THE CARDIOLOGIST
Vol.3, No.1, 2019, pp.34-62

LOGICAL-PROBABILISTIC DIAGNOSIS OF HUMAN CARDIOVASCULAR DISEASES BASED ON BOOLEAN ALGEBRA

Sharif E. Guseynov[1,2,3*], Daniel Chetrari[2], Jekaterina V. Aleksejeva[1,4] Alexander V. Berezhnoy[5], Abel P. Laiju[2]

[1]Institute of Fundamental Science and Innovative Technologies, Liepaja University, Liepaja, Latvia
[2]Faculty of Science and Engineering, Liepaja University, Liepaja, Latvia
[3]"Entelgine" Research & Advisory Co., Ltd., Riga, Latvia
[4]Riga Secondary School 34, Riga, Latvia
[5]Faculty of Information Technologies, Ventspils University of Applied Sciences, Ventspils, Latvia

Abstract. In the present paper, the possibility of applying the logic algebra methods for diagnosing human cardiovascular diseases is studied for an incomplete set of symptoms identified at a given time moment. In the work, the concept of the probability of having a disease is introduced for each particular patient, a formula for calculating probability of disease is proposed, and finally a probabilistic table of the patient's diseases is constructed based on calculated probabilities.

Keywords: *cardiovascular disease, Boolean algebra, medical diagnosis, cause-effect relation.*

Corresponding Author: *Dr. Sc. Math., Professor Sharif E. Guseynov, Liepaja University, Faculty of Science and Engineering, Institute of Fundamental Science and Innovative Technologies, 4 Kr.Valdemar Street, Liepaja LV-3401, Latvia, Tel.: (+371)22341717, e-mail: sh.e.guseinov@inbox.lv*

Manuscript received: *17 November 2018*

1. Introduction and specification of investigated problem

Before turning to the essence of the problem investigated in this paper, let us note that the term "symptom" in medicine refers to a frequently occurring sign of a disease, being detected using clinical research methods and used to diagnose and / or predict a disease. A symptom is called idiopathic or essential if the root cause for its appearance in a patient is unknown and it is not possible to identify and classify the disease causing this symptom. As a rule, in medicine, idiopathic (essential) symptom is stated into a separate independent disease (for instance, idiopathic epilepsy, idiopathic osteoporosis, idiopathic headache, essential hypertension, essential tremor, etc.).

In case if a symptom unconditionally points to the only disease, then it is called a pathognomonic symptom and it is classified into specific symptom category; but if the symptom is accompanied by a number of diseases, then it is classified to the non-specific symptom category. The combination of symptoms that often are faced together in a number of specific diseases and having a commonality of their aetiology and pathogenesis is called symptomatology. In this paper, with the term "symptom", in addition to the abovementioned medical definition and characteristics, we shall consider also the patient history data, hardware diagnostics results, laboratory tests, and so on. In addition, while discussing the detection, non-detection or absence of a symptom, in given paper we shall understand the appropriate medical report of the attending physician or medical researcher, who in each case decides whether this symptom is detected, is not detected or is missing.

Now let us encompass the essence of the problem studied in this paper.

As with any disease, a symptom of the possible incidence of a cardiovascular disease in a patient is confirmed by detection of a set of symptoms typical to the particular disease. However, on the basis of this standalone fact it is impossible to make a definitive diagnosis for the patient and, therefore, to prescribe the appropriate treatment, since

- first, many of the symptoms typical to a given disease may also contain signs of other diseases; for example, pain between the shoulder blades, neck pain, pain in left arm, shoulder pain, wrist pain, jaw pain can be a certain sign of heart disease such as stenocardia, and non-heart diseases such as osteochondrosis or myositis;
- secondly, the lack of detection of other symptoms at the given moment of time does not guarantee that these symptoms do not exist or they will not appear later; moreover, the veiled presence of not yet identified symptoms or the further appearance of new symptoms could impact and lead to fundamental change of the initial diagnosis, made only on the basis of the "initial" symptoms found: it is quite possible that per totality of the "initial" and newly revealed symptoms it follows that:
  - the patient has another cardiovascular disease, and not a cardiovascular disease that was originally diagnosed based on only the "initial" symptoms; for example, if the "initial" symptoms found are chest pain, shortness of breath and sustained fatigue feeling, on the basis of which the patient was diagnosed with atherosclerotic heart disease, the addition to these symptoms of fluttering arrhythmia and frequent fainting indicates that the patient is likely to have another cardiovascular disease, like the valvular heart diseases;
  - the patient has another disease that is not directly related to the cardiovascular system; for example, pain in the left shoulder, neck pain, pain in the lower part of jaw, pain between the shoulder blades and in the left infrascapular region may indicate that the patient has thoracalgia, which is a non-cardiac disease, and caused by squeezing or irritation of the roots of the intercostal nerves;
  - the original diagnosis is correct, but the patient also has another (or a number of other) heart or non-heart disease.

Due to the above described uncertainty in the cause-effect relations between diseases and their accompanying symptoms, a natural question arises: how can a physician make an accurate diagnosis based solely on the symptoms detected for this patient at a given moment of time?

First of all, let us note that the main purpose of establishing a diagnosis is the proper prescription of an appropriate treatment for a patient. Medical diagnostics (i.e. the process of establishing a diagnosis) consists of the following steps (for instance, see Balogh *et al*., 2015; Hoffman, 2014; Syzek, 2018; Melnikov, 1974).

1. Anamnesis, which consists of collecting information about the patient's subjective complaints; about the patient's living conditions; about previous diseases (Anamnesis Morbi); about the history of the development of the disease received from the patient and / or from his relatives; about the signs of the disease described by the patient and / or observed by the physician.

2. Physical examination, which is carried out by the physician during the primary medical inspection, palpation, percussion, auscultation.
3. Extra medical examination, which is performed by physician and qualified medical personnel using laboratory and instrumental methods.
4. Processing and assessment of the collected data.
5. Establishment of the diagnosis.

Proper implementation of stages 1-3 is entirely the prerogative of the higher and middle medical personnel, and is not considered in this paper. In this paper, mainly the fifth stage of medical diagnostics is studied, namely the stage of establishment of the diagnosis. However, due to the extreme importance of the fourth stage, this stage of medical diagnostics is briefly discussed from the perspective of possibility to apply mathematical methods, in particular, from the decision-making theory perspective, and in the context of the possibility of using computers to increase the capabilities of a physician, in particular, allowing to automate the processing as well as to improve the reliability and accuracy of the collected data assessment; to draw up such a treatment plan and to clarify the diagnoses, which could maximize the likelihood of giving the highest positive effect; to minimize the amount of required laboratory and instrumental medical research and procedures for a given patient; to receive an objective assessment of the results of current and further analyzes, etc. (for instance, see Melnikov, 1974; Ledley, 1962; Cutler, 1998; Pendyala & Figueira, 2017 and references given there).

As a rule, the physician begins the fourth stage of medical diagnostics by applying a repetitive process, where a logical analysis is carried out at each stage, allowing gradually to squeeze up the range of possible uncertainties between the possible diseases and the symptoms that are already found or that have not yet been discovered, but their appearance over time is assumed for some reason, for example, on the basis of their long term experience with particular category of patients; because of an epidemy; due to the longstanding extreme weather conditions; due to strong geomagnetical storms (for instance, see Alabdulgade, 2015; Malin & Srivastava, 1979; Stoupel, 1993; Karazyan, 1981); due to the nature and / or psychological condition and / or lifestyle of the patient; etc. As it was shown in the fundamental monograph (Lusted, 1968) and the significant and influential works (Ledley & Lusted, 1959; 1962; Lusted, 1960; Bruce, 1963; Kimura et al, 1963; Patton, 1978; Shusterman, 2004) if any uncertainties remain, then, using Bayes' theorem, useful quantitative estimates can be obtained that demonstrate the extent to which the diagnosis of random external factors and the correlation between diseases and symptoms influencing the diagnosis. As it is known (for instance, see Gnedenko, 2017), the Bayes' theorem is one of the main theorems of the probability theory, which allows one to determine the probability of a given event (A), provided that the probability of another event (B), is statistically interrelated with this event (A):

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)},$$

where $P(A)$ is a priori probability of event A, i.e. $P(A)$ shows (Bayes, 1763) the initial level of confidence to the assumption A before taking into account the previously known information B about the event A (assuming that it is available: $P(B) \neq 0$) or some new observations B of the event A (assuming that new observations have been or will necessarily be made: $P(B)\neq 0$); $P(A|B)$ is a posteriori probability of the event A, i.e. $P(A|B)$ shows a new level of confidence to the assumption A after taking into account previously known information or the new observations data; $P(B|A)$ is a probability of occurrence B for a true A. Obviously, with such an interpretation of the Bayes' theorem, it can be said that the value $P(B|A)\cdot P(A)$ shows how and to what extent the level of confidence to the assumption A is changed by taking into account known information about the event A or the data of new observations of this event.

In other words, having previously known information or new observations regarding an event, according to Bayes' theorem, the likelihood of this event can be clarified: as a part of this study – medical diagnostics dimension, this means that Bayes' theorem allows us to express the conditional probability of the incidence of some disease in patient taking into account the symptom found in him through the unconditional probability of the incidence of this disease and the conditional probability of the symptom found taking into account the disease. In other words, the Bayes' theorem as if allows rearranging the cause (disease) and effect (symptom): according to the detected symptom, the probability is calculated that this symptom was caused exactly by this disease. There are various advanced forms of the Bayes' theorem, which are successfully used in diagnostics, in particular, in medical diagnostics. In Bruce, 1963 and Warner, 1961 the same extended computational form of the Bayes' theorem was successfully applied to diagnose the acquired valvular disease and the congenital heart disease, respectively. In both papers there are the same rather restrictive assumptions: (A1) the symptoms do not depend on each other within the framework of the studied disease; (A2) the diseases themselves are mutually exclusive. Regarding the first assumption (A1), it is appropriate to note that if there is a sufficient amount of data on the coincidence of symptoms in each disease, then the desired independence of the symptoms could be verified by means of using the Pearson's chi-squared (χ^2) test. Assumption (A1) is indeed a limiting assumption, since the symptom independence condition is not often met. However, in Coirnfield, 1964, an authoritative author believes that there are no assumptions that the components of a multidimensional event are independent, and quite categorically asserts that, in medicine, binding the question of the applicability of the Bayes' theorem with the requirement that the symptoms within the studied disease be independent, is a common misconception. As for the second assumption (A2), it is a very restrictive assumption, since in the same period the patient may have several diseases (of course, there are mutually exclusive diseases: for example, hypothyroidism and hyperthyroidism; short-sightedness and long-sightedness; etc.).

Thus, we have briefly discussed some of the problems of the fourth stage of medical diagnostics – the stage of processing and assessment of the collected medical data. In conclusion, it is important to note that correlation methods do not play a significant role in medical diagnostics, in particular, when solving the very diverse problems of the fourth stage, as medical diagnostics is not a study about the principles and practice of classification, systematization and ordering of hierarchical objects or processes. The goal of medical diagnosis is not rooted in coming to an unequivocal indication of the disease on the existing symptoms in the logical way.

As it has been mentioned above, in this paper only the fifth stage of medical diagnostics is considered – the stage of establishing the diagnosis. This paper is limited to the study of only the most common cardiovascular diseases in humans. The mathematical method used in this study is based on the Boolean algebra, which is considered to be a widely known approach to the problem of medical diagnostics.

2. The most common heart diseases and their accompanying symptoms. Brief information on the basic concepts of Boolean algebra

2.1. The most common heart diseases and their accompanying symptoms

In the present paper, we consider the 6 most common heart diseases $D=\{D_i\}_{i=\overline{1,6}}$ and the 21 distinctive symptoms $S=\{S_j\}_{j=\overline{1,21}}$ accompanying these diseases.

