Development of an intelligent system of complex diagnostics of human bioelement status

Sergey A. Miroshnikov
Federal Scientific Center of Biological Systems
Orenburg 460018, Russia
sergey_ru01@mail.ru

Irina P. Bolodurina
Department of Applied Mathematics
Orenburg State University
Orenburg 460018, Russia
ipbolodurina@yandex.ru

Olga Arapova
Department of Applied Mathematics
Orenburg State University
Orenburg 460018, Russia
sando@mail.ru

Denis I. Parfenov
Faculty of Distance Learning Technologies
Orenburg State University
Orenburg 460018, Russia
parfenovdi@mail.ru

Abstract

The creation of an information intelligent preliminary diagnosis system is an important means of monitoring people’s health. Within the framework of the developed intellectual system, preliminary prediction of the level of trace elements content is considered without the sampling of the analysis. As a result of processing of experimental medical data of analyzes of microelement composition, a system of product rules was developed, which is the basis for making diagnostic decisions on the level of trace elements in the human body on the basis of an intelligent system of preliminary diagnosis. The paper considers the problem of the influence of environmental factors on the health of people. The results of the application of the integral index model, generalizing the characteristics of twelve chemical elements, to the assessment of the health status of people employed in industrial enterprises are presented.

1 Introduction

One of the most important state tasks laid down in the state program for the development of the healthcare of the Russian Federation is to ensure an increase in the life expectancy of the population of the country to 74.3 years by 2020. The key factor affecting the life expectancy of the population is the health status of members of civil society. The lifespan of the population directly depends on resources that influence the development of any state. According to the World Health Organization, the largest contribution to the health of the population is made by a group of factors united by the concept of external environment, which includes numerous elements that pollute air, water, food, soil, as well as factors such as noise, vibration, radiation and others. The main source of their origin is the diversified industry. At the same time, studies aimed at studying the influence of

Copyright © by the paper’s authors. Copying permitted for private and academic purposes.
environmental factors on the health status of people working in hazardous industries are becoming particularly relevant [Roc08].

Anthropogenic pollution of the environment is mainly associated with microelements from the group of heavy metals. At present, technogenic microelements are becoming increasingly important. In the immediate vicinity of many industrial enterprises, zones with a high content of various toxic trace elements are formed, including radioisotopes, which pose a threat to health and even human life [Mir13]. The negative impact of these factors on public health is a serious concern for public authorities responsible for both the state of health of the population and the protection of the environment in various states.

In order to comply with the requirements of the state policy in the field of environmental protection, it is necessary to carry out constant monitoring of the health of the population. One of the most important and mandatory conditions for maintaining health is the stability of the chemical composition of the body. Accordingly, deviations in the content of useful chemical elements or the accumulation of toxic elements in the human body can be not only a criterion for environmental ill-being, but also serve as markers at the level of donor diagnosis of abnormalities in health status [Ska15].

In this regard, it is necessary to develop a comprehensive indicator that allows integrating the numerical characteristics of a person's bioelement status into an overall assessment of his health status. An assessment of the microelement status of a person is necessary for an adequate assessment of the situation and the resolution of emerging problems [Kob13].

The definition of human elemental status is possible through the analysis of biological substrates such as blood, urine, hair, and others. Each of these substrates has certain informativeness. Note that the most effective and informative method for testing and diagnosing a person's health status is the analysis of trace elements contained in human hair. According to modern ideas, hair reflects the elemental status over a long period of time. In hair, it is possible to trace the change in the content of an essential or toxic element that occurs when long-term effects of certain factors specific to specific regions, including the environment. It is established that the hair of people living in different countries differs in their chemical composition. This is due to the different contents of chemical elements in drinking water, food, climate, biogeochemical, social, occupational and physiological factors also exert a great influence [Ska15].

However, it should be noted that conducting a full analysis of all trace elements takes a long time and requires significant financial costs. To obtain reliable information, it is necessary to automate the processes of analysis and processing of collected data. In addition, the need to develop an intelligent system for the diagnosis of bioelement status is due to the following factors [Ska15]:

- in a real situation, the doctor has to process a significant amount of clinical, laboratory and instrumental data about the individual characteristics of the patient. Integration of a large number of data, their interpretation, classification of objects is a complex problem;
- the decision has to be taken when an incomplete, uncertain, subjective and contradictory information comes to a–priori;
- the need to minimize financial and time costs for the diagnosis and processing of analysis data.

