Model of integral evaluation of expert knowledge for the diagnosis of lysosomal storage diseases

Nikolay A. Blagosklonov[a], Boris A. Kobrinskii[a]

[a]*Federal Research Center "Computer Science and Control" of RAS, Moscow, Russia*

Abstract

This article proposes an approach to a comprehensive assessment of expert knowledge with using the model. Implemented the ability to account for a fuzzy and incomplete clinical picture of diseases. Based on the hypotheses, differential diagnostic series and comparison of reference models with personal models of new cases are formed, that allows to rate the degree of similarity and identify the disease. A comparative analysis of diagnostic hypotheses was carried out using special algorithms. The study was carried out on the mucopolysaccharidoses as an example, which belong to the class of orphan inherited lysosomal diseases.

1. Introduction

The diagnostics problem of monogenic lysosomal storage diseases that belong to the class of orphan (rare) diseases is a global challenge [1]. The earliest possible detection of this pathology is crucial to development severe changes leading to disability and death. Meanwhile, insufficient level of knowledge in the medical community in regards of the clinical signs of these diseases is often the cause of late identification of diseases [2].

In the present study mucopolysaccharidoses, which are related to lysosomal diseases, and include 15 clinical forms, are considered as an example.

An essential feature of lysosomal diseases is the fuzzy of the symptom's degree of expression in the clinical manifestations. It is determined by the progressive accumulation of macromolecules due to a deficiency of specific enzymes to cause of gene defects. Signs of the disease change progressively with age. Manifestations may be already present in the first year of life. This explains the focus of computer diagnostic systems on early childhood. However, the weakness of previously created and currently used systems is the problem of early detection of the first signs of pathology, and underestimation of the dynamics of changes with children's age [3].

The comprehensive evaluation of the symptoms of diseases in terms of the onset, degree of expression and frequency of occurrence of signs presents a great interest. However, the manifestations of the disease may be incomplete. With this in mind, the attention in the creation of

Russian Advances in Artificial Intelligence: selected contributions to the Russian Conference on Artificial intelligence (RCAI 2020), October 10-16, 2020, Moscow, Russia

nblagosklonov@gmail.com (N.A. Blagosklonov); kba_05@mail.ru (B.A. Kobrinskii)

diagnostic decision support systems was focused on the assessments of the manifestation period and degree of expression of clinical signs, made by the experts. It is necessary to consider the character of the observed changes in the process of identifying diseases. Besides, the complexity of differential diagnosis is determined by the similarity of the clinical manifestations of different lysosomal diseases, in particular of the various types of mucopolysaccharidoses.

The difficulties in forming the knowledge base of the system are determined by the continuum of transitional states, variety of symptoms and a fuzzy clinical picture of lysosomal diseases [4, 5].

2. Problem statement

In clinical practice, physicians in the diagnosis of lysosomal diseases, in particular mucopolysaccharidoses, prior to conducting molecular genetic studies or enzyme analysis, focus on the formation of a differential series of 3 to 5 nosological forms.

At creating an expert system, a number of stages have been identified.

At the first stage, knowledge about the clinical manifestations of diseases were sequentially extracted from the world literature and from experts. The cognitologist synthesized the data presented in the medical literature, and the experts specified the characteristic signs for each from diseases and supplemented them with expert evaluations. Manifestation and degree of expression of signs certainty factors [6] and the coefficient of modality were used to assess the symptoms of diseases. Row scales were developed for: (a) age intervals, (b) periods of manifestation, (c) fuzzy of signs, and (d) modality of signs.

At the second stage, "diseases – signs" matrix was formed, which included expert evaluations for each form of the disease by age group.

At the third stage, a complex evaluation of each symptom and an integrated assessment of the disease, which can be considered a model of diseases, were carried out.

At the fourth stage, six special algorithms for comparative differential diagnosis were developed, tested on a group of mucopolysaccharidoses. This made it possible to provide a choice of the sequence of actions when recognizing a new case.

At the fifth stage, the "disease - signs" matrix was formed, an integrated presentation of the clinical forms of diseases and algorithms for comparative diagnostics, which formed the basis for the formation of the knowledge base.

The sixth stage is focused on differential diagnosis and proposing a number of hypotheses about the disease in the patient.

Schematic representation of the steps is shown in Figure 1.

3. Presentation of expert knowledge

A textological card was created to extract the signs characterizing the clinical picture of diseases from the literature. Great attention was given to the modality of the signs, and the selection of signs that did not require the use of complicated methods for primary diagnosis was carried out. At expert evaluations took into account the fuzzy of the manifestations of those parameters.