Diseases and their designations:

- D_1 – Atherosclerotic heart disease;
- D_2 – Heart attack;
- D_3 – Heart arrhythmias (abnormal heartbeat);
- D_4 – Weak heart muscle (dilated cardiomyopathy);
- D_5 – Heart infection (endocarditis);
- D_6 – Valvular heart diseases.

Symptoms and their designations:

- S_1 – Chest pain;
- S_2 – Chest pressure;
- S_3 – Shortness of breath;
- S_4 – Pain, numbness, weakness or coldness in the legs or arms;
- S_5 – Pain in the neck, jaw, throat, upper abdomen or back;
- S_6 – Nausea;
- S_7 – Fatigue;
- S_8 – Sweating;
- S_9 – Light-headedness;
- S_{10} – A sense of doom;
- S_{11} – Irregular heartbeats / fluttering;
- S_{12} – Dizziness;
- S_{13} – Fainting;
- S_{14} – High blood pressure;
- S_{15} – Swelling of the legs, ankles, abdomen and feet;
- S_{16} – Unusual heartbeat sounds;
- S_{17} – Fever;
- S_{18} – Dry or persistent cough;
- S_{19} – Skin rashes or unusual spots;

- S_{20} – Weight loss;
- S_{21} – Narrowing / improper closing of heart valves

Using information from sources[1], we obtain the following list, which contains information about which diseases are accompanied by which symptoms, and vice versa:

- The disease D_1 be always attended by the symptoms $\{S_1, S_2, S_3, S_4, S_5, S_6, S_7\}$;
- The disease D_2 be always attended by the symptoms $\{S_1, S_3, S_5, S_6, S_8, S_9, S_{10}\}$;
- The disease D_3 be always attended by the symptoms $\{S_1, S_3, S_8, S_9, S_{11}, S_{12}, S_{13}, S_{14}\}$;
- The disease D_4 be always attended by the symptoms $\{S_3, S_7, S_9, S_{11}, S_{12}, S_{13}, S_{15}, S_{16}\}$;
- The disease D_5 be always attended by the symptoms $\{S_3, S_7, S_8, S_{11}, S_{15}, S_{17}, S_{18}, S_{19}, S_{20}\}$;
- The disease D_6 be always attended by the symptoms $\{S_1, S_3, S_7, S_{11}, S_{13}, S_{15}, S_{21}\}$;
- The symptom S_1 always accompanies the diseases $\{D_1, D_2, D_3, D_6\}$;
- The symptom S_2 always accompanies the disease D_1;
- The symptom S_3 always accompanies the diseases $\{D_1, D_2, D_3, D_4, D_5, D_6\}$;
- The symptom S_4 always accompanies the disease D_1;
- The symptom S_5 always accompanies the diseases $\{D_1, D_2\}$;
- The symptom S_6 always accompanies the diseases $\{D_1, D_2\}$;
- The symptom S_7 always accompanies the diseases $\{D_1, D_4, D_5, D_6\}$;
- The symptom S_8 always accompanies the diseases $\{D_2, D_3, D_5\}$;
- The symptom S_9 always accompanies the diseases $\{D_2, D_3, D_4\}$;
- The symptom S_{10} always accompanies the disease D_2;
- The symptom S_{11} always accompanies the diseases $\{D_3, D_4, D_5, D_6\}$;
- The symptom S_{12} always accompanies the diseases $\{D_3, D_4\}$;
- The symptom S_{13} always accompanies the diseases $\{D_3, D_4, D_6\}$;

[1] - NHS inform: A to Z list of common illnesses and conditions including their symptoms, causes and treatments. Available on-line at: https://www.nhsinform.scot/illnesses-and-conditions/a-to-z [Accessed: November 27, 2018];
- 1998-2019 Mayo Foundation for Medical Education and Research (MFMER). Available on-line at: https://www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes/syc-20353118 and https://www.mayoclinic.org/diseases-conditions/endocarditis/symptoms-causes/syc-20352576 [Accessed: December 31, 2018];
- Classification of Diseases (ICD). Available on-line at: https://www.who.int/classifications/icd/en/ [Accessed: November 29, 2018];
- Healthline Media. Available on-line at: https://www.healthline.com/health/atherosclerosis#symptoms [Accessed: December 31, 2018];
- Diseases of the heart and blood vessels: the main signs and primary symptoms. Available on-line at: http://www.aif.ru/health/life/bolezni_serdca_i_sosudov_osnovnye_priznaki_i_pervye_simptomy [Accessed: December 31, 2018].

- The symptom S_{14} always accompanies the disease D_3;
- The symptom S_{15} always accompanies the diseases $\{D_4, D_5, D_6\}$;
- The symptom S_{16} always accompanies the disease D_4;
- The symptom S_{17} always accompanies the disease D_5;
- The symptom S_{18} always accompanies the disease D_5;
- The symptom S_{19} always accompanies the disease D_5;
- The symptom S_{20} always accompanies the disease D_5;
- The symptom S_{21} always accompanies the disease D_6.

2.2. Brief information on the basic concepts of Boolean algebra

As it has been mentioned in the Introduction, the mathematical apparatus used in this work is based on the logical algebra. In order to save the readers who are not familiar or have little knowledge of the fundamental toolset of Boolean algebra, from having to look for relevant information in other sources to understand the mathematical calculations in the fourth and fifth sections of this paper, the necessary information about the basic logical operators on elementary Boolean functions and their properties is given below in a very brief and understandable fashion. The material provided below is more than sufficient for the reader to make him understand the essence of the mathematical approach used. For an in-depth knowledge of discrete mathematics and its applications, in particular, on the Boolean algebra, the fundamental textbook (Rosen, 2012) is highly recommended, which, in our opinion, could be regarded as one of the most remarkable textbooks in this field.

Basic logical operators.

- Conjunction logical operator (denoted as $\wedge$, or &, or using the multiplication symbol, i.e. $x \wedge y$, $x \& y$, $x \cdot y$) matches every two propositions with a new proposition which is true if and only if both source propositions are true. A conjunction is sometimes called a logical multiplication.
- Disjunction logical operator (denoted as $\vee$, i.e. $x \vee y$) matches every two propositions with a new proposition, which is false if and only if both source propositions are false. A disjunction is sometimes called a logical addition.
- Inversion logical operator (denoted as $\neg$ or $\bar{\ }$, i.e. $\neg x$, $\bar{x}$) puts in correspondence to each proposition a new proposition, the value of which is opposite to the original one. Inversion is sometimes called a logical negation.
- Implication logical operator (denoted as $\rightarrow$, i.e. $x \rightarrow y$) is defined as logical binding "if ..., then ...", and is calculated according to the formula $x \rightarrow y = \bar{x} \vee y$, where x is termed as the antecedent of the conditional, y is termed as the consequent of the conditional. An implication matches every two propositions with a new proposition, which is false if and only if where the proposition that is antecedent of the conditional is true and proposition that is a consequent of the conditional is false.
- Equality logical operator (denoted as $\leftrightarrow$ or as $\Leftrightarrow$, i.e. $x \leftrightarrow y$, $x \Leftrightarrow y$) matches every two propositions with a new proposition, which is false if and only if where one of the source propositions is false and the other is true. This definition can be formulated differently: an equality logical operator associates a new

proposition with each two propositions, which is true if and only if both source propositions are simultaneously false or true. Logical equality is calculated according to the formula $x \leftrightarrow y = x \cdot y \vee \bar{x} \cdot \bar{y}.$

- Exclusive disjunction logical operator (other name: modulo-2 addition logical operator; denoted as $\oplus,$ i.e. $x \oplus y$) matches every two propositions with a new proposition, which is false if and only if both of the source propositions are simultaneously false or true. This definition can be formulated differently: the logical exclusive disjunction operator puts in correspondence to every two propositions a new proposition that is true if and only if one of the original statements is false and the other is true. It could be easily noticed that the exclusive disjunction logical operation is the opposite operation with respect to logical equivalence: $x \oplus y = \overline{x \leftrightarrow y}.$ The exclusive disjunction logical operation is calculated according to the formula $x \oplus y = x \cdot \bar{y} \vee \bar{x} \cdot y.$
- Peirce's arrow or joint denial logical operator (denoted as $\downarrow,$ i.e. $x \downarrow y$) matches every two propositions with a new proposition, which is true if and only if both source propositions are false. Peirce's arrow logical operator is calculated according to the formula $x \downarrow y = \overline{x \vee y}.$
- Sheffer stroke logical operator (denoted as $|,$ i.e. $x \,|\, y$) matches every two propositions with a new proposition, which is false if and only if both source propositions are true. Sheffer stroke logical operator is calculated according to the formula $x \,|\, y = \overline{x \cdot y}.$

In the absence of additional brackets, the foregoing logical operators, when executed, have the following priority: at first inversion logical operator, then conjunction logical operator, then equality logical operator or modulo-2 addition logical operator, then disjunction logical operator, then implication logical operator, or Peirce's arrow logical operator, or Sheffer stroke logical operator. For example, the combination $\overline{x \vee \bar{z}} \cdot y \oplus z \to \bar{y} \cdot z \vee \bar{x}$ must be understood as $\left(\left(\overline{(x \vee \bar{z})} \cdot y\right) \oplus z\right) \to \left((\bar{y} \cdot z) \vee \bar{x}\right).$

The set of logical operators $\{\wedge, \vee, \neg\}$ is called a classical basis, and any logical operator can be expressed by the operators of classical basis.

Elementary Boolean functions.

Boolean set, which is denoted as $\mathbf{B},$ is a set consisting of two elements. As these elements are usually taken the numbers 0 and 1, where 0 is interpreted as "false", and 1 is interpreted as "true": $\mathbf{B} = \{0,1\}$ or $\mathbf{B} = \{F, T\},$ where $F =$ "false", $T =$ "true". A function f is called a Boolean function if its arguments take values from the Boolean set $\mathbf{B}$ and, moreover, the values of the function f also belong to the Boolean set $\mathbf{B},$ i.e. $f(x_1, \ldots, x_n) : \mathbf{B}^n \to \mathbf{B},$ where $\mathbf{B}^n = \underbrace{\mathbf{B} \times \ldots \times \mathbf{B}}_{n\ times}.$ Binary set of a length n is called a set $(a_1, \ldots, a_n),$ where $a_i \in \mathbf{B} = \{0,1\},\ \forall i = \overline{1,n}.$ The number of different binary sets having the length n is equal to $2^n,$ i.e. the cardinality (number of all elements) of the set $\mathbf{B}^n$ is equal to 2^n: $\mathrm{card}\left(\mathbf{B}^n\right) = 2^n.$ Hence, various Boolean functions depending on n variables (i.e. various binary sets having the length n) correspond to different binary sets having the length 2^n (i.e. to various Boolean functions depending on n variables). From this verbal proof it follows that the number of all Boolean functions depending on n variables is equal to the number of all binary sets having the length 2^n, i.e. to the cardinality of set $\mathbf{B}^{2^n}$ is equal to 2^{2^n}: $\operatorname{card}\left(\mathbf{B}^{2^n}\right) = 2^{2^n}$.

Now let us list the elementary Boolean functions, the number of which is equal to 11.