In the framework of this study, the goal is to form the knowledge base of the intellectual system for preliminary diagnosis of the level of trace elements in the human body using methods of data mining.

2 Related work

It is known that unfavorable factors of anthropogenic nature predominantly affect the elemental status of the population engaged in harmful industries, primarily mining, metallurgy, machine building [Ska00, Ska02, Say90, Aga00, May00, Boe02].

The level of trace elements in the human body is affected by external and internal factors. An important external factor is a prevalence in the earth’s crust since the main source of chemical elements for humans are food and water, with which chemical elements enter the gastrointestinal tract. In works [Aft12, Ska12-2] it is noted that other ways of penetration of microelements - with inhaled air, through the skin and others are possible. Other external factors are also important: the aggregate state of the natural compounds of microelements, the solubility in water, the form of entry into the body, for example, in plants the elements are in biologically active concentrations, so it is better absorbed by the human body.

Internal factors that affect the content and manifestation of the physiological role of chemical elements include their distribution in various organs and tissues, the form of the presence of chemical elements, which affects the
that the content of microelements in the body is characterized by a vector $X$ of trace elements in human hair in percentiles by means of atomic emission spectrometry. Then we will assume indicators of the bioelement composition of the hair.

The components of the vector within the framework of this study will be considered clinico-laboratory makes it difficult to assess the anthropogenic impact of the state of the environment on a person. The random nature of the changes in its components make it difficult to assess the state of health. This, in turn, accompanied by random fluctuations caused by various factors. The large dimensionality of the state vector and the state by a large number of indicators. In this case, the deviation of indicators from “normal” values is data of human hair, certain difficulties arise. First of all, this is due to the fact that it is necessary to assess mineral metabolism\cite{Ska03, Ska12}.

The microelement status of the organism as a whole. Consequently, hair samples are an integral indicator of previously identified, of particular interest is the study of hair. The content of microelements in the hair reflects the genetic apparatus of people\cite{Nov02}.

3 Materials and methods

To identify the exchange of microelements in the human body and the toxic effects of certain heavy metals, as previously identified, of particular interest is the study of hair. The content of microelements in the hair reflects the microelement status of the organism as a whole. Consequently, hair samples are an integral indicator of mineral metabolism\cite{Ska03, Ska12}.

However, within the framework of this study, it has been established that when analyzing the bioelemental data of human hair, certain difficulties arise. First of all, this is due to the fact that it is necessary to assess the state by a large number of indicators. In this case, the deviation of indicators from “normal” values is accompanied by random fluctuations caused by various factors. The large dimensionality of the state vector and the random nature of the changes in its components make it difficult to assess the state of health. This, in turn, makes it difficult to assess the anthropogenic impact of the state of the environment on a person.

To solve the described problems, we introduce a generalized indicator that is a scalar function of the state vector. The components of the vector within the framework of this study will be considered clinico-laboratory indicators of the bioelement composition of the hair.

In this study, the determination of trace element status is carried out on the basis of analysis of the content of trace elements in human hair in percentiles by means of atomic emission spectrometry. Then we will assume that the content of microelements in the body is characterized by a vector $X = (x^1, x^2, \ldots, x^n)$, the components of which are the percentile values of toxins and vital elements in the hair within the framework of the three groups. For classification into groups, we will use the following percentile scale, currently used by physicians: - the first group - the content of the microelement is up to 25 percentiles, then there is a deficiency of this trace
element; the second group - the content of the microelement is in the range of 25 to 75 percentiles, then the level of content is normal; the third group - the content of the microelement is more than 75 percentiles, then the excess content of the microelement is observed.