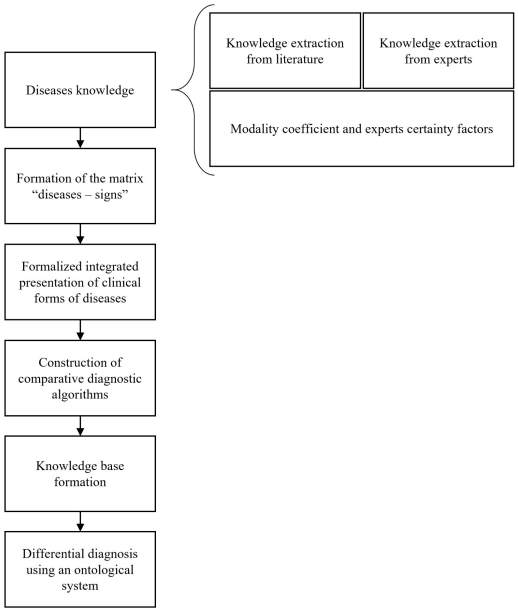

Figure 1: Generalized scheme for creating an artificial intelligence system for diagnosing orphan diseases

At first, all signs found in literary sources were included in the textological cards by a cognitologist. Then diagnostically insignificant characteristics were excluded, and the modality of relevant characteristics was evaluated, with the participation of two experts.

For the group of mucopolysaccharidoses, 22 parameters were selected. Each symptom was accompanied by a coefficient of modality, and two certainty factors – for manifestation and for degree of expression at each from the four age groups.

The scale correction for the expression of signs was adjusted at the results of testing the artificial intelligence system for the differential diagnosis of mucopolysaccharidoses [7].

3.1. Age groups

The physiological characteristics of the development of children and the progressive course of mucopolysaccharidoses division age groups. This allowed more efficient differential diagnosis. In the course of the study, four previously formed age groups of children and adolescents were modified [8]. Table 1 presents age periods and interval scales, allowing to take into account the temporal fuzzy of the manifestation of signs.

In medicine, it is customary to formally consider a person (child) to have not reached the next age before the date of birth (for example, at 2 years 11 months 29 days a child is still considered a two-year-old). The "Intervals" column in Table 1 is presented as decimal numbers. This is necessary in order to avoid erroneous estimates, for 2 years and 6 months is not synonymous with 2.6 years.

Table 1
Age groups

Age groups	Intervals	Possible extension of the age interval
Period up to 1 year	0 – 0.99	0 – 1.25
Period 1-3 years	1 – 3.99	0.75 – 4.25
Period 4-6 years	4 – 6.99	3.75 – 7.25
7 years and older	7 – 17.99	6.75 – 17.99

The column "Possible extension of the age intervals" allows to take into account the fact that a sign can manifest at the time other than indicated in the first column, for example not at 4 years, but at 3 years 10 months. This corresponds to the fuzzy of age-related manifestations of signs. For this, an age "corridor" is used, during which signs of the nearest age group can be considered. Based on the opinion of experts the reserve was determined to be 1/4 year (3 months), since the manifestations in most patients are covered by such an extension of the age intervals.

3.2. Modality coefficients of signs

Modality involves evaluating a concept from a certain point of view [9]. In this study, this is an assessment of the diagnostic significance of the symptom. In this case, we consider modality as an appreciation of the diagnostic significance of the sign. Modality is characterized by a coefficient obtained from experts, taking into account data from various sources, including the frequency of occurrence of the signs in age groups in publications and international databases Genetic and Rare Diseases Information Center - GARD (https://rarediseases.info.nih.gov/) and Human Phenotype Ontology - HPO (https://hpo.jax.org/app/).

Ultimately, modality (M) determines the level of relevance of clinical features on the scale, which is characterized by three gradations. Modality evaluations were introduced in correspondence to the diagnostic role of signs:

- for main symptoms – 5;
- for necessary symptoms – 4;
- for secondary symptoms – 2.

This coefficient allows to determine more accurately the contribution of each individual symptom to the clinical picture of the disease. In cases where the attribute is absent due to the age the concept of modality cannot be used. In these cases the coefficient is replaced by 0. Modality coefficients M_1, M_2, M_3, M_4 are formally labelled in accordance with the presence of the four age groups mentioned above.

3.3. Certainty factors for the manifestation of signs

The period of development of the clinical signs of the disease is called the manifestation of the disease. The significance of the manifestation in the diagnostic process is determined by the fact that due to the individual characteristics of the organism, the character of the development of clinical picture in patients is different. The manifestation and degree of expression of signs in monogenic diseases depends on the level of specific ferment deficiency and gene penetrance.

Thus, the difference in the period of manifestation (m_i) is assessed by a confidence measure (certainty factor), that reflects the level of confidence of the experts that the symptom appears at a given age (age group). A scale in the interval [-1; 1], set as a symmetric function [10], was used to assess the certainty factor for the manifestation of signs:

- The value "-1" characterizes the impossibility of the manifestation of this trait in the specific age group for physiological reasons.
- The value "0" characterizes the norm, either the absence of a sign due to the features of the course of the disease, or the option when the sign has already manifested (in previous age groups).
- Value in the range [0.1; 1] characterizes a confidence measure of experts' that a sign will appear (manifest) in a given age period.