- Functions $f(x) = 0$ and $f(x) = 1$ are called 0 and 1 constants, respectively.
- Function $f(x) = x$ is called the identity function.
- Function $f(x) = \bar{x}$ is called the inverting function.
- Function $f(x_1, x_2) = x_1 \vee x_2$ is called the disjunction.
- Function $f(x_1, x_2) = x_1 \cdot x_2$ is called the conjunction.
- Function $f(x_1, x_2) = x_1 \to x_2$ is called the implication. From the foregoing "Basic logical operators" section, it is already known that $f(x_1, x_2) = x_1 \to x_2 = \bar{x}_1 \vee x_2$.
- Function $f(x_1, x_2) = x_1 \leftrightarrow x_2$ is called the equality function. From the foregoing "Basic logical operators" section, it is already known that $f(x_1, x_2) = x_1 \leftrightarrow x_2 = x_1 \cdot x_2 \vee \bar{x}_1 \cdot \bar{x}_2$.
- Function $f(x_1, x_2) = x_1 \oplus x_2$ is called the exclusive disjunction. From the foregoing "Basic logical operators" section, it is already known that $x_1 \oplus x_2 = \overline{x_1 \leftrightarrow x_2}$ and $f(x_1, x_2) = x_1 \oplus x_2 = x_1 \cdot \bar{x}_2 \vee \bar{x}_1 \cdot x_2$.
- Function $f(x_1, x_2) = x_1 \downarrow x_2$ is called the Peirce function. From the foregoing "Basic logical operators" section, it is already known that $f(x_1, x_2) = x_1 \downarrow x_2 = \overline{x_1 \vee x_2}$.
- Function $f(x_1, x_2) = x_1 | x_2$ is called the Sheffer function. From the foregoing "Basic logical operators" section, it is already known that $f(x_1, x_2) = x_1 | x_2 = \overline{x_1 \cdot x_2}$.

In the theory of Boolean functions the Shannon's theorem plays an important role: Any non-zero Boolean function $f(x_1, \ldots, x_n)$ can be represented as

$$f(x_1, \ldots, x_n) = \bigvee_{\substack{(\sigma_1, \ldots, \sigma_n): \\ f(\sigma_1, \ldots, \sigma_n) = 1}} x_1^{\sigma_1} \cdot \ldots \cdot x_n^{\sigma_n},$$

being called Shannon's decomposition, where $x_i^{\sigma_i} = \begin{cases} x_i & if \quad \sigma_i = 0, \\ \bar{x}_i & if \quad \sigma_i = 1 \end{cases}$ is satisfied for $\forall i = \overline{1, n}$.

The right-hand side of the Shannon's decomposition is called the full disjunctive normal form.

In Boolean algebra, a literal is a variable or its negation. A conjunction of literals is called a product term. If each literal is present only once in a product term, then such product term is called an elementary conjunction. If the number of literals in an elementary conjunction equals the number of variables, on which the Boolean function depends, then such elementary conjunction is called a minterm. Obviously, the full disjunctive normal form (i.e. the right-hand side of the Shannon's decomposition) is the disjunction of all minterms of a Boolean function.

The set of Boolean functions is called a functionally complete set, if any Boolean function can be expressed in terms of the functions of this set. For example, each of the sets $\{\bar{x}, x \vee y, x \cdot y\}$, $\{\bar{x}, x \vee y\}$, $\{\bar{x}, x \cdot y\}$, $\{x \downarrow y\}$, $\{x \mid y\}$, $\{\bar{x}, x \to y\}$, $\{0, x \to y\}$, $\{0, x \vee y, x \leftrightarrow y\}$ is a functionally complete set.

The basic properties of Boolean functions. $\bar{1} = 0$; $\bar{0} = 1$; $0 \vee x = x$; $1 \vee x = 1$; $0 \cdot x = 0$; $1 \cdot x = x$; $\bar{\bar{x}} = x$; $x \vee x = x$; $x \cdot x = x$; $x \vee \bar{x} = 1$; $x \cdot \bar{x} = 0$; $x \to 0 = \bar{x}$; $x \to x = 1$; $x \oplus x = 0$; $x \oplus 0 = x$; $x \oplus 1 = \bar{x}$; $x \mid x = \bar{x}$; $x \circ y = y \circ x$, where $\circ$ means any operator of the listed logical operators $\wedge$, $\vee$, $\oplus$, $|$; $(x \circ y) \circ z = x \circ (y \circ z)$, where $\circ$ means any operator of the listed logical operators $\wedge$, $\vee$, $\oplus$; $x \cdot (y \vee z) = x \cdot y \vee x \cdot z$; $x \vee y \cdot z = (x \vee y) \cdot (x \vee z)$; $x \cdot (y \oplus z) = x \cdot y \oplus x \cdot z$; De Morgan's laws: $\overline{x \vee y} = \bar{x} \cdot \bar{y}$ and $\overline{x \cdot y} = \bar{x} \vee \bar{y}$.

Now, we can turn to the formalization of the problem described by words in the introduction, and then apply the selected mathematical apparatus (logical-probabilistic methods) to the already formalized problem. We especially emphasize that the content of the following two sections (Sections 3 and 4) is not limited to the development of a theoretical-mathematical background for diagnosing only cardiovascular diseases: the developed theoretical background is also suitable for logical-probabilistic diagnosis of other categories of diseases.

3. Formulation of the principal direct facts obtained on the basis of scientific theoretical knowledge and clinical experimental medical data, and their formalization in the language of logical algebra

Let n there be a number of major diseases $D \overset{def}{\equiv} \{D_i\}_{i=\overline{1,n}}$, which are typical to the considered geographical region, for the whole country, or even for the economic and / or political confederation of several countries, for example, for the Southern Federal District of Russia, or for Latvia, or for the European Union, etc. Let m there be a number of all sorts of symptoms $S \overset{def}{\equiv} \{S_j\}_{j=\overline{1,m}}$ of the diseases that the physicians of the given geographical region are facing on the basis of long-term medical examinations of thousands of patients.

Let us introduce the following variables and function:

- the variable

$$x_j = \begin{cases} 1, & \text{if the } j\text{-th symptom is detected in the patient;} \\ 0, & \text{if the } j\text{-th symptom is not detected in the patient,} \end{cases} \tag{1}$$

which characterizes the incidence or absence of a j-th $\left(j = \overline{1,m}\right)$ symptom S_j, is denoted by $x = (x_1, \dots, x_m)$;

- the variable

$$y_i = \begin{cases} 1, & \text{if the patient is sick with the } i\text{-th disease;} \\ 0, & \text{if the patient is not sick with the } i\text{-th disease,} \end{cases} \tag{2}$$

which characterizes the incidence or absence of a i-th $\left(i=\overline{1,n}\right)$ disease D_i, is denoted by $y=\left(y_1,\dots,y_n\right)$;

- for $\forall k\in\mathbb{N}$ and $\forall i\in\{1,\dots,k\}$ the selector function $e_i^k\left(z_1,\dots,z_i,\dots,z_k\right)$, which value matches the value of the variable z_i, i.e. $e_i^k\left(z_1,\dots,z_i,\dots,z_k\right)=z_i$, $\forall i\in\{1,\dots,k\}$, $\forall k\in\mathbb{N}$.

Since the variables introduced above are Boolean variables, the sought for links between the main studied diseases and their detected symptoms can be expressed in the language of the theory of logic algebra. Below we will try to construct these links.

Direct knowledge 1. Let us suppose that it is known that the i-th $\left(i=\overline{1,n}\right)$ disease is always accompanied by K_i $\left(1\le K_i\le m\right)$ symptoms $S^{i,K_i}\overset{def}{\equiv}\left\{S_{i_1},\dots,S_{i_{K_i}}\right\}$, where for $\forall j=\overline{1,K_i}$ there are:

- $i_j\in\{1,\dots,m\}$,
- $S_{i_j}\in S$,
- $S_{i_a}\ne S_{i_b}$ at $a\ne b$.

Remark 1. It is easy to notice that from this direct knowledge it follows that among the symptoms S there may be those that accompany two or more different diseases, i.e. possibility $\bigcap_i S^{i,K_i}\ne\varnothing$ is not excluded: for example, chest pain and shortness of breath, as symptoms accompany diseases such as atherosclerotic heart disease, weak heart muscle, endocarditis, valvular heart diseases; another example, prolonged mild pyrexia, as a symptom, accompanies many diseases, such as tuberculosis, viral hepatitis, thyroid gland diseases, toxoplasmosis, thermoneurosis; etc.

In the language of Boolean algebra, the Direct knowledge 1 (i.e. the fact that the i-th $\left(i=\overline{1,n}\right)$ disease is always accompanied by the symptoms S^{i,K_i} $\left(i=\overline{1,n}\right)$) can be formalized in the following form:

$$f_i\left(x,y\right)=1,\ \forall i=\overline{1,n}, \tag{3}$$

where

$$f_i\left(x,y\right)=F_i\left(e_{i_1}^m\left(x_1,\dots,x_m\right),\dots,e_{i_{K_i}}^m\left(x_1,\dots,x_m\right),e_i^n\left(y_1,\dots,y_n\right)\right)\overset{def}{\equiv}y_i\to\bigwedge_{j=1}^{K_i}x_{i_j}. \tag{4}$$

Direct knowledge 2. In the Remark 1 it has been noticed that among the symptoms S there may be those ones that accompany one or more different diseases. Let us suppose that for each symptom there is exhaustive information about what diseases it accompanies: let the symptom S_j $\left(j=\overline{1,m}\right)$ accompany I_j $\left(1\le K_i\le n\right)$ diseases $D^{j,I_j}\overset{def}{\equiv}\left\{D_{j_1},\dots D_{j_{I_j}}\right\}$, where for $\forall i=\overline{1,I_j}$ there are:

- $j_i\in\{1,\dots,n\}$,
- $D_{j_i}\in D$,

- $D_{j_a} \neq D_{j_b}$ for $a \neq b$.

Then this direct knowledge can be formalized as follows:

$$f_{n+j}\left(x,y\right)=1,\ \ j=\overline{1,m}, \tag{5}$$

$$f_{n+j}\left(x,y\right)=F_{n+j}\left(e_j^m\left(x_1,\dots,x_m\right),e_{j_1}^n\left(y_1,\dots,y_n\right),\dots,e_{j_{I_j}}^n\left(y_1,\dots,y_n\right)\right)\overset{def}{\equiv} x_j \to \bigvee_{i=1}^{I_j} y_{j_i}. \tag{6}$$

<u>Direct knowledge 3</u>. Let us suppose that it is known that if a patient has a list of symptoms S^{i,L_i} $\left(1\le i\le n,\ 1\le L_i\le m\right)$, then a disease D_i $\left(1\le i\le n\right)$ necessarily occurs. In this case, we can write that

$$f_{n+m+i}\left(x,y\right)=1,\ i\in\left\{1,\dots,n\right\}, \tag{7}$$

where

$$f_{n+m+i}\left(x,y\right)=F_{n+m+i}\left(e_{i_1}^m\left(x_1,\dots,x_m\right),\dots,e_{i_{L_i}}^m\left(x_1,\dots,x_m\right),e_i^n\left(y_1,\dots,y_n\right)\right)\overset{def}{\equiv}\bigwedge_{j=1}^{L_i} x_{i_j}\to y_i. \tag{8}$$

So, if we denote by k $\left(k\in\left\{1,\dots,n\right\}\right)$ the number of equalities in (7), then the total number of logical equalities in (3), (5) and (7) is equal to $N\left(k\right)=n+m+k$:

$$f_i\left(x,y\right)=1,\ \forall i=\overline{1,N\left(k\right)},\ k\in\left\{1,\dots,n\right\}, \tag{9}$$

the left part of each statement is an important proposition about the cause-effect relation between the n diseases D studied and the m symptoms S.

Further, it is obvious that to the already existing $N\left(k\right)$ logical connections (7) one can add $2^{N(k)}-N\left(k\right)$ more equalities

$$\bigwedge_{\substack{\{i_j\}\uparrow \\ i_j\in\{1,\dots,N(k)\}}} f_{i_j}\left(x,y\right)=1, \tag{10}$$

each of which has a certain semantic interpretation, and some of which, quite possibly, will bring additional information on the cause-effect relation between diseases and their symptoms.