Then, within the framework of the study, the problem is to construct a scalar function of the state vector $\phi = (X)$ that estimates the value of $s$ (the ideal value corresponding to an absolutely balanced exchange of substances) with a minimum error. The function $\phi = (X)$ is called a generalized exponent. The function $\phi = (X)$ is chosen from the condition of the minimum mean square error of the prediction of the value $s$:

$$E[\varphi(X) - s]^2 \rightarrow \min$$

or in the form of a complete mathematical expectation:

$$E[\varphi(X) - s]^2 = \sum_s p_s E[(\varphi(X) - s)^2|s],$$

where $p_s$ is the probability of state with number $s$.

For $\phi = (X)$ we choose a linear form: $\varphi(X) = \sum_{i=1}^{n} a_i x_i$, $\sum a_i x_i = 1$.

Coefficients of the generalized index $i(i = 1, \ldots, n)$ are determined from the condition for the minimum of the expression (2) on the set $a : \sum a_i = 1$. The method for finding the coefficients of the generalized exponent is described in [Aft12]. Thus, the scalar function of the state vector is defined.

### 4 Practical implementation

The present study was conducted on the example of the database "Results of medical analyzes of microelement composition in the city-megalopolis N". Hair samples were subjected to sample preparation in accordance with the requirements of Guidelines 4.1.1482-03 and 4.1.1483-03 "Determination of chemical elements in biological media and preparations by inductively coupled plasma atomic emission spectrometry and inductively coupled plasma mass spectrometry". The elemental composition of the hair was defined by atomic emission and mass spectrometry (AES and MS) at the Test Laboratory of the ANO BCenter for Biotic Medicine, Moscow (Registration Certificate of ISO 9001: 2000, Number 4017 5.04.06). The biosubstrates were ashed using the MD-2000 microwave decomposition system (USA). The content of elements in the resulting ash was estimated using the Elan 9000 mass spectrometer (Perkin Elmer, USA) and an Optima 2000 V atomic emission spectrometer (Perkin Elmer, USA). The sample size of real medical data of trace element analyzes was 68,000 records, the volume for each microelement is from 3,064 to 3,452 records. The influence of four factors on the level of 23 microelements on the human body was studied: arsenic, phosphorus, aluminum, beryllium, calcium, cadmium, cobalt, chromium, copper, iron, potassium, lithium, magnesium, manganese, sodium, nickel, lead, selenium, silicon, tin, titanium, vanadium, zinc. The following Table 1 gives a summary sample size data for each trace element, broken down by possible values of their level content.

The micronutrients presented have different effects on the human body and in a certain concentration are necessary and safe. The following grouping of microelements is distinguished:

- Minor trace elements: selenium Se, cobalt Co, zinc Zn, manganese Mn, copper Cu.
- Essential trace elements: iron Fe, iodine I, copper Cu, zinc Zn, cobalt Co, chromium Cr, molybdenum Mo, selenium Se, manganese Mg.
- Conditionally essential trace minerals: arsenic As, boron B, bromine Br, fluorine F, lithium Li, nickel Ni, silicon Si, vanadium V, cadmium Cd, lead Pb.
- Vital elements: structural (macro) elements - elements - H, O, N, C; Ca, Cl, F, K, Mg, Na, P, S and 8 trace elements - Cr, Cu, Fe, I, Mn, Mo, Se, Zn.

The full content of essential elements and minimal, not threatening to disrupt the adaptive mechanisms of the body, the presence of toxic and conditionally toxic elements, is one of the most important components of the normal functioning of the human body.

To assess the microelement status, it is necessary to consider the entire complex of trace elements, each of which performs a certain role in ensuring the proper functioning of the organism. The lack of at least one component can lead to various diseases.

