Original article

Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters

Oleg Yu. Atkov (MD, PhD)a, Svetlana G. Gorokhova (MD, PhD)b,⁎, Alexandr G. Sboev (PhD)c, Eduard V. Generozov (PhD)d, Elena V. Muraseyeva (MD, PhD)e, Svetlana Y. Moroshkina d, Nadezhda N. Chernyi c

a N.I. Pirogov Russian State Medical University, Moscow, Russia
b I.M. Sechenov First Moscow State Medical University, Moscow, Russia
c MEPhi National Research Nuclear University, Moscow, Russia
d Research Institute of Physical-Chemical Medicine, Moscow, Russia
e Central Clinical Hospital No. 2 of Russian Railways JSC, Moscow, Russia

Received 18 November 2011; accepted 21 November 2011
Available online 2 January 2012

KEYWORDS
Coronary heart disease; Risk factors; Artificial neural networks

Summary The aim of this study was to develop an artificial neural networks-based (ANNs) diagnostic model for coronary heart disease (CHD) using a complex of traditional and genetic factors of this disease. The original database for ANNs included clinical, laboratory, functional, coronary angiographic, and genetic [single nucleotide polymorphisms (SNPs)] characteristics of 487 patients (327 with CHD caused by coronary atherosclerosis, 160 without CHD). By changing the types of ANN and the number of input factors applied, we created models that demonstrated 64—94% accuracy. The best accuracy was obtained with a neural networks topology of multilayer perceptron with two hidden layers for models included by both genetic and non-genetic CHD risk factors.
 © 2011 Japanese College of Cardiology. Published by Elsevier Ltd. All rights reserved.

Introduction

The mortality rate caused by coronary heart disease (CHD) has been changing since the 1990s and, in some industrialized countries, shows a decline. However, morbidity from myocardial infarction and angina/CHD remains high in some subgroups, the highest being male workers and the elderly [1—6]. It poses a serious problem, and the development of
methods for CHD prediction is of immediate scientific and practical interest.

Several algorithms of risk stratification and diagnostic models for CHD have already been created. They are based on different sets of risk factors, established in epidemiologic studies and randomized controlled trials, such as arterial hypertension, hypercholesterolemia, diabetes mellitus, and smoking [1,7–12]. Some authors suggest using a coronary calcium score [13,14] and retinal vascular signs [15] etc. as additional factors. Recent developments focus on genetic markers for prediction of CHD and recommend including single nucleotide polymorphisms (SNPs) for risk assessment [16–19].

It seems that objective difficulties in CHD detection are caused by the multiplicity of risk factors to be taken into consideration. This makes it necessary to survey the structure of the variable risk factors and to create an efficient classification system. Therefore, computer methods of intelligent data processing are available and applied for this purpose, and, on this basis, expert medical systems have been created.

One of these promising methods is artificial neural networks (ANNs), a highly effective tool used in classification tasks, as well as to solve many important problems, such as signal enhancement, identification, and prediction of signals and factors. The important feature of ANNs is their adaptivity. This enables them to be applied in cases where it is impossible to create a strict mathematical model but where there is a sufficiently representative set of samples. The other important characteristic of neural networks is their capacity to generalize input information and to give correct answers for “unfamiliar” data, which makes them effective in solving complicated classification problems [20,21]. Today, ANNs are applied in clinical and genetic research. Attempts have been made to create diagnostic models for various diseases with the use of ANNs of different topologies [22–25]. It is assumed that using complexes of signs of CHD will allow ANNs to not only diagnose, but also to predict clinically significant events, myocardial infarction being the first.

The aim of this study was to develop an ANNs-based diagnostic model for CHD using the complex of traditional and genetic factors of this disease.

Materials and methods

The study included 487 patients (males = 425, females = 62, mean age: 51.25 ± 9.74 years) hospitalized in Central Clinical Hospital No. 2 of Russian Railways JSC for a coronary angiography to diagnose CHD. All patients underwent uniform standard clinical examinations (laboratory tests, electrocardiogram, Holter monitoring, stress tests, echocardiography etc.), coronary angiography (Advantx, General Electric, Waukesha, WI, USA), and genetic analysis. The diagnosis of CHD was made following both the clinical and coronaryography results. The information obtained from testing and genotyping allowed us to create a database of patients that was subsequently used to diagnose CHD using ANNs.

Risk factors

CHD risk factors included in the analysis were: age, gender, total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, very-low-density lipoprotein (VLDL) cholesterol, triglycerides, cholesterol ratio, fasting plasma glucose, arterial hypertension, diabetes mellitus, current tobacco smoking status, obesity (Quetelet index, body mass index BMI > 30), and a family history of CHD. Additional factors taken into account were profession (locomotive driving), risk of fatal cardiovascular disease according to the European SCORE project scale (SCORE index) [1], left ventricular ejection fraction, and coronary angiography data.