For example, one of the equalities in (10) with $i_j=\overline{1,N\left(k\right)}$ is equality

$$\bigwedge_{i=1}^{N(k)} f_i\left(x,y\right)=1, \tag{11}$$

which can be interpreted as the simultaneous truth of all existing $N\left(k\right)$ propositions (9), and which represents the equation of an implicit function $y=h\left(x\right)$, where $h=\left(h_1,\dots,h_n\right)$, moreover $\bigwedge_{i=1}^{N(k)} f_i\left(x,h_1\left(x\right),\dots,h_n\left(x\right)\right)\equiv 1$. In other words, equality (11) means the simultaneous fulfillment of assumptions (3), (5), (7), and this equality defines the implicit formula for the dependence of diseases D on symptoms S. Therefore, having (9) and (10), it is necessary to determine the explicit form of the dependence of the disease on the symptoms. As it has been noted above, some of the equalities in (10) may appear a tautology, i.e. identically true statements invariant with respect to the values of their components (in other words, "empty") and, therefore, will not bring any additional useful information to the already existing (9).

Let us highlight once again that equations (10), (11) were obtained only under the three assumptions made (see the Direct knowledge 1-3), which are based on both theoretical scientific and (mostly) clinical and experimental medical knowledge acquired by medical researchers and medical practitioners. Therefore, it could be noted that equations (10), (11) represent some interrelated set of experimentally and / or theoretically confirmed facts, describing the principal laws, properties and relations between objects of two types - various diseases D and symptoms S. In other words, the resulting equations (10), (11) carries more global-scale information about the cause-effect relations between diseases and symptoms in general, rather than particular symptoms found in the case of some specific patient.

Remark 2. The Direct knowledge 3 is a strong enough assumption: it is not always possible to list the accompanying symptoms for each disease, by which it could be unambiguously asserted that the patient has this particular disease and not some others. In other words, not every disease has a known set of sufficient symptoms so that formulas (7), (8) can be written and used with complete confidence. For this reason we will not take the Direct knowledge 3 into account in this paper, in particular, in the examples that will be discussed later, starting from here throughout the study we will assume that $S^{i,L_i}=\varnothing$ for $\forall i=\overline{1,n}$.

4. The logical-probabilistic approach for diagnosing the incidence of a disease (or diseases) in a patient according to the symptoms detected at a given time

Let us return again to the functional equation (11), the left side of which, as Boolean function of $N(k)$ variables, is represented by a truth table with $2^{N(k)}$ rows. Having equation (11), which, taking into account the Remark 2, has the form

$$\bigwedge_{i=1}^{N(0)} f_i(x,y)=1,\ \ N(0)=n+m, \tag{12}$$

we will try to diagnose the incidence of a disease (or diseases) in a particular patient according to the previously disclosed symptoms.

Let us suppose that during the medical check, the patient Y had q $(1\le q\le m)$ symptoms $S^q \overset{def}{\equiv} \{S_{t_1},\ldots,S_{t_q}\}$, where for $\forall j=\overline{1,q}$ the following is valid

- $t_j\in\{1,\ldots,m\}$,
- $S_{t_j}\in\{S_1,\ldots,S_m\}$,
- $S_{t_a}\ne S_{t_b}$ for $a\ne b$.

Since $S\cap S^q\ne\varnothing$, then it is natural and goes without saying that part of the symptoms S^q found in a patient Y are contained in at least one of the sets of symptoms S^{i,K_i} $i\in\{1,\ldots,n\}$, accompanying the disease D_i $i\in\{1,\ldots,n\}$. In other words, exactly the q_i $\left(0\le q_i\le \min\{K_i,q\},\ \forall i=\overline{1,n}\right)$ number of symptoms from the set of symptoms S^q, found in a patient Y, belong to the set S^{i,K_i}, moreover the equality is fulfilled

$$q=\left|\sum_{i=1}^{n} q_i-\operatorname{card}\left(S^{\text{Pairwise}\cap}\right)\right|, \tag{13}$$

where the set $S^{\text{Pairwise}\cap} \overset{def}{\equiv} \left\{ S^{q_i} \bigcap S^{q_j} \right\}_{i<j,\, i=\overline{1,n},\, j=\overline{1,n}}$ consists of the symptoms from a set S^q, and each of the symptoms accompanies more than one disease. It is important to note that the set S^q may appear to be such that $S^q \not\subset S^{i,K_i}$ for $\forall i=\overline{1,n}$, in other words, it is quite possible that not all the symptoms found in a patient Y, match or are part of the symptoms accompanying any disease from the set D of underlying principal diseases in question: in example 1, which will be discussed later, four patients are examined, and only for one patient (the second patient) there are detected symptoms being part of the symptoms that always accompany one of the diseases (the second) out of the three considered diseases, and the remaining three patients have symptoms from a set, which is not a subset of any of the two sets of symptoms accompanying the other two diseases, respectively.

So, for the logical-probabilistic diagnosis of the incidence of diseases in a patient who has q symptoms $S^q \subseteq S$, as an input data we have the set $S = \left\{ S_j \right\}_{j=\overline{1,m}}$ and m number; the set $D = \left\{ D_i \right\}_{i=\overline{1,n}}$ and n number; the functional logical equation (12); the set S^q and q number; $q_i \left(0 \le q_i \le \min \left\{ K_i, q \right\},\, \forall i=\overline{1,n} \right)$ numbers, where $q = \left| \sum_{i=1}^{n} q_i - \text{card}\left(\left\{ S^{q_i} \bigcap S^{q_j} \right\}_{i<j,\, i=\overline{1,n},\, j=\overline{1,n}} \right) \right|$. Staying limited only to this input data, the problem is to find the probability of incidence for each disease from the list of diseases $D = \left\{ D_i \right\}_{i=\overline{1,n}}$ (the list of diseases D can be compiled based on some specific criterion, which is currently preferred for some reason; for example, in this paper the "highest prevalence rate" has been chosen as the criterion for selecting principal diseases in the studied geographical area – in a separate locality, in a region, in a city, in a country, etc.). The following procedure for solving the stated problem is proposed.

Step 1. The left side of the functional equation (12), as a Boolean function depending on $m+n$ variables, is minimized (in general case, taking into account the Direct knowledge 3, this function depends on $N(k)$ variables: see Remark 2). For this purpose, there could be used any minimization algorithm. The most popular algorithms are regarded to be the algebraic methods (Blake method, etc.), the Karnaugh maps, and Quine-McCluskey algorithm (for instance, see Karazyan, 1981; Rosen, 2012) and respective references given there). The resulting function is denoted as $f_{\min}(x,y)$. The function $f_{\min}(x,y)$ will be a disjunction of elementary conjunctions, some or all of which may be a minterm. Let us denote by Q the number of elementary conjunctions of the function $f_{\min}(x,y)$.

Step 2. For the function $f_{\min}(x,y)$, an extended diagnostic table is constructed, which consists of:

- from a table whose work space contains Q rows (each row corresponds to one elementary conjunction of a function $f_{\min}(x,y)$) and $m+n$ columns (the number of function variables on which the function $f_{\min}(x,y)$ depends); each row of this table contains the binary code of the corresponding elementary conjunction (if the variable has logical negation, then it is put 0, otherwise it is put 1);

- from a column whose work space contains Q rows; this column is an informative column containing information about a set of symptoms S^q (it is recommended to allocate this column at the far right in the extended diagnostic table), namely, it is necessary to put a "+" sign in those rows of this column where the values of variables x_i, correspond to the discovered symptoms, take the value 1; such rows of this informative column will be called positive rows; the set of positive rows numbers in the informative column will be denoted by X, and the cardinality of this set $\text{card}(X)$ will be calculated (i.e. the number of "+" characters in the informative column).

It is important to emphasize that it is possible that some cells of the diagnostic table for the function $f_{\min}(x, y)$ (i.e. only cells of those columns in the extended diagnostic table that correspond to variables $x_i, \forall i = \overline{1, m}$ and $y_j, \forall j = \overline{1, n},$ but not the cells of an entire extended diagnostic table) remain empty (it is recommended to fill-in such cells with dash sign "–"). This means that elementary conjunctions are present in the expression of $f_{\min}(x, y)$, which are not a minterm because the presence condition for all literals is violated. If these dashes (i.e. empty cells) are located in cells that correspond to variables $y_j, j = \{1, \dots, n\}$, then it is necessary to assume that the corresponding variable can take both the value 0 and 1. This is important and should not be ignored.

<u>Step 3</u>. A new table is constructed – a table of possible patient diseases. To achieve that, all the columns that correspond to variables $x_i, \forall i = \overline{1, m}$ are removed from the extended diagnostic table. As a result, a table whose number of columns is $n+1$ will be obtained: the table will contain columns corresponding to the variables $y_j, \forall j = \overline{1, n},$ and one informative column, in all cells of which a "+" sign is present (let us call this column as the patient column). Further, in the resulting table it is necessary instead of all the dashes "–" to insert $\{0;1\}$ as it has been described in the Step 2. This means that the corresponding variable $y_j, j = \{1, \dots, n\}$ can take both the value 0 and 1. Finally, a new column is added to the resulting table on the far right containing Q rows in which their corresponding numbers are written in the extended diagnostic table.

<u>Step 4</u>. The probability concept for the incidence of the disease $D_i, i = \{1, \dots, n\}$ in a patient as well as the concept of the conditionality measure of the symptoms S^q for each of the diseases $D = \{D_i\}_{i=\overline{1,n}}$ found in the patient are introduced based on the data of the k-th positive row from the table of the patient's possible diseases: $p_i(k) \in [0,1], \forall k \in X$ and $\omega_i(k) \in [0,1], \forall k \in X$, where

$$\sum_{i=1}^{n} \omega_i(k) \cdot p_i(k) = 1 \text{ for } \forall k \in X, \tag{14}$$

$$\sum_{i=1}^{n}\omega_i(k)=1 \text{ and } \sum_{i=1}^{n}p_i(k)=1 \text{ for } \forall k\in X. \tag{15}$$

Equalities (14), (15) interrelate the introduced concepts of $\omega_i(k)$ and $p_i(k)$. For $n=2$ it is easily possible to see and understand the essence of this relationship: $p_1(k)=\dfrac{\omega_1(k)}{2\cdot\omega_1(k)-1}$, $p_2(k)=\dfrac{\omega_1(k)-1}{2\cdot\omega_1(k)-1}$, $\omega_2(k)=1-\omega_2(k)$. It is possible to interpret somewhat differently the $\omega_i(k)$, namely, $\omega_i(k)$ is a measure of the causation of the disease D_i in the occurrence of the symptoms S^q, that were found in the patient. The concepts introduced are the sought for parameters and their values, in general case, should not be set arbitrarily or by guesswork. By means of applying approaches that are based on the principles of ill-posed and inverse problems, and which were proposed in Natrins et al. (2015), Guseynov, (2017), Guseynov, (2015a), Guseynov, (2015b) investigated and successfully applied both in the field of technical diagnostics and in the field of financial and economic forecasting, one can determine the values of the parameters $\omega_i(k),\ \forall i=\overline{1,n}$ based only on the initial data – the set $S=\{S_j\}_{j=\overline{1,m}}$ and m number; the set $D=\{D_i\}_{i=\overline{1,n}}$ and n number; the functional equation (12); the set S^q and q number; $q_i\ \left(0\le q_i\le\min\{K_i,q\},\ \forall i=\overline{1,n}\right)$ numbers. This could be done without making any additional assumptions, which, as a rule, at least significantly limit the scope of the developed approaches. In the absence of additional information of a quantitative or qualitative nature, the assumption $\omega_i(k)\equiv const.,\ \forall i=\overline{1,n}$ is unsubstantiated (it may be either correct or false), and the essence of this assumption is that we assert that the disease in the patient is of equal cause when detected symptoms appear. In this paper, we will also assume that the values of the introduced parameters $\omega_i(k),\ \forall i=\overline{1,n}$ are the same. Then, without loss of generality, it is possible to state that instead of (14) there is an equality

$$\sum_{i=1}^{n}p_i(k)=1 \text{ for } \forall k\in X. \tag{16}$$

If the doctor does not have any additional substantial information, other than the symptoms detected at a given time, then it is quite natural to assume that the events "the patient has the disease $D_i,\ i=\overline{1,n}$" are equally probable events under the conditions that: (a) for $\forall i=\overline{1,n}$ the table of possible diseases $y_i=1$; (b) in the extended diagnostic table, the informative column contains only one "+" sign. Therefore, the introduced probabilities $\{p_i(k)\}_{i=\overline{1,n},\,k\in X}$ are found by the formula

$$p_i(k)=\begin{cases}0 & if\quad y_i=0,\\ \dfrac{\lambda}{2} & if\quad y_i=\{0,1\},\ \left(\text{i.e. the corresponding cell of table is empty (has dash sign "}-\text{")}\right)\\ \lambda & if\quad y_i=1,\end{cases} \tag{17}$$

where λ is unambiguously determined from the equality (16) (see examples of the Section 5).

