World Applied Sciences Journal 26 (9): 1204-1208, 2013
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2013.26.09.13551

# Estimate or the Effectiveness of Intelligent Information System of Early Diagnosis and Prognosis of Cardiovascular Disease

*Olga Alekseevna Efremova, Valeriy Mihailovich Nikitin, Elena Andreevna Lipunova, Denis Aleksandrovich Anohin and Ludmila Aleksandrovna Kamyshnikova*

Belgorod National Research University, Belgorod, Russia

**Submitted:** Sep 25, 2013; **Accepted:** Nov 26, 2013; **Published:** Nov 30, 2013

**Abstract:** This article describes a new approach to the implementation of the training program of risk probability of coronary heart disease (CHD) remote evaluation. It describes an example of training sample formation based on the statistical analysis of medical data. The proposed solution will allow medical personnel to carry out the personification of cardiac services and first of all, preventive orientation. Also a method for assessing the effectiveness of the program of expert system "ARM-Cardiologist" on the results of ongoing clinical trials is provided.

**Key words:** Telecardiology · Decision support system · Workstation · Electronic diagnosis · CHD · Bayes' theorem

## INTRODUCTION

Cardiovascular diseases (CVD) are the most frequent cause of death among the population and the "time factor" is of special importance to treat them. Significant reduction of the risk of this class of diseases can be achieved by means of preventive measures among the population. One of the main requirements for the organization of mass screening of the population is the high throughput of medical offices and minimizing the cost of screening and acceptable precision of medical decisions. The use of telemedicine technologies that allow remote sensing methods to diagnose the cardiovascular pathologies and particularly the dynamics of their development can provide timely medical care to prevent death or disability of the person.

The analysis of the given articles [1-5] devoted to solving the problem of providing remote and automated help to cardiac patients has shown that the existing automated systems do not allow checking the adequacy of the valuation, the training function of the system is absent or limited and the medical history data of the patient is not fully used. Another disadvantage is the fact that among the currently existing information systems there is no database linking together individual databases on diseases, symptoms, course of treatment and medicines. The reported deficiencies make it difficult to apply the existing development in the subject area for practical implementation in the form of intelligent information systems to support medical management solutions (MMS).

Thus, the relevance of this study is due to the need to improve the quality of forecasting and early diagnosis of heart disease through the use of leading predictors of cardiovascular disease, such as the structural and functional parameters of the heart, using modern mathematical techniques of intelligent information technology.

**The Research Methodology:** Is based on the use of modern information technologies and, in particular, systems, human-machine interaction. The proposed approach is used in medical practice, automated generation of medical management solutions (MMS) [6-9]. The research team has proposed the approach to evaluate the effectiveness of intelligent information system of early diagnosis and prognosis of cardiovascular disease. At the first stage, the creation of workstation "ARM-Cardiologist" and the software package, which includes the database (DB) on the diagnostic features of CHD and electronic cardio cases of the surveyed patients.

**Corresponding Author:** Efremova, Belgorod National Research University, Pobeda Street, 85, 308000, Belgorod, Russia.

Next, based on a comprehensive analysis of the various cardio data derived from non-invasive monitoring of the current CVS patient, the estimation of the conditional probability of the risk of coronary heart disease (CHD) is determined, the value of which is necessary for the subsequent generation of MMS. One of the important stages of the technology is to create a database (DB) on the diagnostic features of CHD, electronic cardio cases of the surveyed persons, as well as the formation of training set and training programs for the early diagnosis and prognosis of cardiovascular disease by cardiologist. The characteristic feature of software developed by a team for the "ARM-Cardiologist" is that the training set is updated continuously by incorporating into its structure, data about each of the surveyed persons for which the diagnosis of a system of man-machine interaction (cardiologist + "ARM-Cardiologist") at the stage of clinical testing and operation of the workstation.

## RESULTS

One of the main procedures of the proposed algorithm - is the formation of diagnostic features (DF) of coronary heart disease and the calculation of their values for the individual patient. The algorithm involves the calculation of risk of coronary heart disease conditional probability in a patient, which on the basis of the analysis, cardio data state classification of CVS patient groups according to DF CHD. In this case, under the DP CHD means:

- Symptoms of coronary heart disease (such as chest pain, etc.);
- Risk factors for coronary heart disease (e.g., smoking, etc.);
- The results of the clinical examination;
- Instrumental data (opinions during electrocardiography and echocardiography, etc.).

Then, a mathematical model of the state as a set of calculated CVS group numbers (the values) DP CHD patient is formed.

