

Estimation Medicine for Diseases System to Support Medical Diagnosis by Expert System

Noor T. Mahmood¹

Computer Science Department
Al- Mustansiriyah University
Baghdad, Iraq

Abstract—Researches confirmed that 70 thousand cases of death, which happen yearly in the world, were because of the misprescribing of the drug itself or its dose (overdose or lower dose). Choosing the wrong alternative drug inspired the professionals in the healthcare field to the importance of assigning the best technologies to decrease the percentages of the therapeutic methods in giving the drug to prevent mistakes in prescribing the suitable drug. A system based on Rete Algorithm is proposed where the best-chosen medicine is offered through the suggested system. Selection of Estimation Medicine for Diseases (EMD) System is introduced where the diagnosis is made basically according to the symptoms and the medical history of the patient. This research aims to acquire a good model using this algorithm to obtain more accurate choices of medicine. The system (EMD) is tested by the doctors in Iraqi hospitals and it has been found that there is no other systems that can be compared to EMD system. The accuracy of estimating the appropriate medicine for heart diseases is approximately (87.26%).

Keywords—*Diagnosis; Disease; Medicine; Rete Algorithm; Expert System; Intelligent System*

I. INTRODUCTION

Many of the patients have some troubles when they visit the doctor for identifying the suitable treatment for them. After the diagnosis of the disease have been made by the doctor, another problem is appeared, the drug might be written in a nonspecific way. Since the choice of the drug depends on the disease history of the patient and whether he had any chronic diseases, consequently careful assessment should be done as the drug is given.

The Estimation Medicine for Diseases (EMD) is an intelligent system (Expert System) to identify the suitable treatment to the patient by knowing the disease or the case and the disease history of the patient and the symptoms that the patient is suffering. This process may prevent the mistakes that effects the wrong prescription of the drug which might not cause a significant health damage to many thousands of the patients only but also might lead to death to some of them.

What's done in the proposed approach is a process to enter the information into the system which identifies the correct diagnosis and obtains the accurate diagnosis of patient's diseases as quickly as possible with less cost. In our system selection, the best medicine is proposed for the patient and the doctor. It helps in determining medicine by disease or conditions(symptoms) according to some concepts, rules, and algorithms which are expressed in the next sections

As a comparative view between previous researches and that one; previous researches usually choose two or three side effects and then contrast them to reach drugs with less interference and are less risky for the patient. On the other hand, this research is tested by an expertise doctor in diagnosing the disease depending on the rules of choosing the right drug for the intended patient. Also to realize the side effects of each drug or medication in addition to the previous symptoms taken from the patient prohibits causing pharmacological interference. The system picks the selected medicine that takes less error ratio and higher priority of accuracy for the preferring drugs that have fewer side effects and less conflicts with other medicines.

Collecting information about both diseases and medicines along with the side effects of each drug is made by the assistance of expert doctors from Iraqi hospitals and from reliable international resources in [1][2][3][4].

II. LITERATURE SURVEY

The side effects of the prescribed medications are a common occurrence in the medical history of patients. Jenna, M. R., et. al [5] presented the opportunity to identify new side effects efficiently through electronic healthcare databases. A confirmation of concept method is suggested. It learns common associations and uses this knowledge to automatically refine exposure-outcome associations (i.e. side effect signals). A novel measure is determined. It is named the confounding-adjusted risk value, an accurate absolute risk value of a patient. They show that signals filtering might be possible at a patient level according to association rules acquired by taking into account patients' medical histories.

Chen, Jie, et. al [6] developed an association classification algorithm in which rules discovered can be used to alert medical practitioners when prescribing drugs. They propose two kinds of probability trees that can present clearly the risk of specific adverse drug reactions to prescribers.

Chen, Yu, et. al [7] used association rule mining approach to derive possible side effects due to exposure to multiple drugs at different durations of the pregnancy. They derived sequential temporal rules to discover new information that would not be detected by the traditional analysis method that is currently used by pharmacists.

The optimizing drug dose is a major factor in improving the sustainability of health care. Daughton, C. G., and Ruhoy, I. S. [8] presents the first critical examination of the multi-faceted role of drug dose in reducing the ambient levels of active

pharmaceutical ingredients (APIs) in the environment. Personalized adjustment of drug dose holds the potential for enhancing therapeutic outcomes while simultaneously lowering the incidence of adverse drug events and in lowering patient healthcare cost.

III. EXPERT SYSTEM

An expert system is an application that tries to react like a human expert on a special subject area. This system is used to advise non-experts in situations where a human expert is unavailable [9].