Step 5. The concept of the likelihood of the incidence of the disease $D_i,\ i=\{1,\ldots,n\}$ in patient is introduced based on information across all the positive rows of the patient's possible disease table:

$$p(D_i)=\frac{\sum_{k\in X} p_i(k)}{\text{card}(X)}, \tag{18}$$

where the condition $\sum_{i=1}^{n} p(D_i)=1$ should be met.

5. Application examples

Example 1 (Complete investigation). Let us suppose that three diseases $D=\{D_i\}_{i=\overline{1,3}}$ based on the analysis of seven symptoms $S=\{S_j\}_{j=\overline{1,7}}$ are investigated. Then we can introduce Boolean variables $x=(x_1,\ldots,x_7)$ and $y=(y_1,y_2,y_3)$, which essence is explained above in (1), (2). Further, let it be known that

- the disease D_1 is always accompanied by symptoms S_1, S_2 and S_6, i.e. it could be written
$$S^{1,K_1=3}=\{S_{1_1},S_{1_2},S_{1_3}\}=\{S_1,S_2,S_6\}; \tag{19}$$
- the disease D_2 is always accompanied by symptoms S_1, S_3, S_4 and S_7, i.e. it could be written
$$S^{2,K_2=4}=\{S_{2_1},S_{2_2},S_{2_3},S_{2_4}\}=\{S_1,S_3,S_4,S_7\}; \tag{20}$$
- the disease D_3 is always accompanied by symptoms S_2, S_3, S_5, S_6 and S_7, i.e. it could be written
$$S^{3,K_3=5}=\{S_{3_1},S_{3_2},S_{3_3},S_{3_4},S_{3_5}\}=\{S_2,S_3,S_5,S_6,S_7\}. \tag{21}$$

Then, by virtue of (4), it is possible to write that

$$\begin{cases} f_1(x,y)=F_1\left(e_1^7(x),e_2^7(x),e_6^7(x),e_1^3(y)\right)=y_1\to x_1\cdot x_2\cdot x_6; \\ f_2(x,y)=F_3\left(e_1^7(x),e_3^7(x),e_4^7(x),e_7^7(x),e_2^3(y)\right)=y_2\to x_1\cdot x_3\cdot x_4\cdot x_7; \\ f_3(x,y)=F_3\left(e_2^7(x),e_3^7(x),e_5^7(x),e_6^7(x),e_7^7(x),e_3^3(y)\right)=y_3\to x_2\cdot x_3\cdot x_5\cdot x_6\cdot x_7. \end{cases} \tag{22}$$

From (19)-(21) it follows that

- the symptom S_1 accompanies the diseases D_1 and D_2, i.e. we will have $D^{1,I_1=2}=\{D_{1_1},D_{1_2}\}=\{D_1,D_2\}$;
- the symptom S_2 accompanies the diseases D_1 and D_3, i.e. we will have $D^{2,I_2=2}=\{D_{2_1},D_{2_2}\}=\{D_1,D_3\}$;
- the symptom S_3 accompanies the diseases D_2 and D_3, i.e. we will have $D^{3,I_3=2}=\{D_{3_1},D_{3_2}\}=\{D_2,D_3\}$;

- the symptom S_4 accompanies only the disease D_2, i.e. we will have $D^{4,I_4=1}=\{D_{4_1}\}=\{D_2\}$;
- the symptom S_5 accompanies only the disease D_3, i.e. we will have $D^{5,I_5=1}=\{D_{5_1}\}=\{D_3\}$;
- the symptom S_6, as well as symptom S_2, accompanies only the diseases D_1 and D_3, i.e. we will have $D^{6,I_6=2}=\{D_{6_1},D_{6_2}\}=\{D_1,D_3\}$;
- finally, the symptom S_7, as well as symptom S_3, accompanies only the diseases D_2 and D_3, i.e. we will have $D^{7,I_7=2}=\{D_{7_1},D_{7_2}\}=\{D_2,D_3\}$.

Then, by virtue of (6), it is possible to write that

$$\begin{cases} f_4(x,y)=F_4\left(e_1^7(x),e_1^3(y),e_2^3(y)\right)=x_1\to(y_1\vee y_2); \\ f_5(x,y)=F_5\left(e_2^7(x),e_1^3(y),e_3^3(y)\right)=x_2\to(y_1\vee y_3); \\ f_6(x,y)=F_6\left(e_3^7(x),e_2^3(y),e_3^3(y)\right)=x_3\to(y_2\vee y_3); \\ f_7(x,y)=F_7\left(e_4^7(x),e_2^3(y)\right)=x_4\to y_2; \\ f_8(x,y)=F_8\left(e_5^7(x),e_3^3(y)\right)=x_5\to y_3; \\ f_9(x,y)=F_9\left(e_6^7(x),e_1^3(y),e_3^3(y)\right)=x_6\to(y_1\vee y_3); \\ f_{10}(x,y)=F_{10}\left(e_7^7(x),e_2^3(y),e_3^3(y)\right)=x_7\to(y_2\vee y_3). \end{cases} \tag{23}$$

Taking into account (22) and (23) in (11), we will receive $\bigwedge_{i=1}^{10} f_i(x,y)=1,$ i.e.

$$\left.\begin{aligned} &(y_1\to x_1\cdot x_2\cdot x_6)\cdot(y_2\to x_1\cdot x_3\cdot x_4\cdot x_7)\cdot(y_3\to x_2\cdot x_3\cdot x_5\cdot x_6\cdot x_7)\cdot \\ &\cdot(x_1\to(y_1\vee y_2))\cdot(x_2\to(y_1\vee y_3))\cdot(x_3\to(y_2\vee y_3))\cdot(x_4\to y_2)\cdot \\ &\cdot(x_5\to y_3)\cdot(x_6\to(y_1\vee y_3))\cdot(x_7\to(y_2\vee y_3))=1. \end{aligned}\right\}$$

By simplifying the left side of the last equation, we will get

$$\begin{aligned} f_{\min}(x,y)=&\bar{x}_1\cdot\bar{x}_2\cdot\bar{x}_3\cdot\bar{x}_4\cdot\bar{x}_5\cdot\bar{x}_6\cdot\bar{x}_7\cdot\bar{y}_1\cdot\bar{y}_2\cdot\bar{y}_3\vee x_1\cdot x_2\cdot x_3\cdot x_4\cdot x_5\cdot x_6\cdot x_7\cdot y_2\cdot y_3\vee \\ &\vee x_1\cdot x_2\cdot x_3\cdot x_4\cdot\bar{x}_5\cdot x_6\cdot x_7\cdot y_1\cdot y_2\cdot\bar{y}_3\vee x_1\cdot x_2\cdot x_3\cdot\bar{x}_4\cdot x_5\cdot x_6\cdot x_7\cdot y_1\cdot\bar{y}_2\cdot y_3\vee \\ &\vee x_1\cdot\bar{x}_2\cdot x_3\cdot x_4\cdot\bar{x}_5\cdot\bar{x}_6\cdot x_7\cdot\bar{y}_1\cdot y_2\cdot\bar{y}_3\vee\bar{x}_1\cdot x_2\cdot x_3\cdot\bar{x}_4\cdot x_5\cdot x_6\cdot x_7\cdot\bar{y}_1\cdot\bar{y}_2\cdot y_3\vee \\ &\vee x_1\cdot x_2\cdot\bar{x}_3\cdot\bar{x}_4\cdot\bar{x}_5\cdot x_6\cdot\bar{x}_7\cdot y_1\cdot\bar{y}_2\cdot\bar{y}_3=1. \end{aligned} \tag{24}$$

Equation (24) is an implicit formulation of the dependence of the studied diseases D on the symptoms S. It is important to note that the functional equation (24) has been obtained only under the two assumptions (see Assumption 1 and 2), which, as it has been mentioned above (see the last paragraph before Remark 2), are based on both theoretical scientific and (mostly) clinical and experimental medical knowledge acquired by medical researchers and medical practitioners. Therefore, using this particular problem as an example, it is possible to state that the functional equation (24) represents some interconnected set of experimentally and / or theoretically confirmed facts describing the principal laws, properties and relations for the objects of two types – all sorts of diseases and their symptoms. In other words, in the frames of the considered example, the obtained equation (24) carries more global-scale information about the cause-effect relation between diseases and symptoms in general, rather than particular symptoms found in the case of some specific patient. Having equation (24), we will try to diagnose the incidence of a disease (or diseases) in a particular patient. Let us suppose that during the medical check of four patients there has been revealed that:

- the first patient has symptoms S_1, S_2 and S_3;
- the second patient has symptoms S_1, S_3 and S_4;
- the third patient has symptoms S_1, S_2, S_5 and S_7;
- the fourth patient has symptoms S_3, S_4 and S_6.

Thus, we assemble the initial information, on which basis the procedure of logical-probabilistic diagnosis of diseases in patients described in the Section 4 will be applied. To recall that the source data is formed from:

– the set $S = \{S_j\}_{j=\overline{1,m}}$ and number m;

– the set $D = \{D_i\}_{i=\overline{1,n}}$ and number n;

– the functional equation (12);

– the set S^q and number q;

– q_i $\left(0 \le q_i \le \min\{K_i, q\},\ \forall i = \overline{1,n}\right)$ numbers, where equality (13) is fulfilled.

Since the set $S = \{S_j\}_{j=\overline{1,7}}$ is known, hence, we know the number $m = 7$. Since the set $D = \{D_i\}_{i=\overline{1,3}}$ is known, hence, we know the number $n = 3$. Equation (12) has been derived already. For each of the four examined patients, the set S^q is also known: for the first patient we have information about the discovered symptoms $S^q = \{S_1, S_2, S_3\}$ and, therefore, we know the number $q = 3$; for the second patient we have information about the discovered symptoms $S^q = \{S_1, S_3, S_4\}$ and, therefore, we know the number $q = 3$; for the third patient we have information about the discovered symptoms $S^q = \{S_1, S_2, S_5, S_7\}$ and, therefore, we know the number $q = 4$; for the fourth patient we have information about the discovered symptoms $S^q = \{S_3, S_4, S_6\}$ and, therefore, we know the number $q = 3$. For the complete assembling of initial data (being also called the source data or input information) it is just necessary to calculate the values q_i $\left(0 \le q_i \le \min\{K_i, q\},\ \forall i = \overline{1,n}\right)$ for each patient. Let us find these numbers for each patient.

