Medical Prognosis Generation from General Blood Test Results Using Knowledge-Based and Machine-Learning-Based Approaches

YouJin Kim, Jonghwan Hyeon, Kyo-Joong Oh, and Ho-Jin Choi

School of Computing, KAIST
291 Daehak-ro, Yuseong-gu, Daejeon, Korea
{117kyjin,hyeon0145,aomaru,hojinc}@kaist.ac.kr

Abstract. In this paper, we present two approaches to generate prognosis from general blood test results. The first approach is a knowledge-based approach using ripple-down rules (RDR). The knowledge-based approach with RDR converts knowledge of pathologists into a knowledge base with the minimum intervention of knowledge engineers. The second approach is a machine-learning(ML)-based approach using decision tree, random forest and deep neural network (DNN). The ML-based approach learns patterns of attributes from various cases of general blood test. Our experimental results show that there are indeed some important patterns of the attributes in general blood test results, and they are adequately encoded by the both approaches.

Keywords: clinical decision support system (CDSS), ripple-down rules (RDR), machine learning (ML), prognosis, general blood test

1 Introduction

In modern society, people suffer from stress, busy schedule, unhealthy diet and lack of exercise. In this situation, people are at risk of getting disease such as hypertension, hyperlipidemia, diabetes and obesity. There are no noticeable initial symptom in these illness so patients cannot recognize their bad health condition and do nothing for their own body. As time goes by, illness become worse leading complications and after-effects. Regular check and early detection increase the possibility of full recovery. Therefore, the medical examination which is typical form of preventive medicine has become more important. General blood test is one of the important test involved in medical examination. Through this, we can find various disease checking the number and the shape of cell in blood, hormone abnormality and metabolite value etc.

In this paper, we examine two approaches namely, knowledge-based and machine learning(ML)-based to generate prognoses from general blood test results. The knowledge-based approach is based on the ripple-down rules (RDR) and Induct RDR. The RDR is helpful to construct knowledge bases with the minimum intervention of knowledge engineers. The ML-based approach is based on

decision tree, random forest and deep neural network (DNN). Machine learning methods can learn patterns of attributes automatically from various cases of general blood test. Clinical decision support system (CDSS) based on these approaches is an efficient way to deal with increasing general blood test demand, reducing the pathologist's burden and human error caused by repeating tasks.

There are two contributions in this paper. First, we construct a knowledge base for prognoses generation from real cases of the general blood test which has 685 types of subtests. Second, we train several machine learning models which can classify the prognoses from over 10,000 pathological multi-labeled cases.

The remainder of this paper is organized as follow. Section 2 explores the background for this research and Section 3 explains our methods. In Section 4, we illustrate experiments and in Section 5, Results and analysis are discussed. We conclude our study in Section 6.

2 Background

2.1 Clinical Decision Support System (CDSS)

Decision support system (DSS) involves various computer systems which help decision making process. It edits and reorganizes information by interacting with users and based on this, users could make reasonable decision [1], [2].

Clinical decision support system (CDSS) refers to DSS in health care domain. It gets knowledge from expertise's experience and converts raw medical data to useful information. It helps physician in clinical decision process, reduces medical mistake and improves patient's safety. It is also useful in complex medical decision making process which has ambiguity and contradiction as well as general medical decision making process. It provides reasonable solution and strengthens the rationale of the solution [1], [3], [4], [5], [6], [7]. CDSS is categorized by knowledge-based, ML-based and combination of knowledge-based and ML-based group [8] as shown in Fig. 1.

Knowledge-based CDSS are based on the logic and if-then statements, consisting of knowledge base, inference engine and communication mechanism. The knowledge base contains lots of clinical, medical information represented by a set of rules. The inference engine relate knowledge base rules to input data (new patient data) with if-then statements. Rule-based expert systems and fuzzy logic techniques are types of knowledge-based CDSS [2], [7], [8], [9].