It should be noted that values other than "-1" in total cannot exceed "1" for four age groups.

Evaluations of m_i from previous age periods m_1, m_2, m_3, m_4 complement each other (except for negative values), which allows to obtain a total manifestation certainty factor (m_k):

$$m_k = \sum_{i=1}^{4} m_i \ , \ for \ m_i \geq 0 \tag{1}$$

Thus, the certainty factors for the manifestation allow to indicate the onset of the development of the sign at a certain age, and the total score for age groups reflects the manifestation at a certain age.

3.4. Certainty factors for the degree of expression of signs

With the progression of the lysosomal storage diseases, an important characteristic of the signs that makes up the clinical picture is the degree of their expression. This is a trend to change at various stages of the development of the pathological process.

Thus, the assessment of expression (s) characterizes the confidence of experts in the force of manifestation of this sign in a specific age group. Accordingly, evaluations are given for each age period: s_1, s_2, s_3, s_4.

Unlike the manifestation scale [-1; 1], the scale for assessing the degree of expression of the sign was determined in the interval [0; 10], where "0" corresponds with the situation when the symptom is absent in the patient, and "10" corresponds with the maximal manifestation of the symptom. Changes in degree of expression by age indicate indirectly the rate of the development of the symptoms.

Scale [0; 10] was proposed to obtain mathematically correct estimates when integrating the results on three scales.

pos=[ht]

Table 2
Form of the matrix “diseases – signs”

Sign name	Disease name											
	up to 1 year			1-3 years			4-6 years			7 years and older		
	M_i	m_i	s_i	M_i	m_i	s_i	M_i	m_i	s_i	M_i	m_i	s_i
Sign 1	M_1	m_1	s_1	M_2	m_2	s_2	M_3	m_3	s_3	M_4	m_4	s_4
Sign 2	M_1	m_1	s_1	M_2	m_2	s_2	M_3	m_3	s_3	M_4	m_4	s_4

3.5. Formation of the matrix “diseases – signs”

The form of the matrix “diseases – signs” is displayed in Table 2, in which in the columns are the diseases for four age groups, with indication of coefficients of modality and certainty factors of manifestation and degree of expression of signs for each age group. Signs of disease are arranged line by line.

Appreciations for each clinical form of mucopolysaccharidosis were evaluated by experts when comparing clinical forms differing in the degree of expression This is advisable in order to achieve more reliable description in connection with the similarity and fuzzy of symptoms in the clinical picture of related forms of the disease.

An example of the matrix is displayed in Table 3, filled down for the type VII mucopolysaccharidosis (Sly syndrome).

Table 3
Matrix “diseases – signs” mucopolysaccharidosis type VII (fragment)

Sign name	Mucopolysaccharidosis type VII (Sly syndrome)											
	up to 1 year			1-3 years			4-6 years			7 years and older		
	M_i	m_i	s_i	M_i	m_i	s_i	M_i	m_i	s_i	M_i	m_i	s_i
Growth inhibition	2	0.2	1	2	0.1	4	2	0.2	5	2	0.2	7
Corneal clouding (determined without using a slit lamp)	0	-1	0	4	0.3	2	4	0.1	4	4	0.1	7

4. System of signs evaluation

4.1. Complex evaluation of signs

An operation of multiplying expert evaluations of the manifestation, degree of expression and modality of symptoms in a certain age group was used to obtain a complex evaluation of signs (P_i). It resulted in the following formula:

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

where:

P_i – sign (symptom),

M_i –modality of the sign, characterizing its frequency,

m_i – certainty factor for the manifestation of the sign,

s_i – certainty factor for the degree of expression of the sign.

An array of P_i estimates was obtained by the calculation of all parameters of the disease according to this formula.

This complex evaluations may require revision or clarification only in the course of obtaining new data on the clinical forms of diseases.

4.2. Integrated evaluation of disease

The next stage was implemented to obtain a summary index characterizing the clinical picture of the disease as a whole, considering the previously obtained complex evaluations of signs (P_i). This aggregate of clinical manifestations was called an integrated evaluation (I) of the disease for each age group:

$$I = \sum_{i=1}^{n} P_i \, , \tag{3}$$

where:

I – integrated evaluation of the signs of the disease,

P_i – sign,

i – the number of signs,

n – the aggregate of signs of the disease (group of diseases).

This formula was used in two variations: one for a total score for all the signs that may occur with a particular disease in an age group, and another for a score for signs that occur in a particular case (in a particular patient).

