Implementation of Fuzziness of the Hereditary Diseases Clinical Picture in an Expert Diagnostic System

Nikolay A. Blagosklonov [1,2]

[1] *Federal Research Center "Computer Science and Control" of RAS, Vavilova str. 44, kor.2, Moscow, 119333, Russian Federation*
[2] *Pirogov Russian National Research Medical University, Ostrovitianov str. 1, Moscow, 117997, Russian Federation*

Abstract
The study explores an approach to operating with the fuzziness of clinical manifestations of hereditary diseases, based on a combination of three expert evaluations. The contribution to a portrait of diseases of such components as modality, manifestation and severity is considered. Simultaneous accounting of these attributes and comparison of the manifestations of the disease in a patient with a standard expert description make it possible to carry out differential diagnostics at the pre-laboratory stage of diagnosis.

Keywords
Expert system, certainty factors, confidence measure, orphan diseases, expert knowledge, fuzziness of clinical manifestations

1. Introduction

Orphan diseases have received more and more attention lately. In the Russian Federation, this is reflected in the changes made over the past year to the federal law No. 323-FZ (as amended on December 22, 2020) "On the basics of protecting the health of citizens in the Russian Federation" and the RF government decree of April 26, 2012 N 403 (as amended on June 5, 2020) "On the procedure for maintaining the Federal register of persons suffering from life-threatening and chronic progressive rare (orphan) diseases leading to a reduction in the life expectancy of citizens or their disability, and its regional segment". Separately, it is worth noting the decree of the President of the Russian Federation No. 16 of January 5, 2021 "On the creation of a Fund to support children with severe life-threatening and chronic diseases, including rare (orphan) diseases, "Circle of Kindness"", which reflects the government's priority in helping individuals, most often for children with hereditary diseases. According to Federal Law No. 323-FZ, rare (orphan) diseases are diseases that have a prevalence of no more than 10 cases per 100 thousand population.

The use of computer systems to support medical decision-making can increase the efficiency of diagnosing such diseases. Previously, such solutions have already been created and operated in Russia [1], France [2], England [3], Australia [4]. However, at present, there are no domestic (Russian) functioning systems. Foreign systems that assist the physician at the pre-laboratory stage of diagnosis, with all their obvious advantages, have several significant disadvantages, among which the following can be distinguished:

1. Lack of support for the Russian language, which, in addition to the inconvenience in translating concepts, also entails incomplete comparability of individual terms. This is due to a different approach to the description of clinical manifestations, for which metaphorical naming is inherent in medicine.

Russian Advances in Fuzzy Systems and Soft Computing: Selected Contributions to the 10th International Conference «Integrated Models and Soft Computing in Artificial Intelligence» (IMSC-2021), May 17–20, 2021, Kolomna, Russian Federation
EMAIL: nblagosklonov@frccsc.ru
ORCID: 0000-0002-5293-8469

2. Lack of explanation of the solution proposed by the computer program in the classical sense for artificial intelligence. This is especially true for the Face2Gene system [3], which works based on the principles of neural networks, in which a heat map is built as an explanation when working with patients' faces or a certain percentage of confidence is displayed without explaining how it was obtained when only signs are entered into the system the patient. Not all systems provide the physician with lists of matched and non-matched signs (otherwise, "counter-signs").
3. Focus on a web application, that is, to work with such systems, need a stable connection to the Internet and the transfer of patient data to a third-party foreign server, which is contrary to the policy of working with personal data on the territory of the Russian Federation. And often in medical institutions the ability to access the external network is limited, which is associated with the sensitivity of personal medical information and ensuring the security of the infrastructure.

The absence of a domestic Russian intelligent decision-making support system determines the relevance of this work. When developing the GenDiES system, special attention was paid to the problem of the ambiguity of the clinical picture of various severity of hereditary diseases, especially those manifested in early childhood.