The next stage of the algorithm involves the creation of the training set as a consequence of pre-computed values of the frequency of occurrence of the presence or absence of coronary heart disease for different values of a representative sample of DF patients, who have already been appropriate diagnosed. Then it computes the estimated risk of coronary heart disease by comparing the parameters of the mathematical model of the current state of the CVS of the individual patient with the appropriate parameters generated by training set. Furthermore, [6, 8], the training set as the reference information can be used to identify the dynamics and direction of parameters of the mathematical model the state of CHD patient. This information is needed to clarify the missing values of DF, as well as the personification of the data by assigning a cardiologist respective initial weights the importance of each individual patient DF, which is recorded in his cardio case.

Thus, the whole procedure of training programs "ARM-Cardiologist" by a cardiologist is as follows:

- The personification of the program through the initial appointment and subsequent relevant adjustments of weight coefficient of the importance of each DF. In this case, the initial assignment of values to boost factors DF is done automatically on the basis of the previously set value (in the software interface called "common values"), which reduces the degree of impact of a DF on a calculated risk assessment of coronary heart disease, if a cardiologist does not consider them appropriate. We note that the correct adjustment of the significance of the coefficients of each DF will depend on several factors. First you need to identify incorrect values of DF, probably due to the relatively small sample of the training or wrong DFs and their groups. Significance of the error correction coefficients may also be due to insufficient training of the cardiologist performing during the procedure training. All the significance of the adjustment coefficients is stored in the DF as personal to the individual patient. Personal factors have a greater priority than the general at their mutual mismatch.
- The addition of training set by adding data to it on the patient, in which completed the process of calculating the risk assessment of CHD, the diagnosis is generated on the presence or absence of CHD and obtained the consent of the cardiologist on the adequacy of electronic diagnosis.

Thus, the drawbacks of the prior works [1, 2], implied by the absence or limited learning function by extracting the knowledge of a cardiologist in the process are eliminated.

For the direct calculation of the risk assessment of CHD the following procedures are carried out. In accordance with the classification produced by the state of CVS patient groups DF CHD the sampling rate from the database [3] of change in the risk of coronary heart disease is carried out. In this case, the conditional probability that the patient has CHD is in accordance with (1):

$$P(A_i|B) = \frac{P(A_i) * \prod_j \left( P(B_{jk}|A_i) \right)}{\sum_i \left( P(A_i) * \prod_j \left( P(B_{jk}|A_i) \right) \right)} \quad (1)$$

where,

$P(A_i|B)$ - The conditional probability of the truth of the hypothesis $A_i$ in B event;

$A_i$ - A hypothesis corresponding to the assumption of the presence or absence of CHD in a patient form, where the i- number of hypotheses about the presence or absence of CHD forms: 1 - has CHD 2 – doesn't have CHD;

B - An event corresponding to a situation where the state was classified CVS patient groups diagnostic features;

$P(A_i)$ - A priori probability of truth of a hypothesis $A_i$ is calculated in accordance with the relation:

$$P(A_i) = \frac{\sum_j \sum_k a_{jki} + 1}{\sum_i \left( \sum_j \sum_k a_{jki} + 1 \right)} \quad (2)$$

where,

$a_{jki}$ - the number of elements in the training set equal to the number of patients whose condition was classified CVS in the k- fifth group of the j- DP CHD. It was observed that performance of i- hypothesis;

j - Number of DP CHD;

k - Number of groups of the DP, which was classified by the state of diagnosed CVS patients;

$V_{jk}$ - An event corresponding to a situation where the state were classified CVS particular patient in the k- fifth group of the j- DP;

$P(B_{jk}|A_i)$ - The conditional probability of occurrence of an event ($B_{jk}$) - implementation status classification CVS patients in the k- fifth group of the j- DP with the truth of the i-th hypothesis (Ai). $P(B_{jk}|A_i)$ is calculated according to equation (3):

$$\begin{cases} P\left(B_{j\kappa}|A_t\right) = \frac{1}{2}, \text{ at } k = 1; \\ P\left(B_{j\kappa}|A_t\right) = \frac{a_{jki} + 1}{\sum_k \left(a_{jki} + 1\right)}, \text{ at } k \neq 1. \end{cases} \quad (3)$$

It should be noted that in calculating $P(B_{jk}|A_i)$may be the case when, due to insufficient initial cardio cases fails to classify the patient's condition CVS one of a group DF. To start the process of calculating the risk assessment of coronary heart disease in this situation an additional group is introduced at number 1 (k = 1), "Group is not defined". As for the calculation of $P(B_{jk}|A_i)$ for the first group is not used for the training set and $P(B_{jk}|A_i)$ is assigned a value equal- probability for all alternatives that because of the procedure of normalization factors are not subsequently affect the final score. When receiving the additional information about the patient, allowing classifying his CVS state under previously undefined DF, the above procedure is the introduction of an additional group # 1 is canceled.