In the first variation, this formula was called the reference integrated disease evaluation (I_e), in this case, for each clinical form of mucopolysaccharidosis. The basis for the reference evaluation is a set of signs that was determined at the stage of problem statement for creating an intelligent diagnostic system.

In the second variation, when patients may have a different incomplete set of signs, the concept of a personal integrated evaluation (I_p) was introduced. Thus, the identical formula I was used in calculating both I_e and I_p. However, the number of signs n differ: when calculating I_e, it is 22 for mucopolysaccharidoses, and for I_p it can exist in the range from 1 to 22, depending on the presence of manifestations in the patient.

Thus, in contrast to the semantic similarity metrics for measuring the phenotypic similarity between a new patient and the base of hereditary diseases annotated using HPO [11], we in the intelligent diagnostic system correlate a personal integrated assessment with a reference disease assessment based on the developed formula.

5. Diagnostic hypotheses based on model comparisons

For 15 clinical forms of mucopolysaccharidoses, the use of the model allowed, taking into account four age groups, the formation of 60 integrated reference estimates of I_e. The personal I_p of a particular case is calculated according to the available signs for each clinical form, and a comparison is made with 15 I_e of the same age period as the patient.

A differential diagnostic series of hypotheses, including several forms of mucopolysaccharidoses from the set I_e is formed at this step of the algorithm.

Let's have a look at an example of using the integrated evaluation model for a 3 months old patient (description taken from publication [12]). The patient had the following signs: coarse facial features, kyphoscoliosis, stiffness of large joints, hepatomegaly, splenomegaly, cardiopathy. The result of calculations of I_p and the corresponding values of I_e is presented in Table 4.

Table 4
Calculation according to the clinical case

The clinical form of mucopolysaccharidosis	I_p	I_e
Hurler syndrome	38.0	93.2
Hurler-Scheie syndrome	2.5	19.6
Scheie syndrome	0.0	0.4
Hunter syndrome (severe course)	1.8	14.1
Hunter syndrome (mild course)	0.0	0.6
Sanfilippo syndrome A	0.0	1.6
Sanfilippo syndrome B	0.0	1.6
Sanfilippo syndrome C	0.0	0.0
Sanfilippo syndrome D	0.0	0.0
Morquio syndrome A (rapidly progressing)	0.0	0.8
Morquio syndrome A (slowly progressing)	0.0	0.0
Morquio syndrome B	0.0	0.0
Maroteaux-Lami syndrome (rapidly progressing)	0.0	15.6
Maroteaux-Lami syndrome (slowly progressing)	0.0	0.4
Sly syndrome	3.0	9.5

Despite the patient having an atypical picture of the disease (as a rule, signs of stiffness of large joints and kyphoscoliosis do not occur at this age), the greatest similarity among the differentiated diagnoses was found with type IH mucopolysaccharidosis (Hurler's syndrome). This corresponds with a diagnosis verified by genetic testing.

However, not in all cases, such a direct method of comparison by relevant features is effective. In the medical practice there are patients whose to advance a hypothesis about the disease can only on the basis of signs that are secondary to one or a number of diseases but are of great importance for other differentiable forms. In such clinical cases, false positive hypotheses are possible, when the system may put forward an assumption that is not relevant to this case.

In this regard, differential diagnostic algorithms were proposed and tested, taking into ac-

count the diagnostic significance of the signs for each disease.

6. The approach for choice of differential diagnosis algorithm

It is possible to use different approaches to the practical implementation of the differential diagnosis of orphan diseases. In the course of the study, variants of algorithms were considered, providing for a comparison of the calculated case estimates only among themselves (absolute values), and a comparison in similarity with reference evaluations (relative values), which could either take into account or not take into account the modalities of the signs. As an experiment, options for diagnostic solutions were considered with exclusion of expert certainty factors of signs. This made it possible to obtain comparative diagnostic results under the conditions of using the developed model and in a simplified version by taking into account the patient's signs confirming or excluding a possible diagnosis.

6.1. Differential diagnostic algorithms

In the study developed six different algorithms below, which were then tested to identify the positive and negative aspects of their application.

6.1.1. Algorithm one

For each disease, the sum of complex estimates of the signs of P_i, called I_p, is calculated (independence of their modality). Then the hypotheses were ranked by the sum of I_p from greater to lesser. Thus an ordered set of hypotheses is formed.

6.1.2. Algorithm two

For each disease, the sum of complex estimates of the signs of P_i, that is, I_p (independence of their modality), is calculated. Further, in contrast to algorithm one, the percentage of coincidence of I_p of a particular case with a reference integrated assessment of I_e of a disease in a given age group is calculated. Then the hypotheses are ranked by the percentage of coincidence from larger to smaller.

6.1.3. Algorithm three

For each disease, integral estimates of I_p are calculated taking into account the modality of the signs. Only signs with modalities "main" and "necessary" are taken into account. Then there is a ranking of hypotheses by the sum of I_p from greater to lesser.