ML-based CDSS is a non-knowledge-based CDSS. It learns from clinical experiences finding patterns of attributes in a large amount of medical data [2], [7]. Artificial neural networks, genetic algorithms and decision trees are involved in ML-based CDSS [8].

2.2 Knowledge-Based Approach

Ripple-Down Rules (RDR). Ripple-down rules (RDR) is a system which defines how to represent, infer and acquire rules. In this system, rules are represented as an n-ary tree, where a node corresponds to a rule which contains a set of conditions and a conclusion. The node can have children which are also rules.

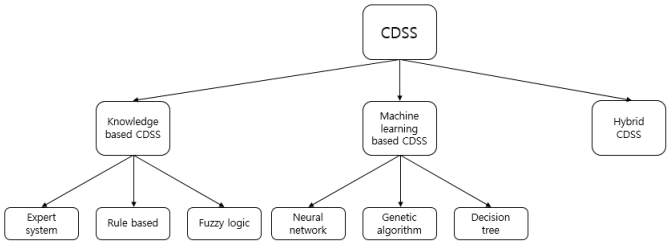

Fig. 1. Classification of CDSS. CDSS is categorized into knowledge-based, ML-based and hybrid system [8].

The inference procedure of RDR is as follow:

1. Evaluate every rule in top level of rule tree.
2. Evaluate rules which is true in previous level.
3. Repeat procedure 2 until there are no more child rules or none of rules is evaluated as true.

Through this procedure, RDR obtains fired rule paths. When the process finished, a conclusion set of the inference consists of the conclusion of the last fired rule in each fired rule path.

RDR provides a systematic way to acquire knowledge so expert can easily modify the knowledge base. This is a crucial difference compared to other methods and makes RDR suitable for CDSS [10], [11].

Induct RDR. A Induct RDR is an algorithm which produces a RDR knowledge base from a large dataset.

Domain experts easily access RDR form knowledge base. However, It requires lots of effort for domain expert to construct a knowledge base from A to Z. In this situation, Induct RDR can help domain expert to build an initial knowledge base. It extracts rules in the form of RDR applying statistical methods to data. After applying Induct RDR, domain experts modify and refine some errors. [10], [11], [12].

2.3 ML Approach

Decision Tree. A decision tree illustrates decision making process as tree shape form. Logical thinking and comparison of various decision paths could be possible with decision tree.

Random Forest. A random forest consists of several decision trees. When it predicts new data, it merges several decision tree's prediction result. Generally, random forest shows good performance avoiding over fitting problem.

Deep Neural Network. A deep neural network imitates a process of human thinking and learns from examples. It consists of an input layer, hidden layers and an output layer and each layer has several neurons. Neurons are connected to neurons in the next layer by weighted connection. In this architecture, neural net analyses data, finds useful meaning and solves various problem [13].

2.4 Multi-Label Classification

Single-label classification problem takes an instance and produces only one output label as shown in Fig. 2 (a). In other words, the problem is to select a label $l \in L$ for each instance $d \in D$. In contrast, multi-label classification problem takes an instance and produces several output labels as shown in Fig. 2 (b). In other words, let D be set of instances and L be set of labels. Then, the problem is to select a set of labels $S \subseteq L$ for each instance $d \in D$.

Pathological diagnosis is a kind of multi-label classification problem because one pathological test sample should produce multiple prognoses as output labels.

(a) Single-label classification (b) Multi-label classification

Fig. 2. Single-label vs. multi-label classification

3 Methods

3.1 Knowledge-Based Approach

We use Induct RDR in constructing an knowledge base reducing expert's burden. We preprocess general blood test results rather than use it directly. First, we quantize numerical test results as low, normal, and high to reduce the complexity of statistical inferences. Also, we merge test codes which indicate the same test (e.g., test code 00011, 00530 indicate glucose test) into single test code.

Because Induct RDR handles only single-label classification problem, we cannot directly apply Induct RDR into the general blood test results which are multi-label classification problem. Therefore, we divide the data into 18 categories that is Anemia, Liver, Blood, Pancreas, Blood sugar, Rheumarthritis, Blood type, Stool, Electrolyte, Syphilis, Hepatitis virus, Thyroid, Infection, Tumor, Kidney, Urine, Lipid and Etc.