It should be noted that the values of $P(A_i)$ and $P(B_{jk}|A_i)$ are calculated beforehand on the basis of training set that is created with statistical data obtained during the survey and survey of the patients. These statistics are presented in a table whose rows correspond to groups of diagnostic features and columns - a proposed diagnosis (presence or absence of coronary heart disease). Data that contains the table represent the number of patients whose condition was classified by CVS in one of the groups of each DF with known CHD. The peculiarity of the formation of the training set used for the calculation of risk assessment for coronary heart disease is that it can replenish it by making patient data by which completed the process of diagnosis of the presence or absence of their coronary heard disease.

For example, let's consider the fragment of training set, built on the basis of national statistics of morbidity in the United States [3], illustrated in Table 1. Substituting the data Table 1 in equation (3) it can be shown that for this example, P(A1) = 0,042; P(A2) = 0,958. From which it follows that the initial probability of detection of CHD in a patient is 0,042. It should be noted that the relation (3),

Table 1: The procedure for determining the structure and presentation of the results of clinical trials

| Surveyed patients | The results of the electronic diagnostics: Positive | Negative | Total |
|---|---|---|---|
| $n_1$ | TP | FN | TP+FN |
| $n_0$ | FP | TN | FP+TN |
| Total | TP+FP | FN+TN | TP+FP+FN+TN |

Note: n1 - the number of patients in the experimental sample for which a cardiologist diagnosed 1st Class disease (CHD);
n0 - the number of patients in the experimental sample for which a cardiologist diagnosed the absence of the disease under study.

whereby the calculated estimate $P(A_i)$, to minimize the degree of influence on the final result of the DFs whose values found using the limited selection of patients. In general, the assessment of a priori probabilities is calculated taking into account the general population statistics detecting coronary heart disease in all populations. The value of $P(A_i)$ is calculated as the ratio of the number of CHD patients to the total number of patients in the training set. In a situation when the training set is not present, the numerical value of $P(A_i)$ is calculated as the ratio of the number of CHD patients to the population of the earth, country, region, city, professional affiliation, etc. Also it should be noted that in calculating $P(A_i)$ on the basis of learning sample data can be a problem, which is that for some information DP has only a portion of the patients and in some groups DF - information is completely missing. In this case, there is the need for a priori probability for each incidence of CHD DF and, consequently, to calculate the average value of $P(A_i)$ for all DFs.

The characteristic feature of computing $P(B_{jk}|A_i)$ in accordance with (3) is that the classification of the state of the patient on the CVS DF in the first group (k = 1), a group for all DF is called "The Group is not defined" is always given a condition:

$$P(B_{j1}|A_i) = 0,5 \quad (4)$$

This approach to the estimation of $P(B_{jk}|A_i)$ allows you to keep the ability to start the process of calculating the risk assessment of coronary heart disease in a situation where according to the individual DF is impossible to classify the patient's CVS (the condition of incomplete input of cardio data). Another feature of the calculation of $P(B_{jk}|A_i)$ is that if $k \neq 1$ there may be situations related uncertainties of the first and second kind. This situation arises when a training sample in any DF groups doesn't have fixed patients. In this situation, the calculation of $P(B_{jk}|A_i)$ is carried out in accordance with the second equation system (3).

**The Evaluation of Performance of the Expert System Program "APM-Cardiologist:** When generating the primary electron CHD diagnosis was made based on the known method of calculation using the diagnostic efficiency in Table 1, filled by the results of theoretical and clinical trials [10]. This method allows us to estimate the extent to which the results of the generation of an electronic program CHD diagnosis results of the reference diagnosis (RD). Thus as a result of RD considered diagnostic results carried cardiologist in respect of one and the same group of patients.

Experimental sample includes 107 items previously diagnosed medical patients. The selection of samples was wearing random. As estimates of the quality of the generation of electronic projects CHD diagnoses were used: diagnostic sensitivity (DSe), the diagnostic specificity (DSp), the predictive value of positive results (PR+), negative predictive value of the results (PR - ), the diagnostic efficiency of decision rule, electronic diagnostics (ED). Quantitative estimates of the mentioned indicators are in accordance with the equation (5):

$$\begin{cases} DSe = TP/n_1; \\ DSp = TN/n_0; \\ PR^+ = TP/(TP+FP); \\ PR^- = TN/(FN+TN); \\ ED = (TP+TN)/(TP+FP+FN+TN). \end{cases} \quad (5)$$

where,

TP- the number of patients in the experimental sample for which the result of electronic diagnostics is a true positive, that is, the diagnosis of "revealed the presence of the disease" and put a cardiologist and program;

FP - ЛПП the number of patients in the experimental sample for which the result of electronic diagnostics are false positive, then there is a diagnosis, "revealed the absence of disease" put the program does not coincide with the diagnosis of a cardiologist;

FN - the number of patients in the experimental sample for which the result of electronic diagnostics are false negative, the diagnosis is a cardiologist and an electronic diagnosis and distinguished physician revealed the presence of disease;

TN - the number of patients in the experimental sample for which the result of electronic diagnostics is a true negative, the diagnosis is a cardiologist and an electronic diagnosis and the same doctor revealed the absence of disease.