For the first patient we have: $S^q=\{S_1,S_2,S_3\}$ and $q=3$. Therefore, we can write that $S^{q_1}=\underbrace{\{S_1,S_2\}}_{\subset S^q}\subset S^{1,3}$, $S^{q_2}=\underbrace{\{S_1,S_3\}}_{\subset S^q}\subset S^{2,4}$, $S^{q_3}=\underbrace{\{S_2,S_3\}}_{\subset S^q}\subset S^{3,5}$, from which it follows that $q_1=2,\ q_2=2,\ q_3=2.$ Let us validate the formula (13):

$$3=q=\left|\sum_{i=1}^{3}q_i-\mathrm{card}\left(\left\{S^{q_i}\cap S^{q_j}\right\}_{i<j,\ i=\overline{1,3},\ j=\overline{1,3}}\right)\right|=\left|(2+2+2)-\mathrm{card}(S_1,S_2,S_3)\right|=6-3=3.$$

For the second patient we have: $S^q=\{S_1,S_3,S_4\}$ and $q=3$. Therefore, we can write that $S^{q_1}=\underbrace{\{S_1\}}_{\subset S^q}\subset S^{1,3}$, $S^{q_2}=\underbrace{\{S_1,S_3,S_4\}}_{\subset S^q}\subset S^{2,4}$, $S^{q_3}=\underbrace{\{S_3\}}_{\subset S^q}\subset S^{3,5}$, from which it follows that $q_1=1,\ q_2=3,\ q_3=1.$ Let us validate the formula (13):

$$3=q=\left|\sum_{i=1}^{3}q_i-\mathrm{card}\left(\left\{S^{q_i}\cap S^{q_j}\right\}_{i<j,\ i=\overline{1,3},\ j=\overline{1,3}}\right)\right|=\left|(1+3+1)-\mathrm{card}(S_1,\varnothing,S_3)\right|=5-2=3.$$

For the third patient we have: $S^q=\{S_1,S_2,S_5,S_7\}$ and $q=4$. Therefore, we can write that $S^{q_1}=\underbrace{\{S_1,S_2\}}_{\subset S^q}\subset S^{1,3}$, $S^{q_2}=\underbrace{\{S_1,S_7\}}_{\subset S^q}\subset S^{2,4}$, $S^{q_3}=\underbrace{\{S_2,S_5,S_7\}}_{\subset S^q}\subset S^{3,5}$, from which it follows that $q_1=2,\ q_2=2,\ q_3=3.$ Let us validate the formula (13):

$$4=q=\left|\sum_{i=1}^{3}q_i-\mathrm{card}\left(\left\{S^{q_i}\cap S^{q_j}\right\}_{i<j,\ i=\overline{1,3},\ j=\overline{1,3}}\right)\right|=\left|(2+2+3)-\mathrm{card}(S_1,S_2,S_7)\right|=7-3=4.$$

Finally, for the fourth patient we have: $S^q=\{S_3,S_4,S_6\}$ and $q=3$. Therefore, we can write that $S^{q_1}=\underbrace{\{S_6\}}_{\subset S^q}\subset S^{1,3}$, $S^{q_2}=\underbrace{\{S_3,S_4\}}_{\subset S^q}\subset S^{2,4}$, $S^{q_3}=\underbrace{\{S_3,S_6\}}_{\subset S^q}\subset S^{3,5}$, from which it follows that $q_1=1,\ q_2=2,\ q_3=2.$ Let us validate the formula (13):

$$3=q=\left|\sum_{i=1}^{3}q_i-\mathrm{card}\left(\left\{S^{q_i}\cap S^{q_j}\right\}_{i<j,\ i=\overline{1,3},\ j=\overline{1,3}}\right)\right|=\left|(1+2+2)-\mathrm{card}(\varnothing,S_6,S_3)\right|=5-2=3.$$

Thus, we have the necessary source data to apply the procedure of logical-probabilistic diagnosis of diseases proposed in the Section 4 for the cases of four patients. Step 1 of this procedure has already been completed: the function $f_{\min}(x,y)$ on the right side of the functional equation (24) is the sought for minimal disjunctive normal form of the function (minimization was carried out manually by the Blake method), which is located on the left side of the functional equation (12). Let us proceed to implementation of Step 2. Applying this step to the initial data of each of the four patients allows us to build an extended diagnostic table for each patient, the work space of which consists of 7 rows (the number of elementary conjunctions in the function $f_{\min}(x,y)$ expression from the functional equation (24)) and of 11 columns, where the first 10 columns correspond to the $10=\underbrace{m}_{=10}+\underbrace{n}_{=3}$ variables on which the function $f_{\min}(x,y)$ depends, and the last 11th column is an informative column about the symptoms that have been discovered in the currently examined patient by the present moment. As a result, we will have four extended diagnostic tables, where the first 10 columns and 7 rows completely coincide. By combining these four extended diagnostic tables into single one, we will obtain one extended diagnostic table for all four examined patients.

Table 1. Extended table for establishing diagnosis

x_1	x_2	x_3	x_4	x_5	x_6	x_7	y_1	y_2	y_3	Patient #1	Patient #2	Patient #3	Patient #4
0	0	0	0	0	0	0	0	0	0				
1	1	1	1	1	1	1	–	1	1	+	+	+	+
1	1	1	1	0	1	1	1	1	0	+	+		+
1	1	1	0	1	1	1	1	0	1	+		+	
1	0	1	1	0	0	1	0	1	0		+		
0	1	1	0	1	1	1	0	0	1				
1	1	0	0	0	1	0	1	0	0				

Remark 3. Brief description of informational columns structure in the Table 1. Columns 11-14 of Tab. 1 are informative, and they refer to the respective four patients. As it has been stated in Step 2 of the Section 4, the "+" sign in the rows of the informative columns indicates that the patient may have those diseases D_i, which corresponding variables y_i possess the value 1. For example, let us consider the 12th column, which refers to the second patient: in this column, the "+" sign is in the second, third and fifth rows, because only in these three rows the values of variables x_1, x_3 and x_4 (they correspond to the symptoms S_1, S_3 and S_4, respectively) are equal to 1. Therefore, the second patient with certain probabilities (generally speaking, with different probabilities; see formulas (16), (17)) may have the diseases D_1, and / or D_2, and / or D_3. The remaining informative columns (i.e. columns 11, 13 and 14) of a Table 1 are compiled and read in a similar way. Let us pay attention to the eighth column of the Table 1, which corresponds to the variable y_1. The second row of this column contains a "–" (a dash sign), and as it has been mentioned in Steps 2 and 3 from the Section 4, this means that the variable y_1 may possess both the value 0 and 1.

It should be noted that everywhere below there will be added a new index j, meaning the number (or, alternatively, the identifier) of the current patient from the rest of the examined patients, in the following designations from the Section 4: instead of the $p_i(k)$ notation, we will use the $p_i^j(k)$ notation, instead of $p(D_i)$ we will use the $p^j(D_i)$ notation, instead of X we will use the X^j notation, instead of the S^q notation we will use the S^{q^j} notation, where in all the new designations the index j identifies the patient for whom the logical-probabilistic diagnostics process is underway.

Now, using the Table 1, we will try to diagnose the disease in each of the four patients. For this purpose, it is necessary to carry out Steps 3-5, described in the Section 4, consequently for each of the four examined patients. Let us start with the first patient.

Performing Step 3 of the logical-probabilistic diagnosis for the incidence of diseases in a patient described in the Section 4 allows us to construct the following table of possible diseases for the first patient:

Table 2. Table of possible diseases for the first patient

y_1	y_2	y_3	Patient #1	k $\left(\text{card}\left(X^1\right)=3\right)$
$\{0;1\}$	1	1	+	2
1	1	0	+	3
1	0	1	+	4

Using the data from a first row (or rather the row, where $k=2$) from the Table 2 in the formula (17) we will receive that $p_1^1(2)=\frac{p_2^1(2)}{2}$, $p_2^1(2)=p_3^1(2)$. Then, by virtue of (16), we have

$$p_1^1(2)=\frac{1}{5},\ p_2^1(2)=\frac{2}{5},\ p_3^1(2)=\frac{2}{5}. \tag{25}$$

Similarly, taking second row data $(k=3)$ from the Table 2 into account in (17), gives us the following results:

$$p_1^1(3)=\frac{1}{2},\ p_2^1(3)=\frac{1}{2},\ p_3^1(3)=0. \tag{26}$$

Taking third row data $(k=4)$ from the Table 2 into account in (17), gives us the following results:

$$p_1^1(4)=\frac{1}{2},\ p_2^1(4)=0,\ p_3^1(4)=\frac{1}{2}. \tag{27}$$

Considering (25)-(27) in formula (18), gives us the following results:

$$p^1(D_1)=\frac{2}{5},\ p^1(D_2)=\frac{3}{10},\ p^1(D_3)=\frac{3}{10}. \tag{28}$$

From (28) it follows that the probability of the fact that the first patient with symptoms $S^{q^1=3}=\{S_1,S_2,S_3\}$ has a disease D_1 is approximately 1.3 times greater than the probability of having diseases D_2 or D_3.

Now let us complete the Steps 3-5 for the second patient. Performing Step 3 allows to construct the following table of possible diseases for the second patient:

Table 3. Table of possible diseases for the second patient

y_1	y_2	y_3	Patient #2	k $\left(\text{card}\left(X^2\right)=3\right)$
$\{0;1\}$	1	1	+	2
1	1	0	+	3
0	1	0	+	5

In a manner similar to how the diagnosing of the diseases has been carried out for the first patient, let us extract and use the data from the Table 3 in the formula (17).

It is relatively easy to make sure that

$p_1^2(k=2)=\frac{1}{5}, \quad p_2^2(k=2)=\frac{2}{5}, \quad p_3^2(k=2)=\frac{2}{5}, \quad p_1^2(k=3)=\frac{1}{2}, \quad p_2^2(k=3)=\frac{1}{2},$
$p_3^2(k=3)=0, \; p_1^2(k=5)=0, \; p_2^2(k=5)=1, \; p_3^2(k=5)=0.$

Taking into account those obtained results in (18), we will receive:

$$p^2(D_1)=\frac{7}{30}, \; p^2(D_2)=\frac{19}{30}, \; p^2(D_3)=\frac{2}{15},$$

from which it follows that the probability indicating that the second patient with symptoms $S^{q^2=3}=\{S_1,S_3,S_4\}$ has a disease D_2 is approximately 2.7 times greater than the probability of having the disease D_1, and exactly 4.75 times more than the probability of having the disease D_3.

Now let us complete the Steps 3-5 for the third patient. Performing Step 3 allows to construct the following table of possible diseases for the third patient:

Table 4. Table of possible diseases for the third patient

y_1	y_2	y_3	Patient #3	$k \; (\text{card}(X^3)=2)$
$\{0;1\}$	1	1	+	2
1	0	1	+	4

Relevant consideration of data from Table 4 in formula (17) gives us the following results:

$p_1^3(k=2)=\frac{1}{5}, \quad p_2^3(k=2)=\frac{2}{5}, \quad p_3^3(k=2)=\frac{2}{5}, \quad p_1^3(k=4)=\frac{1}{2}, \quad p_2^3(k=4)=0,$
$p_3^3(k=4)=\frac{1}{2}.$

Taking into account those results in (18), we will have:

$$p^3(D_1)=\frac{7}{20}, \; p^3(D_2)=\frac{1}{5}, \; p^3(D_3)=\frac{9}{20},$$

from which it follows that the probability indicating that a third patient with symptoms $S^{q^3=4}=\{S_1,S_2,S_5,S_7\}$ has a disease D_3 is approximately 1.3 times greater than the probability of having the disease D_1, and exactly 2.25 times greater than the probability of having the disease D_2.

Finally let us complete the Steps 3-5 for the fourth patient. Performing Step 3 allows to construct the following table of possible diseases for the fourth patient:

Table 5. Table of possible diseases for the fourth patient

y_1	y_2	y_3	Patient #4	$k \; (\text{card}(X^4)=2)$
$\{0;1\}$	1	1	+	2
1	1	0	+	3

Again, taking into account data from the Table 5 in formula (18), we will receive:

$p_1^4(k=2)=\frac{1}{5},\quad p_2^4(k=2)=\frac{2}{5},\quad p_3^4(k=2)=\frac{2}{5},\quad p_1^4(k=3)=\frac{1}{2},\quad p_2^4(k=3)=\frac{1}{2},$ $p_3^4(k=3)=0.$

Substituting these results into formula (18), we will have:

$$p^4(D_1)=\frac{7}{20},\ p^4(D_2)=\frac{9}{20},\ p^4(D_3)=\frac{1}{5},$$

from which it follows that the probability indicating that a fourth patient with symptoms $S^{q^4=3}=\{S_3,S_4,S_6\}$ has a disease D_2 is approximately 1.3 times greater than the probability of having the disease D_1, and exactly 2.25 times greater than the probability of having the disease D_3.