An expert system consists of three parts [9]: (see Figure 1)

A user interface - that accepts a non-expert user and it asks questions of the expert system, and is conferred with advice. The interface should be a simple and easy to use [9].

A knowledge base - It is a group of facts and rules. It is built with information that is supplied by human experts [9].

An inference engine - this acts as a search engine, inspecting the knowledge base for a particular arrangement that matches the user's query [9].

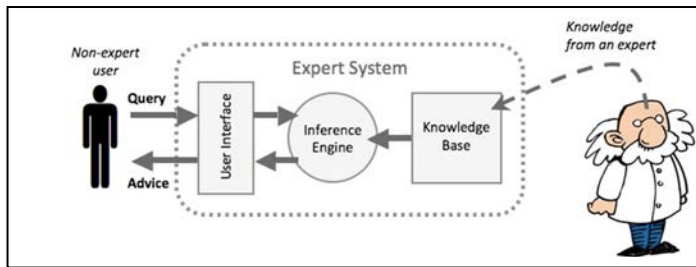


Fig. 1. An Expert system works in block diagram [9]

An example of expert systems is diseased diagnosis (the input would contain medical history information, the symptoms of the patient. Those might be the query, and the answer is a diagnose for the patient's mental or physical pain). However, Nowadays, "conventional" computer programs may achieve some of the typical expert system functionalities. When using, expert systems problems are solved using heuristics or approximate methods which, unlike traditional applications, are not guaranteed to present an optimal solution [9].

IV. GENERAL DESCRIPTION FOR ESTIMATING THE BEST MEDICINE SYSTEM

This section describes the proposed Estimation Medicine for Diseases (EMD) System framework, algorithms, and the applied rules. Figure (2) illustrates a block diagram for selection medicine for each patient's disease.

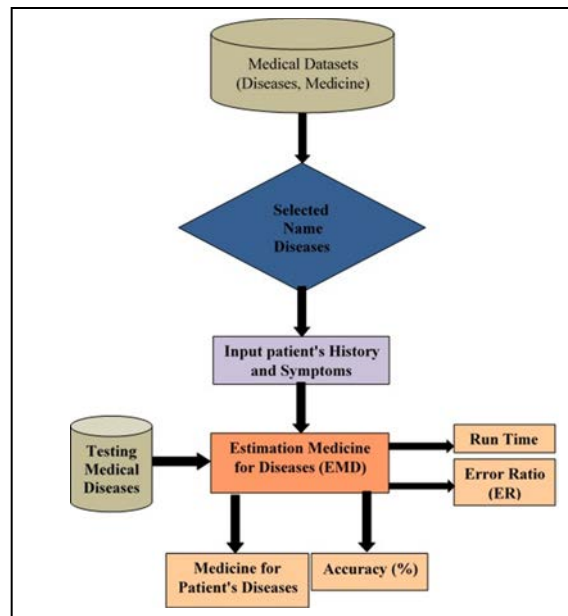


Fig. 2. Block Diagram Estimation Medicine for each Patient's Disease

The method used in the system makes use of dataset for heart diseases along with medicine dataset for medical data. The obtained medical data are utilized in the formation of the expert system. The system utilizes in a convenient way the concepts and algorithms introduced in the following subsections.

A. Reliable rules for safer drug use

This section explains how the doctor selects the best medicine for each patient's disease for conditions by some rules based on safer drug use; these rules are expressed in [10].

Those rules were compiled for more dependable drug use (or nonuse). They were listed particularly from the World Health Organization's General Prescribing Principles for the Elderly. Nevertheless, these rules might be applied to all ages. Doctors and patients involved in drug therapy should know them. The guiding principle is to use those rules as guidelines for patient treatment according to symptoms and patient's medical history. Among those rules are [10]:

- Additional drugs used may make the earlier drugs much more dangerous [10].
- The primary doctor ought to coordinate patient's care and drug use [10].
- The patient should make sure that drug therapy is needed [10].

- Important, believable, and adverse effects of the drug in addition to plausible drugs interactions should be foreseeable in advance [10].
- Any adverse drug reactions patient may have should be taken into consideration early [10].

B. Recognize-Act Cycle

Since the production system is a formulation, that the expert system’s inference engine can process efficiently, they can be used to generate new information. The basic recognize-act cycle consists of four steps [11]:

- 1) Recognize (get) matches.
- 2) Choose a match.
- 3) Act (perform the selected production).
- 4) Return to step (1).