6.1.4. Algorithm four

For each disease, integral estimates of the case of I_p are calculated taking into account the modality of the signs. Only attributes with modalities "main" and "necessary" are taken into account. In contrast to algorithm three, the percentage of coincidence of the patient's attributes

I_p with the reference integrated assessment I_e is calculated for signs with the "main" and "necessary" modalities in this age group. Hypotheses are ranked by the percentage of coincidence from larger to smaller.

6.1.5. Algorithm five

For each disease, the number of signs of any modality "for" and "against" is calculated. If the number of signs denying a hypothesis exceeds a predetermined threshold, then the hypothesis is rejected. The remaining hypotheses are ranked by the number of signs "for".

6.1.6. Algorithm six

For each disease, the number of signs of any modality "for" and "against" is calculated. The percentage of each group is calculated relative to the total number of patient symptoms. If signs denying a hypothesis exceed a certain threshold relative to the total number of patient attributes, then such a hypothesis is not considered. The remaining hypotheses are ranked taking into account the percentage of signs "for" from the total number of signs of the patient.

6.2. Algorithm's testing

Selection of the optimal differential diagnostic algorithm based on the clinical picture of 20 patients from different countries published in journal articles [13, 14, 15, 12, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31]. Signs of patients with orphan diseases (including age and gender) were included in the database. Then, an analysis was made of the results for each algorithm with the allocation of the first 5 hypotheses in each case. If the clinical form of the disease was confirmed by genetic testing in the article and was present among the selected hypotheses, then differential diagnosis using a specific algorithm was considered successful. The experimental results are presented in Table 5. In this test of the fifth algorithm, the threshold for rejecting hypotheses was set with the number of signs "against" 3 or more, and for the sixth algorithm more than 33% of the total number of patient symptoms. Both thresholds are quite soft, but the initial testing was decided to be carried out under such conditions. In the case of insufficient effectiveness of these thresholds, there was the possibility of a further tightening of the selection criteria.

As can be seen from Table 5, the best result was shown for differential diagnostics using the third algorithm for ranking hypotheses by the integral assessment I_p taking into account the modalities of the signs.

Next, an experiment was conducted to increase the accuracy of diagnostics by a pairwise combination of algorithms. According to its results, the most promising was the approach with preliminary exclusion of the hypothesis with the greatest number of "against" signs according to the fifth algorithm, and among the remaining hypotheses, ranking according to one of the first four algorithms.

For algorithm five, at a given threshold for rejecting hypotheses, the correct diagnosis was rejected in 4 cases out of 20. That is, in 20% of cases the correct diagnosis was not made. In this regard, a number of additional tests were carried out in order to clarify the "against" threshold for algorithm five using a dynamic approach.

Table 5
Algorithm's testing results

Patient number	Confirmed diagnosis	A confirmed diagnosis among the first 5 hypotheses based on the results of the algorithm					
		1	2	3	4	5	6
1 [13]	Hunter syndrome	Yes	Yes	Yes	No	Yes	Yes
2 [14]	Maroteaux-Lami syndrome	Yes	Yes	Yes	No	Yes	Yes
3 [15]	Maroteaux-Lami syndrome	Yes	Yes	Yes	Yes	Yes	Yes
4 [12]	Hurler syndrome	Yes	Yes	Yes	Yes	No	No
5 [16]	Hurler syndrome	Yes	Yes	Yes	Yes	Yes	Yes
6 [17]	Sanfilippo syndrome C	No	Yes	No	Yes	Yes	Yes
7 [18]	Sanfilippo syndrome B	Yes	Yes	Yes	Yes	No	No
8 [19]	Maroteaux-Lami syndrome	Yes	Yes	Yes	Yes	No	No
9 [20]	Sly syndrome	Yes	Yes	Yes	Yes	No	No
10 [21]	Sly syndrome	Yes	No	Yes	No	Yes	Yes
11 [22]	Sly syndrome	Yes	Yes	Yes	Yes	Yes	Yes
12 [23]	Maroteaux-Lami syndrome	Yes	No	Yes	No	Yes	Yes
13 [24]	Hunter syndrome	Yes	No	Yes	No	Yes	Yes
14 [25]	Hurler-Scheie syndrome	Yes	No	Yes	Yes	Yes	No
15 [26]	Sly syndrome	Yes	Yes	Yes	Yes	Yes	Yes
16 [27]	Morquio syndrome A	Yes	Yes	Yes	Yes	Yes	Yes
17 [28]	Morquio syndrome A	No	Yes	Yes	Yes	Yes	Yes
18 [29]	Morquio syndrome B	No	Yes	No	Yes	Yes	Yes
19 [30]	Sanfilippo syndrome A	No	Yes	No	Yes	Yes	Yes
20 [31]	Hurler syndrome	Yes	No	Yes	No	Yes	No
Total		16	15	17	14	16	14

This approach was as follows: according to algorithm five, the number of diagnoses in the differential series was calculated, which had 0 signs "against". If the number of hypotheses is less than 5, then the threshold for signs "against" is increased by one to expand the differential series. This process continued until the selection of 5 or more hypotheses.