Therefore, the creation of an intelligent intelligence system for preliminary diagnosis is an important means of monitoring people's health. Within the framework of the developed intellectual system, a preliminary prediction of the level of the trace element content is considered without sampling the analysis.
Table 1: The data of the volume of samples of trace elements by content levels

<table>
<thead>
<tr>
<th>Element</th>
<th>Scope sampling</th>
<th>Level of microelement content</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>num of rec.</td>
<td>%</td>
<td>num of rec.</td>
</tr>
<tr>
<td>As</td>
<td>3451</td>
<td>3340</td>
<td>96,78</td>
<td>111</td>
</tr>
<tr>
<td>P</td>
<td>3451</td>
<td>2073</td>
<td>60,07</td>
<td>509</td>
</tr>
<tr>
<td>Al</td>
<td>3451</td>
<td>3032</td>
<td>87,86</td>
<td>242</td>
</tr>
<tr>
<td>Be</td>
<td>3451</td>
<td>3451</td>
<td>100,00</td>
<td></td>
</tr>
<tr>
<td>Ca</td>
<td>3451</td>
<td>2104</td>
<td>60,97</td>
<td>480</td>
</tr>
<tr>
<td>Cd</td>
<td>3443</td>
<td>3200</td>
<td>92,94</td>
<td>243</td>
</tr>
<tr>
<td>Co</td>
<td>3451</td>
<td>1652</td>
<td>47,87</td>
<td>13</td>
</tr>
<tr>
<td>Cr</td>
<td>3452</td>
<td>1908</td>
<td>55,27</td>
<td>503</td>
</tr>
<tr>
<td>Cu</td>
<td>3452</td>
<td>2396</td>
<td>69,41</td>
<td>451</td>
</tr>
<tr>
<td>Fe</td>
<td>3450</td>
<td>1852</td>
<td>53,68</td>
<td>338</td>
</tr>
<tr>
<td>K</td>
<td>3450</td>
<td>1721</td>
<td>49,88</td>
<td>666</td>
</tr>
<tr>
<td>Li</td>
<td>3451</td>
<td>3398</td>
<td>98,46</td>
<td>19</td>
</tr>
<tr>
<td>Mg</td>
<td>3451</td>
<td>1620</td>
<td>46,94</td>
<td>406</td>
</tr>
<tr>
<td>Mn</td>
<td>3452</td>
<td>1666</td>
<td>48,26</td>
<td>407</td>
</tr>
<tr>
<td>Na</td>
<td>3450</td>
<td>2038</td>
<td>59,07</td>
<td>703</td>
</tr>
<tr>
<td>Ni</td>
<td>3451</td>
<td>3376</td>
<td>97,83</td>
<td>75</td>
</tr>
<tr>
<td>Pb</td>
<td>3451</td>
<td>3156</td>
<td>91,45</td>
<td>295</td>
</tr>
<tr>
<td>Se</td>
<td>3452</td>
<td>1994</td>
<td>57,76</td>
<td>217</td>
</tr>
<tr>
<td>Si</td>
<td>3451</td>
<td>2012</td>
<td>58,30</td>
<td>1011</td>
</tr>
<tr>
<td>Sn</td>
<td>3083</td>
<td>2950</td>
<td>95,98</td>
<td>124</td>
</tr>
<tr>
<td>Ti</td>
<td>3248</td>
<td>3128</td>
<td>96,31</td>
<td>120</td>
</tr>
<tr>
<td>V</td>
<td>3064</td>
<td>2964</td>
<td>96,74</td>
<td>100</td>
</tr>
<tr>
<td>Zn</td>
<td>3452</td>
<td>1917</td>
<td>55,53</td>
<td>268</td>
</tr>
</tbody>
</table>

When creating model support for the diagnostic system, methods of data mining are used: the algorithm for constructing the decision tree of ID3 and its improved version of C4.5 and the method of direct logical inference based on qualitative medical data.

One of the main problems of decision-making in weakly structured areas, such as medicine, ecobiomedicine, immunology, is the problem of processing and analyzing large amounts of information. In this regard, there is a need to create intelligent systems. The software implementation of the intellectual system is carried out using the means “1C: Enterprise 8.2”. The scheme of the data flows of the intelligent prediagnostic system is shown in Fig. 1.