The obtained results can be combined into the single table, which is convenient for reading:

Table 6. Probabilistic table of patients' diseases

Patient	Detected symptoms	The probability of disease			$\sum_{i=1}^{3} p(D_i)$
		$p(D_1)$	$p(D_2)$	$p(D_3)$	
Patient #1	$\{S_1,S_2,S_3\}$	0.4	0.3	0.3	1
Patient #2	$\{S_1,S_3,S_4\}$	0.23(3)	0.63(3)	0.13(3)	1
Patient #3	$\{S_1,S_2,S_5,S_7\}$	0.35	0.2	0.45	1
Patient #4	$\{S_3,S_4,S_6\}$	0.35	0.45	0.2	1

As it can be seen from the Table 6, according to the symptoms detected by this moment, none of the four examined patients can be diagnosed with complete certainty (i.e. with probability equal to 1) that he/she has one or another disease out of three D_1, D_2, D_3. The reason for this is not, as it looks at first sight, because the symptoms found $S^{q^1=3}=\{S_1,S_2,S_3\}$ (in the first patient), $S^{q^2=3}=\{S_1,S_3,S_4\}$ (in the second patient), $S^{q^3=4}=\{S_1,S_2,S_5,S_7\}$ (in the third patient), $S^{q^4=3}=\{S_3,S_4,S_6\}$ (in the fourth patient) do not coincide with any of the clinically-proven clinical sets of symptoms $S^{1,K_1=3}=\{S_1,S_2,S_6\}$, $S^{2,K_2=4}=\{S_1,S_3,S_4,S_7\}$, accompanying the diseases D_1, D_2, D_3, respectively. Indeed, let us assume that the new patient has symptoms $\{S_1,S_2,S_6\}$, which fully coincide with the theoretical-clinical set of symptoms $S^{1,K_1=3}$. Then, as it follows from the Table 1, the corresponding table of possible diseases of this patient is:

Table 7. Table of possible diseases for the new patient

y_1	y_2	y_3	New patient	k $\left(\text{card}\left(X^{\text{New patient}}\right)=4\right)$
$\{0;1\}$	1	1	+	2
1	1	0	+	3
0	1	0	+	4
1	0	0	+	7

In a manner similar to how the corresponding possible diseases tables have been handled for the patients 1-4, let us take into account data from the Table 7 in formula (17). As a result, it is possible to receive:

$$p_1^{\text{New patient}}(k=2)=\frac{1}{5},\qquad p_2^{\text{New patient}}(k=2)=\frac{2}{5},\qquad p_3^{\text{New patient}}(k=2)=\frac{2}{5},$$

$$p_1^{\text{New patient}}(k=3)=\frac{1}{2},\qquad p_2^{\text{New patient}}(k=3)=\frac{1}{2},\qquad p_3^{\text{New patient}}(k=3)=0,$$

$$p_1^{\text{New patient}}(k=4)=\frac{1}{2},\qquad p_2^{\text{New patient}}(k=4)=0,\qquad p_3^{\text{New patient}}(k=4)=\frac{1}{2},$$

$$p_1^{\text{New patient}}(k=7)=1,\quad p_2^{\text{New patient}}(k=7)=0,\quad p_3^{\text{New patient}}(k=7)=0.$$

Substituting these results into formula (18), we will have:

$$p^{\text{New patient}}(D_1)=0.55,\quad p^{\text{New patient}}(D_2)=0.225,\quad p^{\text{New patient}}(D_3)=0.225,$$

from which it follows that the probability indicating that a new patient with the detected symptoms, which fully coincide with the symptoms that always accompany the disease D_1, has a disease D_1, is about 2.4 times more than the probability of having a diseases D_2 or D_3. Moreover, this probability is approximately 1.15 times less than the probability of having a disease D_2 in the second patient, where the detected symptoms $\{S_1,S_3,S_4\}$ are only part of the set of symptoms $S^{2,K_2=4}=\{S_1,S_3,S_4,S_7\}$, accompanying the disease D_2.

<u>Example 2 (Formalization and preparation of the source data)</u>. The present example contains only

- the formalization of the Direct knowledge 1 and Direct knowledge 2 described in the Section 3 concerning the information in the Subsection 2.2 on the 6 most common heart diseases and the 21 accompanying symptoms;
- the preparation of the source data necessary for the carrying out the logic-probabilistic diagnosis of the presence of diseases in a patient as proposed in the Section 4.

Let us emphasize once again that the information specified in the Subsection 2.2 is valid medical information obtained on the basis of scientific-theoretical and clinical-experimental medical knowledge (see [23]-[27] and respective numerous sources given there).

Formalization of the Direct knowledge 1 gives the following results, which are an analogue of (3), (4):

$$f_1(x_1,x_2,x_3,x_4,x_5,x_6,x_7,y_1)=x_1\cdot x_2\cdot x_3\cdot x_4\cdot x_5\cdot x_6\cdot x_7\vee\bar{y}_1=1,$$

$$f_2(x_1,x_3,x_5,x_6,x_8,x_9,x_{10},y_2)=x_1\cdot x_3\cdot x_5\cdot x_6\cdot x_8\cdot x_9\cdot x_{10}\vee\bar{y}_2=1,$$

$$f_3(x_1,x_3,x_8,x_9,x_{11},x_{12},x_{13},x_{14},y_3)=x_1\cdot x_3\cdot x_8\cdot x_9\cdot x_{11}\cdot x_{12}\cdot x_{13}\cdot x_{14}\vee\bar{y}_3=1,$$

$$f_4(x_3,x_7,x_9,x_{11},x_{12},x_{13},x_{15},x_{16},y_4)=x_3\cdot x_7\cdot x_9\cdot x_{11}\cdot x_{12}\cdot x_{13}\cdot x_{15}\cdot x_{16}\vee\bar{y}_4=1,$$

$$f_5(x_3,x_7,x_8,x_{11},x_{15},x_{17},x_{18},x_{19},x_{20},y_5)=x_3\cdot x_7\cdot x_8\cdot x_{11}\cdot x_{15}\cdot x_{17}\cdot x_{18}\cdot x_{19}\cdot x_{20}\vee\bar{y}_5=1,$$

$$f_6(x_1,x_3,x_7,x_{11},x_{13},x_{15},x_{21},y_6)=x_1\cdot x_3\cdot x_7\cdot x_{11}\cdot x_{13}\cdot x_{15}\cdot x_{21}\vee\bar{y}_6=1.$$

Formalization of the Direct knowledge 2 gives the following results, which are an analogue of (5), (6):

$$f_7(x_1,y_1,y_2,y_3,y_6)=\bar{x}_1\vee y_1\vee y_2\vee y_3\vee y_6=1,\quad f_8(x_2,y_1)=\bar{x}_2\vee y_1=1,$$
$$f_9(x_3,y_1,y_2,y_3,y_4,y_5,y_6)=\bar{x}_3\vee y_1\vee y_2\vee y_3\vee y_4\vee y_5\vee y_6=1,$$
$$f_{10}(x_4,y_1)=\bar{x}_4\vee y_1=1,\quad f_{11}(x_5,y_1)=\bar{x}_5\vee y_1\vee y_2=1,\quad f_{12}(x_6,y_1,y_2)=\bar{x}_6\vee y_1\vee y_2=1,$$
$$f_{13}(x_7,y_1,y_4,y_5,y_6)=\bar{x}_7\vee y_1\vee y_4\vee y_5\vee y_6=1,$$
$$f_{14}(x_8,y_2,y_3,y_5)=\bar{x}_8\vee y_2\vee y_3\vee y_5=1,$$
$$f_{15}(x_9,y_2,y_3,y_4)=\bar{x}_9\vee y_2\vee y_3\vee y_4=1,\quad f_{16}(x_{10},y_2)=\bar{x}_{10}\vee y_2=1,$$
$$f_{17}(x_{11},y_3,y_4,y_5,y_6)=\bar{x}_{11}\vee y_3\vee y_4\vee y_5\vee y_6=1,\quad f_{18}(x_{12},y_3,y_4)=\bar{x}_{12}\vee y_3\vee y_4=1,$$
$$f_{19}(x_{13},y_3,y_4,y_6)=\bar{x}_{13}\vee y_3\vee y_4\vee y_6=1,\quad f_{20}(x_{14},y_3)=\bar{x}_{14}\vee y_3=1,$$
$$f_{21}(x_{15},y_4,y_5,y_6)=\bar{x}_{15}\vee y_4\vee y_5\vee y_6=1,$$
$$f_{22}(x_{16},y_4)=\bar{x}_{16}\vee y_4=1,\quad f_{23}(x_{17},y_5)=\bar{x}_{17}\vee y_5=1,$$
$$f_{24}(x_{18},y_5)=\bar{x}_{18}\vee y_5=1,\quad f_{25}(x_{19},y_5)=\bar{x}_{19}\vee y_5=1,$$
$$f_{26}(x_{20},y_5)=\bar{x}_{20}\vee y_5=1,\quad f_{27}(x_{21},y_6)=\bar{x}_{21}\vee y_6=1.$$

The implicit formula for the dependence of the 6 diseases studied on the existing 21 symptoms is the following equation, which is an analogue of the functional logical equation (12):

$$\begin{aligned}&(x_1\cdot x_2\cdot x_3\cdot x_4\cdot x_5\cdot x_6\cdot x_7\vee\bar{y}_1)\cdot(x_1\cdot x_3\cdot x_5\cdot x_6\cdot x_8\cdot x_9\cdot x_{10}\vee\bar{y}_2)\cdot(x_1\cdot x_3\cdot x_8\cdot x_9\cdot x_{11}\cdot x_{12}\cdot x_{13}\cdot x_{14}\vee\bar{y}_3)\cdot\\&\cdot(x_3\cdot x_7\cdot x_9\cdot x_{11}\cdot x_{12}\cdot x_{13}\cdot x_{15}\cdot x_{16}\vee\bar{y}_4)\cdot(x_3\cdot x_7\cdot x_8\cdot x_{11}\cdot x_{15}\cdot x_{17}\cdot x_{18}\cdot x_{19}\cdot x_{20}\vee\bar{y}_5)\cdot\\&\cdot(x_1\cdot x_3\cdot x_7\cdot x_{11}\cdot x_{13}\cdot x_{15}\cdot x_{21}\vee\bar{y}_6)\cdot(\bar{x}_1\vee y_1\vee y_2\vee y_3\vee y_6)\cdot(\bar{x}_2\vee y_1)\cdot(\bar{x}_{20}\vee y_5)\cdot(\bar{x}_{21}\vee y_6)\cdot\\&\cdot(\bar{x}_3\vee y_1\vee y_2\vee y_3\vee y_4\vee y_5\vee y_6)\cdot(\bar{x}_4\vee y_1)\cdot(\bar{x}_5\vee y_1\vee y_2)\cdot(\bar{x}_6\vee y_1\vee y_2)\cdot(\bar{x}_7\vee y_1\vee y_4\vee y_5\vee y_6)\cdot\\&\cdot(\bar{x}_8\vee y_2\vee y_3\vee y_5)\cdot(\bar{x}_9\vee y_2\vee y_3\vee y_4)\cdot(\bar{x}_{10}\vee y_2)\cdot(\bar{x}_{11}\vee y_3\vee y_4\vee y_5\vee y_6)\cdot(\bar{x}_{12}\vee y_3\vee y_4)\cdot\\&\cdot(\bar{x}_{13}\vee y_3\vee y_4\vee y_6)\cdot(\bar{x}_{14}\vee y_3)\cdot(\bar{x}_{15}\vee y_4\vee y_5\vee y_6)\cdot(\bar{x}_{16}\vee y_4)\cdot(\bar{x}_{17}\vee y_5)\cdot(\bar{x}_{18}\vee y_5)\cdot(\bar{x}_{19}\vee y_5)=1.\end{aligned}\tag{29}$$