Genotyping

Genotyping was performed by an allele-specific primer extension of multiplex amplified products and detection with a matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy on an AutoPhlex II MALDI-TOF MS (Bruker Daltonics, Billerica, MA, USA). The analyzed panel includes 14 SNPs localized in genes involved in CHD pathogenesis: lipoprotein lipase [LIPC 250G/A (rs2070895) and LIPC 514C/T (rs1800588)], nitric oxide synthase [NOS298D (rs1799983)], methylenetetrahydrofolate reductase [MTHFR A223V (rs1801133)], angiotensin-converting enzyme [ACE Alu Ins/Del I–D (rs4646994)], angiotensinogen [AGT M235T (rs699)], and AGT T174M (rs4762) variants, angiotensin II type 1 receptor [AGTR A1166C (rs5186)], plasminogen activator inhibitor-1 [PAI-1 5G/4G (rs1799869)], and C-reactive protein [CRP-1 (rs1800947)], CRP-2 (rs1417938), CRP-3 (rs1205), CRP-4 (rs3093068), and CRP-5 (rs11308641).

Artificial neural network

The neural network model was created using a multilayer perceptron, the multi-level neural feedforward network taught by the statistical back propagation of error. The sets of variable parameters were selected in order to adjust ANN models by pairwise correlation between the database parameters and CHD diagnosis. Accuracy of models was improved by a genetic algorithm with different optimization parameters [26] including number of neurons in the hidden layer, number of inputs to the neural network, and slope coefficient of activation functions. NeuroSolutions 5.0 development environment (NeuroDimension Inc., Gainesville, FL, USA) was used to check the possibilities of optimization. The ANN was created by inputting the specified variable factors. The task to solve was of two-class classification: “1” — CHD, “0” — healthy. A total of 287 examples were used for teaching, 100 for cross-validation, and 100 for testing.

Results

Based on the results of the clinical instrumental research, CHD caused by coronary atherosclerosis was diagnosed in
Table 1  Accuracy of ANN models for CHD diagnosis. a

<table>
<thead>
<tr>
<th>Model</th>
<th>Factors</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Age, profession, diabetes, arterial hypertension, smoking, obesity, family anamnesis of CHD, glucose, cholesterol</td>
<td>64</td>
</tr>
<tr>
<td>II</td>
<td>Age, profession, diabetes, arterial hypertension, smoking, obesity, family anamnesis of CHD, glucose, cholesterol</td>
<td>77</td>
</tr>
<tr>
<td>III</td>
<td>Age, profession, diabetes, arterial hypertension, smoking, obesity, family anamnesis of CHD, glucose, total cholesterol, HDL, LDL, VLDL, triglycerides, cholesterol ratio</td>
<td>83</td>
</tr>
<tr>
<td>IV</td>
<td>Age, profession, diabetes, arterial hypertension, smoking, obesity, heredity, glucose, total cholesterol, HDL, LDL, VLDL, triglycerides, cholesterol ratio, coronary angiography data</td>
<td>91</td>
</tr>
<tr>
<td>V</td>
<td>NOS, ACE, AGT-235, AGT-174, AGTR, CRP-1, CRP-2, CRP-3 SNPs</td>
<td>90</td>
</tr>
<tr>
<td>VI</td>
<td>NOS, ACE, AGT-235, AGT-174, AGTR, CRP-1, CRP-2, CRP-3 SNPs, SCORE index</td>
<td>83</td>
</tr>
<tr>
<td>VII</td>
<td>NOS, ACE, AGT-235, AGT-174, AGTR, CRP-1, CRP-2, CRP-3 SNPs, coronary angiography data</td>
<td>89</td>
</tr>
<tr>
<td>VIII</td>
<td>NOS, ACE, AGT-235, AGT-174, AGTR, CRP-1, CRP-2, CRP-3 SNPs, SCORE index, coronary angiography data</td>
<td>93</td>
</tr>
<tr>
<td>IX</td>
<td>NOS, ACE, AGT-235, AGT-174, AGTR, CRP-1, CRP-2, CRP-3 SNPs, HDL, LDL, glucose</td>
<td>90</td>
</tr>
<tr>
<td>X</td>
<td>NOS, ACE, AGT-235, AGT-174, AGTR, CRP-1, CRP-2, CRP-3 SNPs, age, smoking, obesity, family anamnesis of CHD, HDL, LDL</td>
<td>88</td>
</tr>
</tbody>
</table>

CHD, coronary heart disease; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; VLDL, very-low-density lipoprotein cholesterol; SNP, single nucleotide polymorphisms.

a In all cases, the multilayer perceptron neuronal network topology with two buried 4-neuron layers was used.

