2007).
- Gao, J., Denzinger, J., and James, R.C.” A Cooperative Multi-agent Data Mining Model and Its Application to Medical Data on Diabetes”, (2005).
- Gunnbjörnsdóttir, M. I., Omenaas, E., Gíslason, T., et al., Obesity and Nocturnal Gastro-Oesophageal Reflux are Related to Onset of Asthma and Respiratory Symptoms. *Eur open Respir Journal 24*, pp.116–121, 2004.
- Gupta, S., Kumar, D., Sharma, A, ”Data Mining Classification Techniques Applied For Breast Cancer Diagnosis and Prognosis”, (2011).
- Ha, S.H., Joo, S.H.” A Hybrid Data Mining Method for the Medical Classification of Chest Pain”, International Journal of Computer and Information Engineering 4:1, pp. 33-38, (2010).
- Habrard, A., Bernard, M., Jacquenet, F.” Multi-Relational Data Mining in Medical Databases”, Springer-Verlag, LNAI 278, (2003).
- Hosseinkhah, F., Ashktorab, H., Veen, R., Owrang O, M. M.” Challenges in Data Mining on Medical Databases IGI Global”, pp. 502-511, (2009).
- Huang, M.-J., Chen, M.-Y., Lee, S.-C.” Integrating data mining with case-based reasoning for chronicdiseases prognosis and diagnosis, Expert Systems with Applications 32”, pp.856–867, (2007).
- Jimenez, M. L., Santamarı, J. M., Barchino, R., Laita, L., Laita, L.M., Gonza´lez, L. A., Jones, M. D., and Brown, A. E., Trzeciak, S., Shapiro, N. I., Garrett, J. S., Heffner, A. C., and Kline, J. A, “The effect of a quantitative resuscitation strategy on mortality in patients with sepsis: a meta-analysis,” *Crit. Care Med.*, 2008.
- Karaolis, M., Moutiris, J.A., Papaconstantinou, L., Pattichis, C.S,” Association Rule Analysis for the Assessment of the Risk of Coronary Heart Events”, (2009).
- Kavitha, K.S., Ramakrishnan, K.V., Singh, M. K,” Modeling and design of evolutionary neural network for heart disease detection”, IJCSI International Journal of Computer Science
- Issues, Vol. 7, Issue 5, September 2010, ISSN (Online): 1694-0814, pp. 272-283, (2010).
- Kika, A. Cico, B. Alimehmeti, R.” Using Machine Learning for Preoperative Peripheral Nerve Surgical Prediction”, (2010).
- Kumar, D.S., Sathyadevi, G., Sivanesh, S.” Decision Support System for Medical Diagnosis Using Data Mining”, (2011).
- Kumari, M., Godara, S, “Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction, IJCST ISSN: 2229- 4333 Vol. 2, Iss ue 2, June 2011.
- Kusiak, A.” Decomposition in Data Mining A Medical Case Study , B.V. Dasarathy (Ed.), Proceedings of the SPIE Conference on Data Mining and Knowledge Discovery: Theory, Tools, and Technology III”, Vol. 4384, SPIE, Orlando, FL, pp. 267-277, April 2001.
- Lee, CH., Chen, JC,et.al. A Novel Data Mining Mechanism Considering Bio-Signal and Environmental Data with Applications on Asthma Monitoring, Comput Methods Programs Biomed, 2010.
- Marir, F and Watson, I., case based reasoning: A Review, Cambridge University Press, *The Knowledge Engineering Review*, Vol. 9, No.4 : pp. 355-381,1994.

- McCormick, T.H. Rudin, C., Madigan, D.” A Hierarchical Model For Association Rule Mining Of Sequential Events: An Approach To Automated Medical Symptom Prediction 21. I.Ullah: Data Mining Algorithms And Medical Sciences”, (2012).

- Misra, B.B., and Dehuri, S., “Functional Link Artificial Neural Network for Classification Task in Data Mining”, Journal of Computer Science 3 (12): 948-955, 2007.