2. Specifics of orphan diseases

Lysosomal hereditary metabolic diseases related to orphan [5] are considered as a prototype of the GenDiES system. This group includes about 50 diseases (according to various classifications), the largest subgroups are mucopolysaccharidoses (15 clinical forms), mucolipidoses (8 clinical forms) and gangliosidoses (7 clinical forms) [6]. Like other similar hereditary diseases, they are characterized by a number of features that must be taken into account when building a system. The most significant are:

1. Age-related dynamics of changes in the clinical picture, which is manifested by the progression of individual signs, but also the possible transition of some manifestations to the opposite. An example of the progression of severity is the increasing “coarsening” of the patient's facial features as they grow older from the moment of the primary manifestation of this symptom (manifestation). An example of a transition is a violation of the function of large joints, namely “looseness”, or increased mobility, giving way to “stiffness”, up to complete immobility.
2. The fuzziness of the verbal characteristics of clinical signs, which is expressed in linguistic comparisons or comparisons, without specifying specifics, for example, an enlarged head (macrocephaly).
3. The complexity of reproducing in a linguistic form difficult-to-perceive figurative characteristics. Quite often, a well-established metaphorical expression is used to describe the external manifestations of hereditary diseases. For example, “trefoil-shaped skull”, “gargoyle-like facial features”, “beak-shaped nose”, etc. It is not possible in many cases to decompose such descriptions into separate components, that is, smaller manifestations. Facial signs have been described in sufficient detail by medical researchers and each element of the face is represented by its own description. But when the patient's image is decomposed into separate components, the complexity of external manifestations is lost, or in other words, the “portrait of the disease”.
4. The possible absence of relevant signs in the clinical picture, which is due to the extraordinarily strong contribution of the individual characteristics of each patient, from ethnicity to the combination of the diagnosed disease with the developmental characteristics of the individual (due to heredity) and manifestations of other diseases (possibly of a different genetic nature). In clinical practice, deviations from the complete description of diseases in the literature are quite common.

These features of the manifestation of hereditary diseases lead to the need to develop a computer decision support system that could function in conditions of incomplete and sometimes contradictory information without a significant decrease in the accuracy of the hypotheses being issued. The solution to such a problem is possible using the approaches of fuzzy logic proposed by Lotfi Zadeh [7].

3. Confidence measures in signs

When creating a system that takes into account the possibility of working with fuzzy features, the confidence factors [8] modified by B.A. Kobrinsky [9] were used to characterize the manifestation and severity of signs, as well as a scale developed during the study to assess modality.

Knowledge extraction took place in two stages. At the first stage, the work was carried out with literary data. The cognitologist analyzed monographs, articles, online databases on hereditary diseases with an emphasis on the external manifestations of the disease. Knowledge from the literature was presented in the form of a specially developed form – a textological card [10], which took into account the age of manifestation of the trait, its level of severity and frequency of occurrence in the population. At the second stage, knowledge was extracted from experts. The experts analyzed textological cards and, based on their own experience in orphan diseases, identified a limited set of signs for each group of diseases. The final set of signs with expert assessments for four age groups, including the characterization of modality considering the scale, confidence measures in manifestation and severity of signs, was a structured formalized description of lysosomal storage diseases (matrix diseases – signs). This knowledge representation included 22, 21, 27 signs and 3960, 1764, 2592 expert evaluations for mucopolysaccharidoses, gangliosidoses and mucolipidoses, respectively.

Thus, the developed system is based on expert knowledge that reflects the modality of the sign, confidence in its manifestation and the level of severity at a given age. An integrated approach to the presentation of knowledge about the external manifestations of the disease in different age periods is necessary to overcome the problem of the ambiguity of the clinical picture in orphan diseases and help the physician at the pre-laboratory stage of making a preliminary diagnosis.

4. Model construction

To consider, the possible ambiguity of the manifestations of orphan diseases, a model of the disease was developed based on expert assessments. In general, this model is a reference clinical description of the disease, which considers the contribution of each sign to the formation of the clinical picture. This is achieved because not only the fact of the presence of a feature is assessed, but also its "weigh", which is the product of expert assessments. This approach made it possible to comprehensively integrate expert assessments of modality, manifestation and severity within each age group and to consider the disease as a synthesis of various attributes of each sign. Therefore, the disease model is called the integral disease assessment.