Next, ranking was carried out according to the first to fourth algorithms. The results of this experiment testing algorithms are presented in Table 6.

As can be seen from Table 6, the best result is obtained according to algorithm two, when ranking is carried out according to the percentage of coincidence I_p of a particular case with a reference estimate of I_e. Improving diagnostics with a combined approach for this algorithm took place in 3 cases. This allowed us to get 18 correct diagnoses out of 20 and, as a result, made it possible to achieve 90% accuracy.

In addition, it should be noted that the use of the primary rejection of hypotheses on the grounds of "against" excluded the correct diagnosis not in four, but only in 1 out of 20 cases, that is, only in 5% of cases there was no possibility of making an accurate diagnosis.

Thus, the process of differential diagnosis using the developed models and algorithms is as

Table 6
Algorithm's testing results

Patient number	Confirmed diagnosis	A confirmed diagnosis among the first 5 hypotheses based on the results of the algorithm			
		1	2	3	4
1 [13]	Hunter syndrome	Yes	Yes	Yes	Yes
2 [14]	Maroteaux-Lami syndrome	Yes	Yes	Yes	No
3 [15]	Maroteaux-Lami syndrome	Yes	Yes	Yes	Yes
4 [12]	Hurler syndrome	Yes	Yes	Yes	Yes
5 [16]	Hurler syndrome	Yes	Yes	Yes	Yes
6 [17]	Sanfilippo syndrome C	Yes	Yes	Yes	Yes
7 [18]	Sanfilippo syndrome B	Yes	Yes	Yes	Yes
8 [19]	Maroteaux-Lami syndrome	No	No	No	No
9 [20]	Sly syndrome	Yes	Yes	Yes	Yes
10 [21]	Sly syndrome	Yes	Yes	Yes	Yes
11 [22]	Sly syndrome	Yes	Yes	Yes	Yes
12 [23]	Maroteaux-Lami syndrome	Yes	Yes	Yes	Yes
13 [24]	Hunter syndrome	Yes	No	Yes	No
14 [25]	Hurler-Scheie syndrome	Yes	Yes	Yes	Yes
15 [26]	Sly syndrome	Yes	Yes	Yes	Yes
16 [27]	Morquio syndrome A	Yes	Yes	Yes	Yes
17 [28]	Morquio syndrome A	Yes	Yes	Yes	Yes
18 [29]	Morquio syndrome B	No	Yes	No	Yes
19 [30]	Sanfilippo syndrome A	No	Yes	No	Yes
20 [31]	Hurler syndrome	Yes	Yes	Yes	Yes
Total		17	18	17	17

follows:

- At the first step, the patient's clinical picture analyzes the number of "for" and "against" in relation to each hypothesis.
- In the second step, how many hypotheses are counted have 0 signs "against". If 5 or more, then the differential diagnosis process continues.
- In the third step, the percentage of coincidence calculated for the patient I_p with the reference I_e for the selected hypotheses is compared.
- In the fourth step, the hypotheses are ranked and the percentage of coincidence of the integrated estimates of I_p from maximum to minimum, the first 5 are selected.
- In the fourth step, the hypotheses are ranked by the percentage of coincidence of the integrated estimates of I_p and I_e from maximum to minimum, the first 5 are selected.

7. Differential diagnosis using an ontological system

The prototype of the differential diagnostic system is created on the basis of the IACPaaS platform (https://iacpaas.dvo.ru/). Knowledge of orphan diseases, including expert assessments, is presented in the form of an ontology. The solver will produce a list of ranked hypotheses at the output of the system. As an explanation, the physician will receive a list of signs that served as the cause of this hypothesis or hypoteses. Symptoms are grouped into 4 blocks according to modalities: main, necessary, secondary and antisymptoms (signs "against" for this disease). In the future, it is planned to request additional signs of the patient, information about which is necessary to increase the likelihood of the hypothesis put forward.

8. Conclusion

The model for the integrated evaluation of expert knowledge for the differential diagnosis of hereditary orphan lysosomal storage diseases has been implemented. The model is based on an integrated approach to three parameters: modality, manifestation and degree of expression of signs. It is the basis for comparing a new object with previously formed reference variants of known clinical forms.

Expert evaluations (modality coefficients and certainty factors) were used to reflect the fuzzy of the pathological changes. Linguistic and interval scales were introduced to characterize the fuzzy signs and time boundaries of the manifestation of symptoms at different age periods. During the study, differential diagnostic algorithms were developed for subsequent inclusion in the IACPaaS platform. As a result of the experiment, the most effective combination was selected among six algorithms. The experiment of differential diagnosis of mucopolysaccharidoses was carried out using the clinical data of patients from articles in Russian and English.