As it has been already mentioned in the Section 4, for a logical-probabilistic diagnosis of the incidence of diseases in a patient, according to the symptoms that have been identified till the present moment, it is necessary to have the input information. Let us suppose that examined patient has symptoms $S^q=\{S_1,S_3,S_7,S_8,S_{10},S_{13},S_{15},S_{19},S_{20},S_{21}\}$ (consequently, we know $q=10$). In a manner similar to how it has been done in the Example 1, we will find the numbers $q_i\ \left(0\le q_i\le\min\{K_i,q\},\ \forall i=\overline{1,n}\right)$, which indicate how many of the discovered symptoms in a patient are part of the list of symptoms S^{i,K_i}, $i=\{1,\dots,6\}$, which always accompany the diseases D_i, $i=\{1,\dots,6\}$. Using the data specified in Subsection 2.1, we can write:

$$S^{1,7}=\{S_1,S_2,S_3,S_4,S_5,S_6,S_7\},\quad S^{2,7}=\{S_1,S_3,S_5,S_6,S_8,S_9,S_{10}\},$$
$$S^{3,8}=\{S_1,S_3,S_8,S_9,S_{11},S_{12},S_{13},S_{14}\},\quad S^{4,8}=\{S_3,S_7,S_9,S_{11},S_{12},S_{13},S_{15},S_{16}\},$$
$$S^{5,9}=\{S_3,S_7,S_8,S_{11},S_{15},S_{17},S_{18},S_{19},S_{20}\},\quad S^{6,7}=\{S_1,S_3,S_7,S_{11},S_{13},S_{15},S_{21}\}.$$

Then it is obvious that

– $S^{q_1}=\underbrace{\{S_1,\ S_3,S_7\}}_{\subset S^{q=10}}\subset S^{1,7}\Rightarrow q_1=3,$

– $S^{q_2} = \underbrace{\{S_1, S_3, S_8, S_{10}\}}_{\subset S^{q=10}} \subset S^{2,7} \Rightarrow q_2 = 4,$

– $S^{q_3} = \underbrace{\{S_1, S_3, S_8, S_{13}\}}_{\subset S^{q=10}} \subset S^{3,8} \Rightarrow q_3 = 4,$

– $S^{q_4} = \underbrace{\{S_3, S_7, S_{13}, S_{15}\}}_{\subset S^{q=10}} \subset S^{4,8} \Rightarrow q_4 = 4,$

– $S^{q_5} = \underbrace{\{S_3, S_7, S_8, S_{15}, S_{19}, S_{20}\}}_{\subset S^{q=10}} \subset S^{5,9} \Rightarrow q_5 = 6,$

– $S^{q_6} = \underbrace{\{S_1, S_3, S_7, S_{13}, S_{15}, S_{21}\}}_{\subset S^{q=10}} \subset S^{1,7} \Rightarrow q_6 = 6.$

Let us validate the condition (13): due to the fact that

$$\left\{S^{q_i} \cap S^{q_j}\right\}_{i<j,\ i=\overline{1,6},\ j=\overline{1,6}} = \{S_1, S_3, S_1, S_3,$$
$$S_3, S_7, S_3, S_7, S_1, S_3, S_7, S_1, S_3, S_8, S_3, S_3, S_7, S_8, S_1, S_3, S_3, S_{13}, S_3, S_8, S_3, S_{13},$$
$$S_3, S_7, S_{15}, S_3, S_7, S_{13}, S_{15},\ S_3,\ S_7,\ S_{13},\ S_{15}\},$$

we have $\text{card}\left(\left\{S^{q_i} \cap S^{q_j}\right\}_{i<j,\ i=\overline{1,6},\ j=\overline{1,6}}\right) = 37.$ Hence,

$$10 = q = \left|\sum_{i=1}^{6} q_i - \text{card}\left(\left\{S^{q_i} \cap S^{q_j}\right\}_{i<j,\ i=\overline{1,6},\ j=\overline{1,6}}\right)\right| = |27 - 37| = 10.$$

Thus, the required input information is entirely ready, and it consists of the following data: $S = \{S_j\}_{j=\overline{1,21}}$ and $m = 21;$ $D = \{D_i\}_{i=\overline{1,6}}$ and $n = 6;$ $S^q = \{S_1, S_3, S_7, S_8, S_{10}, S_{13}, S_{15}, S_{19}, S_{20}, S_{21}\}$ and $q = 10;$ $q_1 = 3,$ $q_2 = 4,$ $q_3 = 4,$ $q_4 = 4,$ $q_5 = 6,$ $q_6 = 6;$ the functional equation (29).

Therefore, to determine the probabilities of the incidence of disease $\{D_i\}_{i=\overline{1,6}}$ in a patient, it is required to follow Steps 1-5 of the procedure of logical-probabilistic diagnosis of the incidence of disease in a patient.

6. Conclusions

In the present paper, the possibility of applying the logic algebra methods for diagnosing human cardiovascular diseases is studied for an incomplete set of symptoms identified at a given time moment. The study is carried out in the presence of three assumptions regarding the cause-effect relations between diseases and their symptoms. These assumptions are based on both scientific theoretical knowledge and (mainly) clinical experimental medical data acquired by medical researchers and practitioners. In the current research, by means of applying the Boolean algebra theory language, the aforementioned cause-effect relation is constructed, which results in a system of logical equations, where its conjunction represents a certain interrelated set of experimentally and / or theoretically confirmed facts describing the principal laws, properties and relations for two types of objects – various sorts of diseases and their symptoms. In other words, the resulting functional equation (a conjunction of logical functions taken from a system of equations) carries more global-scale information about the cause-effect relation between diseases and symptoms in general, rather than particular symptoms found in the case of some specific patient.

Furthermore, having obtained the functional equation as well as received an up-to-date information about the symptoms detected in some particular patient case, the following objects are constructed: (a) an extended diagnostic table, which consists of a truth table for the corresponding functional equation logical function, and a table whose columns contain empirical data about the discovered symptoms for the particular patients' cases; (b) a table of possible diseases of the patient, which is synthesized from the constructed extended diagnostic table. Then, the concept of the probability of having a disease is introduced for each particular patient, a formula for calculating probability of disease is proposed, and finally a probabilistic table of the patient's diseases is constructed based on calculated probabilities.

Acknowledgements

The authors of this paper would like to express their gratitude to a colleague, Anna Suleymanova (Riga, Latvia) for useful discussion and for substantial assistance, which resulted in the formalization of the Direct knowledge 2.

References

Alabdulgade, A., Maccraty, R., Atkinson, M., Vainoras, A., Berškienė, K., Mauricienė, V... & Daunoravičienė, A. (2015). 1753. Human heart rhythm sensitivity to earth local magnetic field fluctuations. *Journal of Vibroengineering*, *17*(6).

Balogh, E.P., Miller, B.T., & Ball, J.R. (2015). Improving Diagnosis in Health Care. Committee on Diagnostic Error in Health Care; Board on Health Care Services; Institute of Medicine; The National Academies of Sciences, Engineering and Medicine.

Bayes, Th. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 370-418.

Bruce, R.A. (1963). Computer diagnosis of heart disease. *Proceedings of the Fifth IBM Medical Symposium*, October 07-11, 1963, Endicott, USA, 77-98.

Coirnfield, J. (1964). Bayes' theorem. *Proceedings of the Sixth IBM Medical Symposium*, Endicott, USA, 163-196.

Cutler, P. (1998). *Problem solving in clinical medicine: From data to diagnosis*. Lippincott Williams & Wilkins.

Gnedenko, B.V. (2017). *Theory of Probability*. Routledge.

Guseynov, S.E., Urbah, A.I. & Andreyev, S.A. (2015b). On one Approach for Stable Estimate of Technical System Efficiency. In *Proceedings of the 10th International Scientific and Practical Conference, III*, p 108.

Guseynov, Sh.E., Aleksejeva, J.V. & Andreyev, S.A. (2015a). On one regularizing algorithm for comprehensive diagnosing of apparatus, engines and machinery. In *Advanced Materials Research*, 1117, 254-257.

Guseynov, Sh.E., Aleksejevs, R., Aleksejeva, J.V., Gasimov, Y.S. (2017). Evaluating attractiveness of the central and the eastern European countries by using index approach for the strategic decision making process related to expansion of the financial service markets. *Advanced Mathematical Models & Applications, 2*(3), 167-214.

Hoffman, J. (2014). Malpractice risks in the diagnostic process. 2014 Annual Benchmarking Report. Cambridge, USA: CRICO Strategies Press, a division of the Risk Management Foundation of the Harvard Medical Institutions.

Karazyan, N.N. (1981). The dependence of myocardial infarction on the activity of the magnetic field of the Earth. *Blood Circulation*, *XIV*(1), 19-21.

Kimura, E., Mibukura, Y. & Miura, S. (1963). Statistical diagnosis of electrocardiogram by theorem of Bayes. *Japanese Heart Journal*, *4*(5), 469-488.

Ledley, R.S. & Lusted, L.B. (1962, December). Medical diagnosis and modern decision making. In *Amer. Math. Soc. Symposia in Applied Mathematics*, 14, 117.

Ledley, R.S., & Lusted, L.B. (1959). Reasoning foundations of medical diagnosis. *Science*, *130*(3366), 9-21.

Lusted, L.B. (1960). Logical analysis in medical diagnosis. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, June 20-July 30, 1960, University of California, Berkeley, USA, Vol. 1: Contributions to the Theory of Statistics, 903-923.

Lusted, L.B. (1968). *Introduction to Medical Decision Making*. Springfield, USA: Charles C. Thomas Publishing.

Malin, S.R.C. & Srivastava, B.J. (1979). Correlation between heart attacks and magnetic activity. *Nature*, *277*(5698), 646.

Melnikov, V.G., Popov, A.A., Yanenko, V.M. (1974). *Medical diagnosis automation*. In: Cybernetics Encyclopedia, Vol.1, (Eds.: V.M.Glushkov, N.M.Amosov, I.A.Artemenko), Moscow, Nauka, 28-29 (in Russian).

Natrins, A., Deikis, I., Guseynov, Sh.E., Leshinskis, K., Sarnovichs, A. (2015). Fundamental Principles of Creation of Attractiveness Principles of Central and Eastern European Countries for Evaluation of Customer Segment and Financial Services Potential. Monography. Riga, Latvia: Business and Financial Research Centre Press, 107 p.

Patton, D.D. (1978, October). Introduction to clinical decision making. In *Seminars in nuclear medicine*, 8(4), 273-282, WB Saunders.

Pendyala, V. S. & Figueira, S. (2017, April). Automated Medical Diagnosis from Clinical Data. In *Big Data Computing Service and Applications (BigDataService), 2017 IEEE Third International Conference*, 185-190.

Rosen, K.H. (2012). *Discrete Mathematics and its Applications*. New York, USA: McGraw-Hill Publishing.

Shusterman, I.L. (2004). Discrete-logical decision making methods in informational medical diagnostic systems. Ph.D. thesis, Ufa, Russian Federation: Ufa State Aviation Technical University Press, 149 p. (in Russian).

Stoupel, E. (1993). Sudden cardiac deaths and ventricular extrasystoles on days with four levels of geomagnetic activity. *Journal of Basic and Clinical Physiology and Pharmacology*, *4*(4), 357-366.

Strzemecki, T. (1992). Polynomial-time algorithms for generation of prime implicants. *Journal of Complexity*, *8*(1), 37-63.

Syzek, T. (2018). The Diagnostic Process: Rediscovering the Basic Steps. Available on-line at: https://blog.thesullivangroup.com/rsqsolutions/diagnostic-process-rediscovering-basic-steps [Accessed: December 31, 2018].

Warner, H.R., Toronto, A.F., Veasey, L.G. & Stephenson, R. (1961). A mathematical approach to medical diagnosis: application to congenital heart disease. *JAMA*, *177*(3), 177-183.