327 (67.2%) patients. A total of 160 (32.8%) patients had an intact coronary artery wall and no evidence of CHD.

Correlation analysis between the specified variable factors and CHD allowed us to choose 32 factors associated with the disease. These factors were analyzed by ANNs.

Approaches using different ANN types and a variable number of input factors (from 5 to 10) led to the models with 64—94% diagnostic accuracy. The best result (94%) was achieved in a multilayer perceptron (MLP) model with two buried layers and 10 factors (profession, LDL, HDL, triglycerides, cholesterol rate, SCORE index, left ventricular ejection fraction, family CHD history, coronary arteriography data, PAI gene). On the other hand, the same ANN type with 5 factors (coronary arteriography data, cholesterol rate, SCORE index, left ventricular ejection fraction, PAI gene) had a lower diagnostic accuracy (78%). However, the same 10 factors analyzed by other ANN types yielded a 79% result. This suggests that the diagnostic accuracy depends on the ANN type and the number of variable factors.

The next step included an analysis of 10 models of CHD prediction formed by different combinations of examination blocks (demographic characteristics, CHD history, laboratory tests, echocardiography and coronary angiography data, genes) by the MLP with two buried 4-neuron layers. The number of variable input factors ranged from 8 (models I and V) to 15 (model IV). The results are shown in Table 1. The minimal accuracy of 64% was obtained in model I, which included 8 non-genetic factors. Models IV, V, VIII, and IX, which had significantly different sets of variables, showed >90% diagnostic accuracy. Model IV included only non-genetic factors; model V included only eight SNPs (NOS, ACE, AGT-235, AGT-174, AGTR, CRP-1, CRP-2 and CRP-3); model VIII included the same SNPs in combination with coronary angiography data and SCORE index; model IX included the same SNPs and HDL, LDL, and glucose. Thereby the optimal accuracy was more often achieved when a diagnostic model included genetic factors.

Discussion

One of the modern approaches to the solution of classification problems is intelligent data processing based on resolving optimization tasks using ANNs. Our results suggest that ANNs may also be used to create a highly accurate and effective model for CHD prediction.

In this research we have examined the influence of different parameters, teaching methods, and neural network topologies on the accuracy of CHD diagnosis. We revealed that the optimal topology to resolve this classification task is a MLP with two buried layers. The optimal input parameters are 8—10 of the most significant factors: the use of all factors seems to excessively complicate the model, while, at the same time, smaller numbers do not provide essential information for resolving this problem. It is of interest to note that, in some models, the accuracy of CHD diagnosis was lower than 90% in spite of including coronary arteriography data, which are considered the “gold standard” for CHD diagnosis in clinical practice.

Publications on the possibilities of cardiologic application of artificial intelligence, ANN, first of all, are usually limited to different combinations of traditional laboratory and instrumental methods [22–25]. Therefore, the novelty of our study is the inclusion of genes as risk factors in order to estimate capabilities of CHD prediction. In this connection,
it is important to note that the accuracy of the diagnosis based on genetic factors only was found to be almost equal to that in the model combining 15 non-genetic factors (90% and 91%, respectively). The most informative (93%) was the model that included 8 SNPs, SCORE index, and coronary arteriography data. Taking into account the fact that genetic information is a permanent parameter, it can be assumed that analysis of NOS, ACE, AGT-3, AGT-4, AGTR, CRP-1, CRP-2, and CRP-3 SNPs enables prediction of CHD with a high accuracy.

The diagnostic accuracy can be improved not only by increasing the number of genetic markers, but also by their precise selection. Good candidates for consideration may be genes involved in vascular cell growth, apoptosis, and inflammation, or others associated with CHD, e.g. ABCA1 (ATP-binding cassette-transporter A1), CYP1A2 (cytochrome P450), ADRB group (β-adrenoreceptors), HSF1 (heat shock factors) [27], CLOCK, and BMAL1 [28]. The last ones are of interest because they are associated with the regulation of circadian rhythms, including those in the cardiovascular system [29].

The obtained experience in the development of neural network models for CHD prediction constitutes a basis for the design of applied software products for the diagnosis of cardiovascular diseases, including screening tests.

Conclusion

The purpose of this study was to create CHD diagnostic models with appropriate analytical characteristics using MLP ANNs. The best accuracy was obtained in models that included both genetic and non-genetic factors associated with the disease. The models of >90% accuracy may serve as the basis for the development of software tools for diagnosis and prediction of CHD.

References