The model that takes fuzzy transitional pathological conditions [5] may improve the efficiency of disease recognition with the incomplete and fuzzy clinical picture.

References

[1] M. R. Bellettato, R. Tomanin, M. Scarpa, Pathophysiological aspects of lysosomal storage disorders, in: R. Parini, G. Andria (Eds.), Lysosomal Storage Diseases: Early Diagnosis and New Treatments, John Libbey Eurotext, Montrouge, 2010, pp. 31 – 42.

[2] S. Tomatsu, T. Fujii, M. Fukushi, T. Oguma, T. Shimada, M. Maeda, K. Kida, Y. Shibata, H. Futatsumori, A. M. Montaño, R. W. Mason, S. Yamaguchi, Y. Suzuki, T. Orii, Newborn screening and diagnosis of mucopolysaccharidoses, Molecular Genetics and Metabolism 110 (2013) 42 – 53. doi:https://doi.org/10.1016/j.ymgme.2013.06.007.

[3] K. Kawamoto, C. A. Houlihan, E. A. Balas, D. F. Lobach, Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success, BMJ 330 (2005) 765 – 772.

[4] S. Bavisetty, W. W. Grody, S. Yazdani, Emergence of pediatric rare diseases, Rare Diseases 1 (2013) e23579.

[5] B. A. Kobrinskii, Fuzzy and reflection in the construction of a medical expert system, Journal of Software Engineering and Applications 13 (2020) 15 – 23. URL: http://www.scirp.org/journal/PaperDownload.aspx?paperID=99099. doi:10.4236/jsea.2020.132002.

[6] B. A. Kobrinskii, Certainty factor triunity in medical diagnostics tasks, Scientific and Technical Information Processing 46 (2019) 321 – 327.

[7] V. Gribova, A. Kleschev, P. Moskalenko, V. Timchenko, L. Fedorisdiev, E. Shalfeeva, The IACPaaS cloud platform: features and perspectives, 2017 Second Russia and Pacific Conference on Computer Technology and Applications (RPC) (2017) 80 – 84.

[8] B. A. Kobrinskii, N. S. Demikova, N. A. Blagosklonov, Knowledge engineering in construction of expert systems on hereditary diseases, in: S. O. Kuznetsov, G. S. Osipov, V. L. Stefanuk (Eds.), Artificial Intelligence, volume 934, Springer International Publishing, Cham, 2018, pp. 35 – 45.

[9] A. A. Ivin, A. L. Nikiforov, Dictionary of logic, Humanitarian publishing center VLADOS, Moscow, 1997. (In Russian).

[10] I. Z. Batyrshin, Towards a general theory of similarity and association measures: similarity, dissimilarity and correlation functions, Journal of Intelligent and Fuzzy Systems 36 (2019) 2977 – 3004.

[11] S. Köhler, M. H. Schulz, P. Krawitz, S. Bauer, S. Dölken, C. E. Ott, C. Mundlos, D. Horn, S. Mundlos, P. N. Robinson, Clinical diagnostics in human genetics with semantic similarity searches in ontologies, The American Journal of Human Genetics 85 (2009) 457 – 464.

[12] L. S. Namazova-Baranova, N. D. Vashakmadze, M. A. Babaykina, E. N. Basargina, N. V. Zhurkova, A. K. Gevorkyan, L. M. Kuzenkova, T. V. Podkletnova, K. V. Zherdev, O. B. Chelpachenko, T. D. Degtyareva, Effectiveness of modern methods of treating type I mucopolysaccharidosis patients, Pediatric pharmacology (Pediatricheskaya farmakologiya) 11 (2014) 76 – 79. doi:10.15690/pf.v11i6.1220, (In Russian).

[13] N. D. Vashakmadze, L. S. Namazova-Baranova, A. K. Gevorkyan, L. M. Kuzenkova, A. D. Khristochevskiy, L. M. Vysotskaya, A. S. Dadashev, Mucopolysaccharidosis type II, Pediatric pharmacology (Pediatricheskaya farmakologiya) 8 (2015) 66 – 68.

[14] A. Jurecka, E. Zakharova, V. Malinova, E. Voskoboeva, A. Tylki-Szymańska, Attenuated osteoarticular phenotype of type VI mucopolysaccharidosis: a report of four patients and a review of the literature, Clinical rheumatology 33 (2013) 725 – 731. doi:10.1007/s10067-013-2423-z.

[15] O. V. Paramei, S. S. Zhilina, Eye manifestations of Maroteaux-Lami syndrome, The Russian Annals of Ophthalmology (Vestnik oftalmologii) 120 (2004) 41 – 42. (In Russian).