The integral assessment of the disease is the sum of complex assessments of signs within the age group:

$$I = \sum_{i=1}^{n} P_i \,, \tag{1}$$

where:

I – integrated assessment of signs of the disease,

P_i – complex assessment of the sign,

i – the number of features,

n – a set of signs of a disease (group of diseases).

The complex assessment of the sign was calculated by the formula:

$$P_i = M_i \cdot m_i \cdot s_i \,, \tag{2}$$

where:

P_i – complex assessment of the sign,

M_i – the modality of a feature, characterizing its frequency,

m_i – the confidence measure in the manifestation of a trait,

s_i – the confidence measure in the severity of the trait.

The disease model has found application in two possible modifications: as a reference (etalon) model of the disease (I_e), which consisted of all the symptoms described by experts, and as a case model, obtained as the sum of only those symptoms that were present in the patient (I_p).

5. System prototype

To obtain a prototype of the system, a benchmarking algorithm was implemented. The task that this algorithm solves is to compare the case with the reference in order to determine the list of the most appropriate hypotheses for the patient's clinical picture. However, due to the need to handle the fuzziness, an add-on was made. To consider situations when the percentage of coincidence of a case with the standard may be very high, but at the same time the patient has a large number of “extra” features that were not described by experts in the reference variant of the disease, a dynamic threshold was introduced for "features not related to the hypothesis". The implementation of this approach is shown in Figure 1.

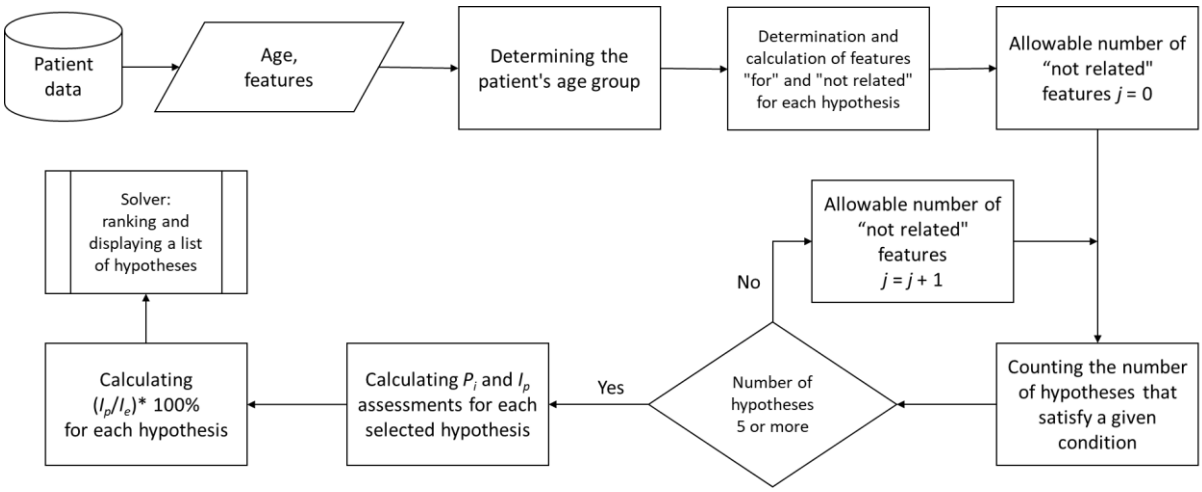

Figure 1: System operation diagram

The superstructure with a dynamic threshold is based on the idea that all the signs described by the physician while examining the patient refer to the manifestations of one hereditary disease. Next, the algorithm checks whether it is possible to select the required number of hypotheses for subsequent ranking. In this case, the number of minimum necessary hypotheses was determined in the amount of five. This is due to the optimal number of preliminary diagnoses that a physician can check using referral for molecular genetic testing to confirm the final diagnosis and, at the same time, is consistent with the series issued in most foreign systems and the previously functioning Russian system. If it is impossible to form a minimum differential diagnostic series of hypotheses with a fixed condition, then the number of admissible features increases by one due to the features of the reference description. Then a check is made. This cycle continues until it becomes possible to form a differential diagnostic series, consisting of the required number of hypotheses.