![Figure 1: Scheme of data flows of an intelligent system for the diagnosis of human bioelement status.](image-url)

We define the input parameters in more detail in the systems. The patient’s personal data is the data of the patient’s medical record. The area of residence is a micro-district in the city, where a person lives. Within the framework of this work, the town of the city N and the special significance of the area of residence are...
considered, as there are factories on the territory of the city\cite{Not06}. In terms of its industrial potential, the metropolitan city of N is considered the largest center in its region. About 20\% of engineering equipment for this huge region is produced at the enterprises of the city of the city N and the region, among them - metalworking and woodworking machines. Along with heavy industry, electric power and non-ferrous metallurgy are rapidly developing. The unique enterprises of the city of the city N and the region produce tin and gold, chemical concentrates, rare metals and nuclear fuel. These types of products are in high demand not only at home and abroad. The following plants are located on the territory of the city: aircraft, metal structures, plastics, building materials, instrumental, cable, brick, foundry instrumentation, chemical, electromechanical, jewelry and others.

The next input parameter is the age group that determines the age of the person. It is advisable to consider it in the age range since the content of trace elements differs in children and in older people due to age-related changes in the human body. Age is presented in groups: up to 3 years, from 3 to 16 years, from 17 to 49 years and over 50 years.

The type of food has a significant effect on the content of microelements. Distinguish the following types of nutrition: breast, breast and lure, mixed, vegetarian, milk-vegetable, meat, diet, hypoallergenic diet. Due to physiological differences, it is necessary to consider the gender of the patient, who takes two values - male or female.

The intelligent preliminary diagnosis system implements the following functions: input of personal data about the diagnosed patient, which includes the patient’s full name and other personal data; extraction of signs (factors) with the purpose of revealing the level of the content of microelements and forming conclusions on diagnostic solutions.

The intelligent system contains: a database and knowledge in which data and knowledge about diagnosed such as last name, first name, patronymic are stored; year of birth; floor; place of residence, type of food; mathematical apparatus of the mechanism of logical inference; a system of production rules; results of express diagnostics of the subjects.

After obtaining the value of informative parameters (factors) for the object to be diagnosed, they are stored in the database and used to formulate conclusions on the diagnosis of the content of trace elements in the human body. The conclusion about the level of the content of trace elements is formed on the basis of the production systems of knowledge, presented in the form of multiple rules. As a result of the operation of the inference engine block, there is a definition of particular dependencies between objects or events.

The dependencies found are represented in the form of rules and can be used both for better understanding of the nature of the analyzed data and for predicting the occurrence of events.

As the most adequate model for constructing production rules, according to the results of ROC analysis, a decision tree was chosen based on the algorithm C4.5. Evaluation of the effectiveness of the obtained decision rules on the examination sample showed fairly good results - with a reliability of 0.8 the objects of the class are “correctly increased”, with a probability of 0.86 - objects of the ”norm” class and with a probability of 0.75 objects of the class “lowered”.

The following Table 2 presented the basis of the production model is laid down the rules obtained by the algorithm C4.5 for each of the microelements.

For each patient, a new fact is processed according to the rules system with the output of the result on the level of trace elements content. At the same time, for each trace element, the service information-the rule number-is indicated.

Based on the results of data analysis, a list of microelements is formed, the level of content of which differs from the value of ”norm”. For this element, recommendations are made for the doctor to refer the patient to an additional examination.

Obtained predicted diagnostic data for all trace elements for each specific patient, the doctor uses for further planning of the survey and the appointment of treatment.

5 Conclusion

Thus, the research suggests a technique implemented in the form of an intellectual support system for making diagnostic decisions on the level of trace elements in the human body on the basis of data analysis using Data Mining methods.

The suggested methodology can form the basis of an environmental information system for decision support, in the process of which an integral indicator of public health is formed, reflecting an assessment of the health of
people working in enterprises with a high anthropogenic load. The application of this environmental information system will allow for constant monitoring of the medical and environmental situation in industrial complexes, and in the development of measures adequate to the current state of the environment, to prevent situations that pose a threat to human health, and thus the environment.

Within the framework of the constructed intellectual system, production rule systems based on decision trees have been formed. Developed by an intelligent system for predicting the level of trace elements in the human body without taking samples, doctors will be able to determine the level of certain trace elements at the time of treatment. This will make it possible to choose the optimal scheme for sampling and analysis of the analyzes for trace elements, based on the individual medical profile of the patient.

6 Acknowledgements
The research was conducted with the support of the Russian Science Foundation (project no. 14-16-00060).

References