However, some categories are still multi-label classification. For example in case of the electrolyte category, a patient may have several abnormalities such

as Na, K, Ca, etc. To handle this, we duplicate test value and split one data instance to several instances which have same test value but different prognosis as shown in Fig. 3.

After preprocessing the data as mentioned above, we apply Induct RDR for each category independently and then, we merge 18 knowledge bases into a single knowledge base.

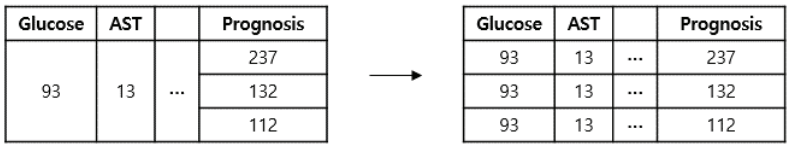

Fig. 3. Transformation from multi-label to single-label classification. We duplicate test value for each prognosis to make one data instance have one prognosis.

3.2 ML-Based Approach

In case of ML-based approach, we preprocess data in other way. For this approach, test value needs to be transformed to vector form. We categorize test value by five categories and for each category, process data maintaining its original meaning. Table 1 shows conversion result of each class data.

Each patient has several prognoses based on the blood tests they take. Therefore, this could be considered as a multi-label classification, having patient's age and sex, blood test results as input and prognoses as output. It is important to choose appropriate machine learning methods to solve this problem because not all machine learning methods effectively solve multi-label classification. In this paper, we use decision tree, random forest and deep neural network.

Decision tree are constructed through Gini impurity and random forest consists of 200 decision trees. Deep neural network is comprised of four fully connected layers. Each layer has 1,041, 1,024, 512 and 256 neurons respectively and output vector has 232 elements as shown in Fig. 4. Generally, softmax function is used as the last activation function in deep neural network to classify something. Instead however, we use a sigmoid function because this is a multi-label classification that each label needs probability distribution. Sigmoid gives n probability distributions and softmax gives one probability distribution when there are n labels. Therefore, we use sigmoid function as the last activation function. We assume that each element of output vector represents the probability of whether each prognosis is involved in the output or not to solve multi-label classification problem by deep neural network. If the probability is higher than 0.5, we include the prognosis in the output.

Table 1. Conversion table for each test result category in the preprocessing step. *before* means raw result value and *after* indicates preprocessed value. Class *a* consists of nominal value and numeric value. Class *b* contains sign of inequality. Class *c* is comprised of two parts (value, class). Class *d* is nominal type data such as blood type. Class *e* has no value.

Category	**Preprocessing**	
	Before	**After**
a	Non-Reactive 1.0	Non-Reactive
b	<0.5	0.5
c	1.23 1.0	1.23
d	A, B, O, AB	1000, 0100, 0010, 0001
e	“ ”	empty

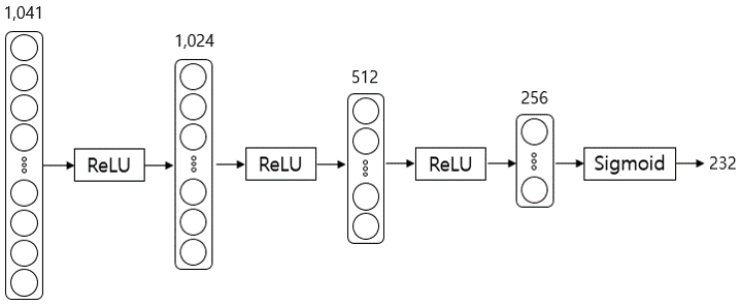

Fig. 4. Deep neural network comprised of four fully connected layers. Each layer has 1,041, 1,024, 512, 256 neurons respectively and output vector has 232 elements. ReLU is used as activation function and sigmoid is adopted for the last activation function.