With regard to accounting for fuzziness, this approach allows, at the stage before comparing the case with the reference descriptions, filter out hypotheses that contain an excessive number of false positive features, while having a lower percentage of overall coincidence with the reference description.

It is advisable to provide the physician with the possibility of choosing the size of the list of hypotheses, for example, in the number of three, five, ten or without limitation. However, it should be noted that the smaller the list of differential diagnostic hypotheses issued by the system, the less likely it is that the correct diagnosis will be included in it. By increasing the list of preliminary diagnoses, the likelihood of entering it a correct diagnosis will increase, but at the same time it may be in a lower position than inaccurate ones.

For the selected hypotheses, integrated case scores are calculated, which are subsequently compared with the reference descriptions and ranked by percentage of agreement from highest to lowest. The ranked list is displayed to the physician indicating:

- signs of the patient, which coincided with the etalon;
- signs of the patient that were absent in the etalon (if any);

- signs from the etalon that were absent in the patient, with a proposal to conduct an additional examination for the presence of such signs.

This approach was tested on twenty case reports of mucopolysaccharidosis from journal articles. Both domestic and foreign publications were selected for the experiment. Descriptions of clinical manifestations were brought to a unified terminology used in the prototype of the GenDiES system.

The criterion for the success of the prototype was the entry of the verified diagnosis (based on molecular genetic analysis according to the publication data) into the ranked list of five hypotheses issued by the system. According to the results of the experiment, the diagnostic accuracy using the prototype of the GenDiES system was 90%.

6. Conclusion

Considering the variability of external manifestations, including the direct fact of manifestation and the strength of the severity of signs in rare hereditary diseases, plays a decisive role in the development of the system.

The use of expert knowledge in relation not only to the facts of the presence or absence of signs in diseases, but based on complex assessments of modality, manifestation, and severity, allows to obtain an integrated assessment of the disease. The implemented model can function under conditions of limited and sometimes inaccurate information about the clinical manifestations of hereditary diseases.

The result of testing the system on the example of descriptions of cases of mucopolysaccharidoses in the literature, taking into account the fuzziness of phenotypic manifestations based on expert assessments, showed an efficiency of 90%.

7. References

[1] B. A. Kobrinskii, Retrospective analysis of medical expert systems, Artificial Intelligence News 2 (2005) 6-17.

[2] S. Ayme, M. Caraboenf, J. Gouvernet, GENDIAG: A computer assisted syndrome identification system, Clinical Genetics 28 (1985) 410–411.

[3] J. E. Allanson, C. Cunniff, H. E. Hoyme, J. McGaughran, M. Muenke, G. Neri, Elements of morphology: standard terminology for the head and face, American Journal of Medical Genetics Part A 149A (2009) 6–28. doi:10.1002/ajmg.a.32612.

[4] A. Fryer, POSSUM (Pictures of Standard Syndromes and Undiagnosed Malformations), Journal of Medical Genetics 28 (1991) 66–67.

[5] Orphanet, 2021. URL: https://www.orpha.net/consor/cgi-bin/index.php.

[6] G. La Marca, Lysosomals, in: N. Blau, M. Duran, K. M. Gibson, C. D. Vici (Eds.), Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases, Springer, Berlin Heidelberg, 2014, pp. 785-793. doi:10.1007/978-3-642-40337-8_52.

[7] L. A. Zadeh, Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic, Fuzzy Sets and Systems 90 (1997) 111-127. doi:10.1016/S0165-0114(97)00077-8.

[8] E. H. Shortliffe, B. G. Buchanan, A Model of Inexact Reasoning in Medicine, in: B. G. Buchanan, E. H. Shortliffe (Eds.), Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project, Addison-Wesley Publ. Co., London, Amsterdam, Sydney, 233-262.

[9] B. A. Kobrinskii, Certainty factor triunity in medical diagnostics tasks, Scientific and Technical Information Processing 46 (2019) 321–327. doi:10.3103/S0147688219050046.

[10] B. A. Kobrinskii, N. A. Blagosklonov, Hybrid approach to knowledge extraction: textual analysis and evaluations of experts, Open Semantic Technologies for Intelligent Systems 2 (2018) 191-195.