Table 2: The results verify the effectiveness of electronic diagnostic coronary artery disease based on clinical trials

| Surveyed patients | The results of the electronic diagnostics | | |
|---|---|---|---|
| | Positive | Negative | Total |
| $n_1$ | 49 | 4 | 53 |
| $n_0$ | 8 | 46 | 54 |
| Total | 57 | 50 | 107 |

When calculating the values in accordance with (5) we used the results of clinical trials, the procedure for determining the structure and representations are illustrated in Table 1.

These figures are calculated on the basis of clinical trials and shown in Table 2.

Table 2 illustrates the results of the pilot study the effectiveness of electronic diagnostic coronary heart disease, based on the data obtained from the Table 1.

In accordance with relation (5) and the data illustrated in Table 2, the following calculations were made of quality electronic projects generate diagnoses of coronary heart disease:

$$\begin{cases} DSe = 49/53 = 0{,}925; \\ DSp = 46/54 = 0{,}852; \\ PR^{+} = 49/57 = 0{,}86; \\ PR^{-} = 46/50 = 0{,}92; \\ ED = 95/107 = 0{,}888. \end{cases} \quad (6)$$

## CONCLUSION

The efficiency and the possibility of training programs developed intellectual assessment of the conditional probability of CHD are shown. This study was carried out on the example of the training set and is formed according to the results of the statistical analysis of medical data describing the state of the CVS on the diagnostic features of coronary heart disease. The efficiency of the information system of early diagnosis and prognosis of cardiovascular disease based on the results of clinical trials was observed. The implementation of the proposed research team of technical solutions to improve the accuracy of early diagnosis and prognosis of coronary heart disease.

## ACKNOWLEDGEMENTS

This article was prepared under a grant to conduct research in priority areas of socio-economic development of the Belgorod region in 2013-2014.

## REFERENCES

1. Matthew, J.L. and R.L. Nydick, 2008. The analytic hierarchy process in medical and health care decision making: A literature review. European Journal of Operational Research, 189: 194-207.
2. Bahill, A.T., K. Bharathan and R.F. Curlee, 1995. How the testing techniques for a decision support system changed over nine years. IEEE Transactions on Systems, Man and Cybernetics, 25(12): 1533-1542.
3. Genest, J., R. McPherson and J. Frohlich, 2009. Canadian Cardiovascular Society / Canadian guidelines for the diagnosis and treatment of dyslipidemia and prevention of cardiovascular disease in the adult—recommendations. Canadian Journal of Cardiology, 25(10): 567-579.
4. Thomas, S.V., J.R. Kurup and A. Kuruvilla, etc., 2001. An expert system for the diagnosis of epilepsy: results of a clinical trial. Natl. Med. J. India, 14(5): 274-276.
5. MYCIN: A Quick Case Study. Databases and Artificial Intelligence. Date Views 08.07.2013 http://cinuresearch.tripod.com/ai/www-cee-hw-ac-uk/_alison/ai3notes/section2_5_5.html.
6. Nikitin, V.M., E.A. Lipunova, O.A. Efremova, D.A. Anokhin, A.A. Merzlikin and I.N. Kurshubadze, 2013. Algorithm for primary diagnosis of ischemic heart disease for a program of expert system ARM- cardiologist. Scientific Journal of Belgorod State University. Series "Medicine. Pharmacy, 4 (147): 187-194.
7. Nikitin, V.M., M.I. Postelnic, E.A. Lypunova, D.A. Anokhin and I.I. Ivanov, 2011. Information model of the heart. International Journal of applied and Fundamental Research, 10: 13.
8. Anokhin, D.A., V.M. Nikitin, V.V. Lomakin, I.I. Ivanov and A.A. Merzlikin, 2011. Intelligent information system by cardiac diseases and drug - ARM-cardiologist: a certificate of official. registration of the computer # 2011617001 Russian Federation. Appl. # 2011615294 15.07.2011, publ. 08.09.2011.
9. Anokhin, D.A., V.M. Nikitin, V.V. Lomakin, I.I. Ivanov and A.A. Merzlikin, 2011. Database information system to support decision- making in the ARM-Cardiologist with relational links between data on diseases, symptoms and syndromes, treatments, medicines: a certificate that official. Registration Database # 2011620419 Russian Federation. Appl. # 2011620281 20.04.2011, publ. 08.06.2011.
10. Omelchenko, V.P. and A.A. Demidov, 2001. Workshop on Medical Informatics: tutorial, Rostov-on -Don, pp: 304.