[16] O. Gabrielli, L. A. Clarke, A. Ficcadenti, L. Santoro, L. Zampini, N. Volpi, G. V. Coppa, 12 year follow up of enzyme-replacement therapy in two siblings with attenuated mucopolysaccharidosis I: the important role of early treatment, BMC Medical Genetics 17 (2016).

[17] H. J. Huh, J. Y. Seo, S. Y. Cho, C.-S. Ki, S. Y. Lee, J. W. Kim, H. D. Park, D.-K. Jin, The first korean case of mucopolysaccharidosis IIIC (Sanfilippo syndrome type C) confirmed by biochemical and molecular investigation, Annals of Laboratory Medicine 33 (2013) 75 – 79.

[18] Y.-E. Kim, H.-D. Park, M.-A. Jang, C.-S. Ki, S.-Y.Lee, J.-W. Kim, S. Y. Cho, D.-K. Jin, A novel

mutation (c.200T>C) in the NAGLU gene of a Korean patient with mucopolysaccharidosis IIIB, Annals of laboratory medicine 33 (2013) 221 – 224. doi:10.3343/alm.2013.33.3.221.

[19] J. A. Guio, G. AdolfoGiraldo-Ospina, Impact of enzyme replacement therapy in a patient younger than 2 years diagnosed with Maroteaux-Lamy syndrome (MPS VI), Journal of Inborn Errors of Metabolism and Screening 5 (2017) 1 – 8. doi:10.1177/2326409817718849.

[20] S. Nampoothiri, M. Kappanayil, H. Ravindran, V. Sunitha, Sly disease: mucopolysaccharidosis type VII, Indian pediatrics 45 (2008) 859 – 861.

[21] Y. Yamada, K. Kato, K. Sukegawa, S. Tomatsu, S. Fukuda, S. Emura, S. Kojima, T. Matsuyama, W. S. Sly., N. Kondo, T. Orii, Treatment of MPS VII (Sly disease) by allogeneic BMT in a female with homozygous A619V mutation, Bone Marrow Transplantation 21 (1998) 629 – 634.

[22] A. C. Sewell, J. Gehler, G. Mittermaier, E. Meyer, Mucopolysaccharidosis type VII (β-glucuronidase deficiency): a report of a new case and a survey of those in the literature, Clinical genetics 21 (1982) 366 – 373.

[23] S. S. Ibatova, T. T. Kerimbayev, G. N. Kasenova, Case report of mucopolysaccharidosis type VI with a brief literature review, Journal "Neurosurgery and Neurology of Kazakhstan" 45 (2016) 36 – 41. (In Russian).

[24] N. I. Averianova, T. I. Rudavina, N. A. Domnina, Thrombocytopenia syndrome in a child with type II mucopolysacharidosis, Perm Medical Journal 31 (2014) 110 – 114. (In Russian).

[25] E. K. Ryskulova, E. G. Khusnutdinova, A. E. Babushkin, G. Z. Israfilova, R. M. Mukhametshina, A clinical case of Hurler-Scheie syndrome, Point of view. East-West (2016) 67 – 69. (In Russian).

[26] P. Dubot, F. Sabourdy, G. Plat, C. Jubert, C. Cancés, P. Broué, G. Touati, T. Levade, First report of a patient with MPS type VII, due to novel mutations in GUSB, who underwent enzyme replacement and then hematopoietic stem cell transplantation, International Journal of Molecular Sciences 20 (2019) 5345.

[27] K. Ramphul, S. G. Mejias, Y. Ramphul-Sicharam, Morquio syndrome: a case report, Cureus 10 (2018) e2270.

[28] S. N. Biswas, S. Patra, P. P. Chakraborty, H. Barman, Mucopolysaccharidosis type IVA (Morquio A): a close differential diagnosis of spondylo-epiphyseal dysplasia, BMJ Case Reports 2017 (2017) bcr–2017. doi:10.1136/bcr-2017-221156.

[29] Y. B. Sohn, H. D. Park, S. W. Park, S. U. Kim, S.-Y. Cho, A. R. Ko, C.-S. Ki, S. Y., D.-K. Jin, A Korean patient with Morquio B disease with a novel c.13_14insA mutation in the GLB1 gene, Annals of clinical and laboratory science 42 (2012) 89 – 93.

[30] E. M. Ribeiro, A. C. Brusius-Facchin, S. Leistner-Segal, C. A. da Silva, I. V. Schwartz, Cardiac disease as the presenting feature of mucopolysaccharidosis type IIIA: a case report, Molecular Genetics and Metabolism Reports 1 (2014) 422 – 424.

[31] T. Gurumurthy, S. Shailaja, S. Kishan, M. Stephen, Management of an anticipated difficult airway in Hurler's syndrome, Journal of Anaesthesiology, Clinical Pharmacology 30 (2014) 558 – 561.