It matches the premise patterns of the rules against elements in the working memory. If there is more than one rule that can be applied, use a conflict resolution strategy to choose one to apply. Early production systems spent over 90% of their time doing pattern matching, but there is now a solution to this efficiency problem which is the Rete algorithm [11].

C. Rete Algorithm

The traditional approach for recognize-act cycle matches all Elements (E) in working memory along with all premises (P) in all rules (R). It then checks E×P×R possibilities in each cycle. Nevertheless, the rules have structural similarity, since parts of their conditions are common. Where the application of any similar rule changes the working memory slightly (temporal redundancy) [11].

Rete Algorithm uses these facts to improve efficiency (‘rete’ is Latin for ‘net’). First, the initial conflict set is generated through the matching algorithm by connecting the network for all condition parts. After that, only the changed elements in working memory are fed into the network to distinguish the changes in the conflict set [11].

D. Estimation Medicine for Diseases (EMD) algorithm

The above concepts and foresight are consolidated into a single unified algorithm called Estimation Medicine for Diseases (EMD) which performs the main system functionality.

Algorithm: Estimation Medicine for Diseases (EMD)

Input: Diseases Table (D), Medicine Table (M), Disease Name for Doctor Diagnosis (DN).

Output: Medicines Name (MsN), Error Ratio(ER).

Begin:

- Input patient's History and Symptoms (PHS).
- For each (D)That has name (DN)
- Find Diseases Symptoms (DS)
- Find Similar Symptoms between(PHs) and (DS) and put it on new array (SS)
- (If SS is found in M Then MsN()→ Count +1)//Compare (SS) with (M) for (DN)

- Sort MsN()
- Select the minimum Error Ratio
- If the value of MsN is equal, then select the high priority from (M)

End

V. MEDICAL DATASET (DISEASES, MEDICINE)

The data was collected from an Iraqi hospital. Patients' files and information were extracted from the Statistics Division at Al-Kindi Teaching Hospital. The data for each disease consists of medical history, symptoms, laboratory analysis, etc. Those have been entered into the database. Each disease has ten medicines that were fed into the system. (See Table I). In this table, each column represents a single disease with its well-known ten medicines. The following are some of the heart diseases:

- High Blood Pressure (I10).
- Angina pectoris (I20).
- Myocardial Infarction (I21).
- Atrial Fibrillation (I48).
- Heart Failure (I50).

Note that I10, I20, I21, I48, and I50 represent the statistical identifiers of the diseases.

TABLE I. DISEASES AND MEDICINES

ID	I10(HighBlood Pressure)	I20(Angina pectoris)	I21(Myocardial Infarction)	I48(Atrial Fibrillation)	I50(Heart Failure)
1	Diazoxide	Glyceryl_trinitrate	Atenolol	Digoxin	Captopril
2	Hydralazine	Aspirin	Metoprolol	Dronedarone	Cilazapril
3	Sodium Nitroprusside	Clopidogrel	Acebutolol	Warfarin	Enalapril maleate
4	Methyldopa	Prasugrel	Propranolol Hydrochloride	Acenocoumarol	Lisinopril
5	Captopril	Heparin	Labetalol Hydrochloride	Phenindione	Valsartan
6	Lisinopril	Streptokinase	Pindolol	Diltiazem	Telmisartan
7	Moexipril Hydrochloride	Urokinase	Sotalol Hydrochloride	Verapamil	Olmesartan Medoxomil
8	Fosinopril Sodium	Tenecteplase	Timolol maleate	Propranolol Hydrochloride	Losartan potassium
9	Enalapril	Propranolol Hydrochloride	Carvedilol	Acebutolol	Irbesartan
10	Cilazapril	Atenolol	Bisoprolol	Atenolol	Eprosartan

A large amount of data is required to train the system since poor of data does not give high accuracy in diagnosis. Each represented disease has its specific table. Each table consists of some attributes. These attributes are different from one table to another. All data are taken from the patient's file, which was read by doctors, who helped me and taught me how to read and understand each patient's file.

Training datasets consist of two groups: the first is made of five tables where a table is specified for each medical case in

which the patient data are taken from the patient files. These are the symptoms and patient medical history. While the second category contains five tables, where a table is designated for each disease. This table has ten drugs commonly identified with that disease, yet each drug has a set of side effects. The name of the medicine is the scientific name as it is an international standard.

Most of the researches are different from this one because in these researches the acquisition of data is made from the internet and they are almost based on one disease. The results of these researches indicate that the disease may exist or not and they don't give the accurate medicine for the patient and as a result they cannot tell all the symptoms that the patient suffers from and hence they cannot prevent the misprescribing of drugs"

VI. TYPICAL SYSTEM OPERATIONS

One user interface of the system explained in this section is shown in Figure (3). This window is called as one of some sub-windows. Learning in the system depends on the data stored in the system (training dataset) and data tested by the system.