- Nordquist, J., Sundberg, K.,et.al. Case-Based Learning in Surgery: Lessons Learned.*World Journal Surgery*, 2012.
- Osofisan, A.O., Adeyemo, O.O., Sawyerr, B.A., Eweje, O.” Prediction of Kidney Failure Using Artificial Neural Networks”, (2011).

- Palaniappan, S., Awang, R.” Intelligent Heart Disease Prediction System Using Data Mining Techniques “, (2008).

- Parvathi, R. Palaniammalì, S.” An Improved Medical Diagnosing Technique Using Spatial” pp. 169-181, (2011).

- Prasad, B.D.C.N., Prasad, P.E.S.N. K., Sagar, Y,” A Comparative Study of Machine Learning Algorithms as Expert Systems in Medical Diagnosis (Asthma)”, (2011).

- Riedler, J., Braun-Fahrländer, C., Eder, W., Schreuer, M., WaserM., Maisch, S., Carr, D., Schierl, R., Nowak, D., Mutius, V. E, and The AlEX Study Team, Exposure to farming in early life and development of asthma and allergy: a cross-sectional survey. *Journal Lattice 358*, pp.1129–1133, 2001.
- Sakthimurugan, T., Poonkuzhali, S.” An Effective Retrieval of Medical Records using Data Mining Techniques, International Journal Of Pharmaceutical Science And Health Care”, ISSN: 2249-5738. 2(2), pp. 72-78, (2012).

- Satyanandam, N., Satyanarayana, Dr. Ch., Riyazuddin, Md., Shaik, A,” Data Mining Machine Learning Approaches and Medical Diagnose Systems: A Survey”, Volume2, Issue3, 2012.

- Savic, E., Potic, J., Babovic, Z., Rakocevic, G., Strineka, V., Dobrota, M., Milutinovic, V.” Sensor Nets and Data Mining in Medical Applications”, (2011).

- Shapiro, G. G., Among Young Children Who Wheeze, Which Children will have Persistent Asthma. Journal Allergy Clin Immunol 118(3), pp. 562–564, 2006.

- Shantakumar, B.Patil., Kumaraswamy, Y.S, “Intelligent and EffectiveHeart Attack Prediction System Using Data Mining and Artificial Neural Network, European Journal of Scientific Research” ISSN 1450-216X Vol.31 No.4 (2009), pp.642-656 EuroJournals Publishing, Inc. 2009.

- Shin, B. and Cole, SL., A New Symptom-Based Questionnarire for Pediatric the Presence of Asthma, *Journal Investing AllergolImmunol*, ,Vol.20(1), pp. 27-34, 2010.
- Shukla, A., Tiwari, R., Kaur, P,” Knowledge Based Approach for Diagnosis of Breast Cancer”, IEEE International Advance Computing Conference, (IACC 2009).

- Soni, J., Ansari, U., Sharma, D., Soni, S,“Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction”, (2011).

- Srinivas, K., Kavihta Rani, B., Govrdhan, Dr. A, ” Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks”, (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 02, pp 250-255, (2010).

- Tan, P., Steinbach, M., Kumar, V., Introduction to Data Mining, Pearson Addison Wesley, 2006.

- Wasan, S.K., Bhatnagar, V., Kaur, H.” The Impact Of Data Mining Techniques On Medical Diagnostics” , Data Science Journal, Volume 5, pp. 119-126, (2006).

- Yang, C., Street, W. N., Der-Fa Lu, Lanning, L, ”A Data Mining Approach to MPGN Type IIRenal Survival Analysis”, (2010).

- Zolnoori, M., Fazelzarandi, M., Moin, M., Heydarnejad, H. Computer-Aided Intelligent System for Diagnosing Pediatric Asthma, *Journal Medicine System*, 2010.
- Zolnoori, M., Zarandi, F.,et.al. Computer Aided Intelligence System for Diagnosing Pediatric Asthma. *journal of medical systems,* 2010.