4 Experiments

4.1 General Blood Test Data

In this paper, we use anonymized dataset provided from Seegene Medical Foundation (http://www.seegene.co.kr) which consists of patient information, blood test results and prognoses.

General blood test has 685 types and consist of several subtest (blood sugar, kidney-gout-arthritis, liver function, electrolyte, lipid-cardiovascular system, hepatitis, venereal disease, iron, blood, blood type, pancreas, inflammation, urine, thyroid gland and tumor marker). Prognoses are made by experts and have 232 types.

Dataset consist of 14,479 data elements and one data element as shown in Fig. 5, consists of *Age, Sex, Test name, Test result* and *Prognosis* for each patient. *Test name* and *Test result* are came from tests each patient take. *Prognosis* is created considering all the test results one patient take.

In 14,479 number of data, we used 8610 data in RDR (training: 7,610, testing: 1,000) and we used all data when training and testing deep neural network.

Age			**Sex**		
47			M		
Test name	**Test result**	**Test name**	**Test result**	**Test name**	**Test result**
Creatinine	0.89	Triglyceride (TG)	77	Hbe Ab	Positive < 0.10
Glucose (FBS)	92	HDL-Cholesterol	60	AFP (α-fetoprotein)	2.89
AST (SGOT)	30	Hemoglobin (Hb)	15.8	HBV DNA정량(Realtime PCR)_01	< 116
ALT (SGPT)	23	HBs Ag(정밀)	Positive 4274.23	HBV DNA정량(Realtime PCR)_02	< 20
Γ-GTP (GGT)	37	HBs Ab (Anti-HBs) 정밀	Negative 1.01	HBV DNA정량(Realtime PCR)_03	< 0.0004
Cholesterol	222	HBeAg	Negative 0.29		
Prognosis					
① 공복 혈당이 정상입니다. 당뇨검사의 결과가 정상입니다. ② 신장기능 검사 결과는 정상입니다. ③ 간 기능 검사의 결과는 정상입니다. ④ 지질검사의 결과 정상 입니다. ⑤ B형간염항원 양성으로 B형간염 보균상태입니다. ⑥ 빈혈이 없습니다. ⑦ 종양표지자검사는 정상입니다. 종양 표지자는 종양 등에서 생성되어 혈액이나 체액으로 분비되는 여러 종류의 물질입니다.					

Fig. 5. Illustration of blood test results and prognoses by human experts. Prognoses are written in Korean. From the top, each comment sentence means that, ① Fasting blood sugar is normal. Result of diabetes mellitus test is normal. ② Result of renal function test is normal. ③ Result of liver function test is normal. ④ Result of lipid test is normal. ⑤ Hepatitis B antigen positive, hepatitis B carrier state. ⑥ Anemia is none. ⑦ Result of tumor marker test is normal. Tumor marker is created from tumor and is substance secreted to blood or body fluid.

4.2 Experiments for the Knowledge-Based Approach

After preprocessing the data as mentioned before, we apply Induct RDR for each category independently and then, we merge 18 knowledge bases into a single

knowledge base. We use 7,610 general blood test data to create an knowledge base and evaluate the constructed knowledge base with unused 1,000 general blood test data.

4.3 Experiments for the ML-Based Approach

We construct decision tree through Gini impurity and we make random forest using 200 decision trees. Deep neural network are composed of four layers and we trained the neural network 100 times. We compare performance of decision tree, random forest and deep neural network and all evaluation was done using 5-fold cross validation.

5 Results and Analysis

5.1 Results for the Knowledge-Based Approach

We use an additional rate and a missing rate to evaluate the RDR-based expert system. The additional rate represents how many prognoses are over-generated compared to original prognoses. The missing rate represents how many prognoses are under-generated compared to original prognoses.

In conclusion, as shown Table 2, we can get 96.01% as the additional rate and 46.91% as the missing rate. We group the general blood test data into 18 categories so, 18 independent knowledge bases are generated. Therefore, the expert system should generate at minimum, 18 prognoses. We think this is the reason of the high additional rate.