As an example, the system works to diagnose that the patient has heart diseases by using functional (EMS) strategies. The system deals with heart disease, which is diagnosed as the patient's disease. Doctor selects any heart disease, as shown in Figure (3) where the patient is suffering from one of them, but it is unknown for his doctor at this moment.

A sample medical history data is (patient's age: more than fifty years (age), Increase of blood cholesterol (cholesterol), Hypertension (HT), Respiration Rate (RespRate) and Blood Pulse High (BPH)). The symptoms are (Shortness of Breath, Dizziness, Weakness, Rapid heartbeats, Palpitation, Headache and General malaise).

The doctor in the hospital diagnosis that patient is suffering from High Blood Pressure and Myocardial Infarction. When his medical history and symptoms are entered into the system, the result would be selecting the best medicine for diagnosis to High Blood Pressure (I10) and Myocardial Infarction (I21). Hence, the doctor has selected medicine by the system (EMS) depending on all information and side effects of that medicine, symptoms, and history of this patient.



Fig. 3. A heart disease diagnosis and selected medicine by (EMD)

VII. EXPERIMENTAL RESULTS

This section illustrates the experimental results of the proposed approach described in the previous sections.

A. Settings

The experimental results also include Rete Algorithm for Medicine. The results are organized into a table for heart diseases. Brief descriptions and the diagnosis were given by the doctors and also by the proposed system—Estimation Medicine for Diseases (EMD) that offers selecting the best medicine for a given disease.

These approaches are running on a laptop (running Core i5 CPU@ 1.7MHz, RAM 4GHz on Windows 7 Home Basic), SQL Server 2012 for building Database and Visual Basic.NET 2012,2013 (VB.NET) environment as the programming language used in the implementation.

B. Results

The proposed system deals with many types of heart diseases. System experiments on the applied fed data. Each stored table is used to test different disease cases. The system was configured over five categories representing the selected heart diseases. Table I demonstrate this training dataset.

The system has been tested by entering symptoms and medical history of each patient. Data are entered into the built structure, then calculations of the probability of occurrences of the disease for each patient are made. This result is the likelihood of disease for each patient along with the selected medicine. The doctors can choose two of the top up possibilities of these medicines (i.e. the higher priority among them). The performance of the system is expressed in Table II. Where we can see twenty-two patients for heart diseases.

After that accuracy is calculated by (true (cases) / (true (cases) + false (cases))) to get the results for the system model of (EMD). The performance of the EMD can be seen for each disease. Hence, the accuracy rate of the given model of (EMD) is approximately (87.26%) in heart diseases. It can be deduced that the rate of accuracy of the system (EMD) is the best average results. These results are shown in table II while the running time is presented in Table III.

TABLE II. EMD PERFORMANCE FOR THE SELECTED MEDICINE

Disease code	Disease name	Number of patients used in testing	The selected medicine for disease	
			Accuracy (%)	Error Ratio (%)
I10	High Blood Pressure	22	81.8%	18.2%
I20	Angina pectoris	22	86.36%	13.64%
I21	Myocardial Infarction	22	90.9%	9.1%
I48	Atrial Fibrillation	22	90.9%	9.1%
I50	Heart Failure	22	86.36%	13.64%
Average			87.26%	12.74%

TABLE III. TRAINING DATASET RUNNING TIME FOR HEART DISEASES AND MEDICINE DATASET

Model for Heart Diseases	Time (Sec)
SBM	0.30

Training dataset running time for heart diseases and medicine dataset is (0.30) sec because it is only a time to choose the best medicine in the system yet it is not for diagnosing the disease. The system is an assistant to the doctor, and it is a complementary tool for the medical diagnosis.

Figure (4) demonstrates a chart indicating the Performance of (EMD) for Heart Diseases. Training datasets consist of two groups: the first is made of five tables where a table is specified for each medical case in which the patient data are taken from the patient files. These are the symptoms and patient medical history. While the second category contains five tables, where a table is designated for each disease. This table has ten drugs commonly identified with that disease, yet each drug has a set of side effects. The name of the medicine is the scientific name as it is an international standard.

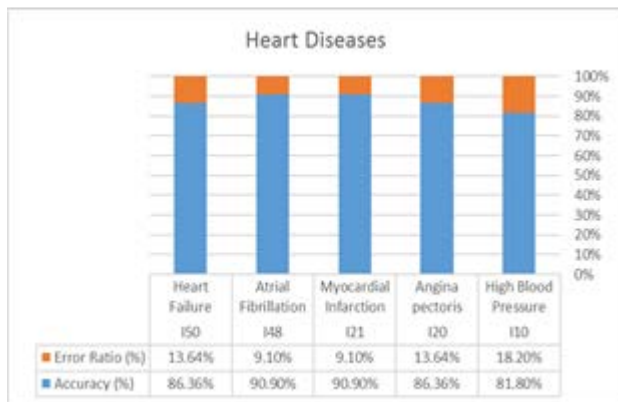


Fig. 4. The performance of (EMD) for Heart Diseases

VIII. CONCLUSIONS

In this paper, an approach based on Rete Algorithm is used where the best-chosen medicine through the system (EMD) is suggested. We want to reach into the best model using the algorithm to obtain a more accurate selection for an optimal medicine for diagnosis. The doctor can select medicine by the system (EMD) depending on all information and side effects of that medicine, symptoms and medical history for a patient. The system (EMD) is tested by the doctors in Iraqi hospitals. The accuracy of selecting the medicine to heart diseases is approximately (87.26%).

This accuracy is obtained because the principles of nominating drugs by a doctor for a given patient are based on two factors. The first factor is the symptoms experienced by the patient along with his medical history, while the second one is the chemical interactions that happen inside the patient's body. Our system depends on the first factor in which the selection of medicine is made to avoid the interfering with other medicaments for a certain disease. It doesn't concentrate on more of other suspicious diseases but simply treats the current symptoms that are experienced by the patient.

REFERENCES

- [1] U.S. Pharmacopeia National Formulary, "United States Pharmacopeia (U.S. Pharmacopeia National Formulary)," Us Pharmacopeia .Pck Slp edition, ISBN-13, January 1, 2015.
- [2] U K Stationery Office, "British Pharmacopoeia (BP) 2016," Stationery Office Books (TSO), ISBN-13, August 24, 2015.
- [3] Sid M. Wolfe, "Worst Pills, Best Pills: A Consumer's Guide to Avoiding Drug-Induced Death or Illness," Gallery Books; Revised ed. edition, ISBN-13, January 4, 2005.
- [4] BMJ Group , Pharmaceutical Press , "BNF 69: March 2015 - September 2015 (British National Formulary)," Pharmaceutical Pr, 1 edition, ISBN-13, March 31, 2015.
- [5] Jenna M. Repts, Uwe Aickelin, Jiangang Ma, Melbourne, Yanchun Zhang, "Refining Adverse Drug Reactions using Association Rule Mining for Electronic Healthcare Data," arXiv:1502.05943v1 [cs.DB] , 20 Feb 2015.
- [6] Jie Chen , Hongxing He ,JiuyongLi , Huidong Jin , Damien McAullay ,Graham Williams, Ross Sparks and Chris Kelman, "Representing Association Classification Rules Mined from Health Data," R. Khosla et al., Eds. Springer-Verlag Berlin Heidelberg, 2005, pp. 1225–1231.
- [7] Yu Chen, Lars Henning Pedersen, Wesley W. Chu, Jorn Olsen, "Drug Exposure Side Effects from Mining Pregnancy Data," ACM SIGKDD Explorations Newsletter - Special issue on data mining for health informatics, Vol. 9 Issue 1, June 2007 pp. 22 - 29 .
- [8] Christian G. Daughton , Ilene Sue Ruhoy, "Lower-dose prescribing: Minimizing "side effects" of pharmaceuticals on society and the environment," Elsevier B.V. , CC BY-NC-ND license,2012.
- [9] Nwigbo Stella N , Agbo Okechuku Chuks,"Expert System: A Catalyst in Educational Development in," Human Resource Management Academic Research Society, Proceedings of the 1st International Technology, Education and Environment Conference, 2011.
- [10] Brian R. Walker BSc MD FRCPE FRSE , Nicki R Colledge BSc (Hons) FRCPE , Stuart H. Ralston MD FRCP FMedSci FRSE , Ian Penman BSc MD FRCPE, "Davidson's Principles and Practice of Medicine: With STUDENT CONSULT Online Access, 22e (Principles & Practice of Medicine (Davidson's)) 22nd Edition," Churchill Livingstone, ISBN-13, 22 edition ,February 15, 2014.
- [11] Robert B. Doorenbos," Production Matching for Large Learning Systems," Robert B. Doorenbos, Thesis for the degree of Doctor of Philosophy, CMU-CS-95-113 ,January 31, 1995.