Table 2. Results of the knowledge-based approach

Missing rate	**Additional rate**
46.91%	96.01%

5.2 Results for the ML-Based Approach

We use accuracy, precision, recall and F1-measure to evaluate the ML-based system. However, we cannot use general accuracy, precision, recall and F1-measure because multi-label classification problem partially correct concept [14]. Therefore, we use Godboles definition [15] to properly evaluate.

Let T, h and Z_i be defined as

T: dataset consisting of n data elements, (x_i, Y_i), $1 \leq i \leq n$
h: multi-label classifier

Z_i: classifier h's classification result, $h(\mathrm{x}_i)$.

Then, accuracy, precision, recall and f1-measure are defined as

$$Accuracy = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \bigcap Z_i|}{|Y_i \bigcup Z_i|} \tag{1}$$

$$Precision = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \bigcap Z_i|}{|Z_i|} \tag{2}$$

$$Recall = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \bigcap Z_i|}{|Y_i|} \tag{3}$$

$$F1 - measure = \frac{1}{n}\sum_{i=1}^{n}\frac{2|Y_i \bigcap Z_i|}{|Y_i| + |Z_i|} \tag{4}$$

Furthermore, we use hamming-loss to take into account the prediction error (mistakenly predicting wrong label) and the missing error (missing out correct label).

Let I and k be defined as below

I: indicator function
k: the number of labels dataset T has.

Then, hamming-loss is defined as

$$Hamming - loss = \frac{1}{kn}\sum_{i=1}^{n}\sum_{l=1}^{k}\left[I\left(l \in Z_i \wedge l \notin Y_i\right) + I\left(l \notin Z_i \wedge l \in Y_i\right)\right] \tag{5}$$

We compare performance of decision tree, random forest and deep neural network. All evaluation was done using 5-fold cross validation and evaluation results can be checked in Fig. 6 and Fig. 7. We can find that deep neural network has better performance than decision tree and random forest in every metric. We think that if we have more data, then deep neural network would show the better performance.

In Fig. 8, we show prognosis by human experts and prognosis generated by deep neural network. The 5th, 8th and 15th prognosis from the top are in *real prognosis* but are not in *predicted prognosis*. This shows missing error. On the other hand, the last prognosis from the top is contained in*predicted prognosis* but not in *real prognosis* indicating prediction error. There are some errors but we can check deep neural net quite well predict.

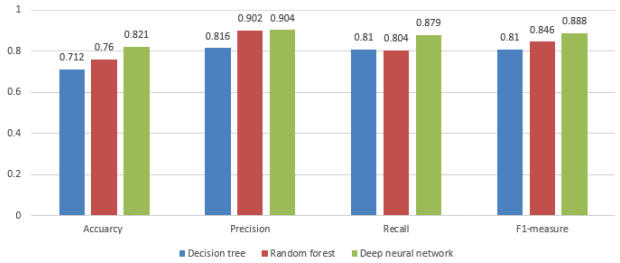

Fig. 6. Results of the ML-based methods. *Deep neural network* shows better performance than *decision tree* and *random forest* in every metric.

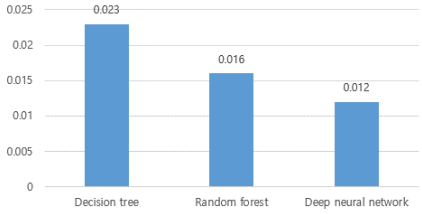

Fig. 7. Hamming-loss of the ML-based methods. *Deep neural network* has the lowest hamming-loss which means it is better than others.

6 Conclusion

In this paper, we investigated two approaches for generating preliminary medical prognosis from patients' general blood test results.

The first approach was the RDR (and the induct RDR) framework to form a knowledge base for medical prognosis. By applying the RDR framework to CDSS, we can expect to build an environment where human experts (doctors) can freely modify the knowledge base with the reduced help of knowledge engineers when the need arises.

As the second approach, we utilized various machine learning algorithms including deep neural network to classify relevant prognoses using the pathological cases. Our experimental results showed that there are indeed some important patterns of the attributes in the blood test results, and they are adequately learned by this approach.

Acknowledgments. This work was supported by the Industrial Strategic Technology Development Program, 10052955, Experiential Knowledge Platform Development Research for the Acquisition and Utilization of Field Expert Knowledge, funded by the Ministry of Trade, Industry & Energy (MOTIE), Korea.

Actual prognosis	Predicted prognosis
B형간염항체 양성으로 B형간염 면역상태입니다.	B형간염항체 양성으로 B형간염 면역상태입니다.
C형간염항체 음성입니다. C형간염이 없습니다.	C형간염항체 음성입니다. C형간염이 없습니다.
간 기능 검사의 결과는 정상입니다.	간 기능 검사의 결과는 정상입니다.
갑상선기능 검사 결과 정상입니다.	갑상선기능 검사 결과 정상입니다.
공복 혈당이 정상이고 당화혈색소(HbA1c)도 정상입니다. 당뇨검사의 결과가 정상입니다.	
류마티스인자(RF)가 정상입니다. 류마티스관절염의 30%에서 음성을 보이므로 임상증상과 영상의학적 소견 등을 참고하여야 합니다.	류마티스인자(RF)가 정상입니다. 류마티스관절염의 30%에서 음성을 보이므로 임상증상과 영상의학적 소견 등을 참고하여야 합니다.
빈혈이 없으며 철분도 부족하지 않습니다.	빈혈이 없으며 철분도 부족하지 않습니다.
소변에서 결정체가 발견되었습니다. amp.urate결정체는 정상에서도 발견될 수 있습니다.	
신장기능 검사 결과는 정상입니다.	신장기능 검사 결과는 정상입니다.
요산의 농도가 정상입니다. 요산은 통풍의 원인입니다.	요산의 농도가 정상입니다. 요산은 통풍의 원인입니다.
전해질 관련검사의 결과는 정상입니다.	전해질 관련검사의 결과는 정상입니다.
종양표지자검사는 정상입니다. 종양 표지자는 종양 등에서 생성되어 혈액이나 체액으로 분비되는 여러 종류의 물질입니다.	종양표지자검사는 정상입니다. 종양 표지자는 종양 등에서 생성되어 혈액이나 체액으로 분비되는 여러 종류의 물질입니다.
지질검사의 결과 정상 입니다.	지질검사의 결과 정상 입니다.
혈구질환 관련 검사의 결과 정상입니다.	혈구질환 관련 검사의 결과 정상입니다.
혈액형은 AB형 Rh 양성입니다.	
	소변 백혈구 양성입니다. 요로감염과 관련이 있습니다. 여성의 경우는 질분비물에 의한 오염이 흔합니다.

(a) Prognosis by human experts vs. prognosis predicted by DNN

Actual prognosis	Predicted prognosis
Hepatitis B antibody positive, hepatitis B immune state.	Hepatitis B antibody positive, hepatitis B immune state.
Hepatitis C antibody negative, hepatitis C is none.	Hepatitis C antibody negative, hepatitis C is none.
Result of liver function test is normal.	Result of liver function test is normal.
Result of thyroid gland function test is normal.	Result of thyroid gland function test is normal.
Fasting blood sugar is normal and glycosylated hemoglobins(HbA1c) is normal. Result of diabetes mellitus test is normal.	
Rheumatoid Factor(RF) is normal. 30% of rheumatoid arthritis is negative so clinical manifestation, radiological finding and etc. should be considered.	Rheumatoid Factor(RF) is normal. 30% of rheumatoid arthritis is negative so clinical manifestation, radiological finding and etc. should be considered.
Anemia is none and limatura ferri is not deficiency.	Anemia is none and limatura ferri is not deficiency.
Crystal is found in urine. amp.urate crystal can be found in normal.	
Result of renal function test is normal.	Result of renal function test is normal.
Concentration of uric acid is normal. Uric acid is cause of gout.	Concentration of uric acid is normal. Uric acid is cause of gout.
Result of electrolyte related test is normal.	Result of electrolyte related test is normal.
Result of tumor marker test is normal. Tumor marker is created from tumor and is several substance secreted to blood or body fluid.	Result of tumor marker test is normal. Tumor marker is created from tumor and is several substance secreted to blood or body fluid.
Result of lipid test is normal.	Result of lipid test is normal.
Result of hemocyte disease related test is normal.	Result of hemocyte disease related test is normal.
Blood type is AB Rh positive.	
	Urine leukocyte positive. This is related to urinary tract infection. In cased of women, contamination caused by vaginal discharge is common.

(b) Translated version of (a)

Fig. 8. Illustration of predicted results for a patient

References

1. Foster, D., McGregor, C., El-Masri, S.: A survey of agent-based intelligent decision support systems to support clinical management and research. In: the 2nd International Workshop on Multi-Agent Systems for Medicine, Computational Biology, and Bioinformatics, pp. 16-34 (2005)
2. Abbasi, M.M., Kashiyarndi, S.: Clinical Decision Support Systems: A discussion on different methodologies used in Health Care. Marlaedalen University, Sweden (2006)
3. Smith, A.E., Nugent, C. D., McClean, S. I.: Implementation of intelligent decision support systems in health care. Journal of management in medicine, 16(2), 206-218 (2002)
4. Jao, C.S., Hier, D. B.: Clinical decision support systems: An effective pathway to reduce medical errors and improve patient safety. INTECH Open Access Publisher

(2010)
5. Rokach, L., Maimon, O.: Data mining with decision trees: theory and applications. World scientific (2014)
6. Harold C.S, Michael C. Higgins, Douglas K. Owens: Medical Decision Making, 2nd Edition, John Wiley and Sons, Ltd. (2013)
7. Allen J. Moses, Marcus Lieberman, Irving Kittay, Jorge A. Learreta: Computer-Aided Diagnoses of Chronic Head Pain: Explanation, Study Data, Implications, and Challenges. The Journal of Craniomandibular & Sleep Practice, 24(1), 60-66, Taylor & Francis (2006)
8. Aljaaf, A.J., Al-Jumeily, D., Hussain, A. J., Fergus, P., Al-Jumaily, M., Abdel-Aziz, K.: Toward an optimal use of artificial intelligence techniques within a clinical decision support system. In: Science and Information Conference(SAI), pp. 548-554. IEEE (2015)
9. Ye, Y., Diao, X., Jiang, Z., & Du, G.: A Knowledge-Based Variance Management System for Supporting the Implementation of Clinical Pathways. In: Management and Service Science, 2009. MASS'09. International Conference on pp. 1-4, IEEE (2009)
10. Compton, P., Edwards, G., Kang, B., Lazarus, L., Malor, R., Menzies, T., Sammut, C.: Ripple down rules: possibilities and limitations. In: the 6th AAAI Knowledge Acquisition for Knowledge-Based Systems Workshop, Calgary, Canada, University of Calgary, pp. 6-1, AAAI (1991)
11. Kang, B., Compton, P., Preston, P.: Multiple classification ripple down rules: evaluation and possibilities. In: the 9th knowledge acquisition for knowledge based systems workshop (1995)
12. Gaines, B.R., Compton, P.: Induction of ripple-down rules applied to modeling large databases. Journal of Intelligent Information Systems, 5(3), 211-228, Springer (1995)
13. Berner, E.S.: Clinical decision support systems. New York: Springer Science + Business Media, pp. 3-22, Springer (2007)
14. Sorower, M.S.: A literature survey on algorithms for multi-label learning. Oregon State University, Corvallis (2010)
15. Godbole, S., Sarawagi, S.: Discriminative methods for multi-labeled classification. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 22-30, Springer, Berlin Heidelberg (2004)