Uncertain Model for Classification of Children's Neurological Diseases

Olga Rybytska\textsuperscript{a}, Ivanna Dronyuk\textsuperscript{a,b} and Andriy Pushnyk\textsuperscript{c}

\textsuperscript{a} Lviv Polytechnic National University, Bandery 12, Lviv, 79013, Ukraine
\textsuperscript{b} Jan Dlugosz University in Czestochowa, Waszyngtona 4/8, Czestochowa, 42200, Poland
\textsuperscript{c} Lviv National Medical University, Pekarska str.,69, Lviv,79010, Ukraine

Abstract

The dataset of thirty-three parameters about one hundred children from three to seven years old is considered. The biggest part of children (about 70\%) have some neurological diseases. The dataset contains such parameters: body characteristics, physical and emotional functioning, physical health, communications, relationships in the family, and some results of medical analysis. Standard statistical analysis was realized for the preprocessing. Based on the obtained results the k-mean method was proposed for solving classification tasks. The Silhouette method and Elbow method were realized. Clusters corresponding to groups of indicators of Psychosocial Health Summary Score, Physical Health Summary Score, Total score (0-100\%), and nutritional status of healthy children and children with neurological pathology were obtained. The calculation was made according to the PedsQLTM 2.0 questionnaire and was calculated using a special licensed computer program Dietplan 7. An uncertain model for classification is proposed.

Keywords

Children's neurological diseases, classifications of diseases, improving diagnostics with AI, Silhouette method, Elbow method

1. Introduction

One of the most important signs of the well-being of society is the health of its members, and first of all, the health of children. Modern research data indicate a tendency to decrease in children's health indicators, particularly in Ukraine. Diseases showing the highest growth rates include damage to the central and peripheral nervous system and primarily psychomotor disorders. The count of children's neurological diseases is increasing nowadays. The effectiveness of treatment heavily depends on the timely and correct diagnosis. The development of AI technologies and their application for medicine has made a real breakthrough in the diagnosis and treatment of neurological diseases in children.

It does not seem possible to establish the exact number (in percent of the child-age population) for the purpose of comparisons by regions and the level of welfare of countries using today's existing methods of diagnosis. This is due, in particular, to different possibilities in organizing samples, methods, and criteria of randomization, diagnostic criteria, and instruments used. In some studies, cases of pervasive developmental disorders associated with severe mental retardation, congenital abnormalities of brain development, epilepsy, and prematurity are not taken into account. There is often a lack of registration of children who demonstrate accelerated development, which is not characteristic of their age. This explains why the prevalence of Autism Spectrum Disorders (ASD) in particular in China (6.4 per 10,000) is significantly lower than in Asia, Europe, and North America.
(the average prevalence of ASD is about 1%) [1]. In South Korea, where schoolchildren were studied, this indicator was 2.6% (3.7% among boys and 1.5% among girls) [2].

The rate of incidence of ASD in Ukraine is increasing. Thus, from 2006 to 2017, the number of newly diagnosed cases of the disease increased 8.5 times (by the end of 2017 it was 998) [3]. Over the last decade, a stable trend of growth in the share of children with developmental disorders has been recorded in Ukraine. According to information from the Center for Medical Statistics of the Ministry of Health of Ukraine for 2017, among the children, the incidence and prevalence of central and peripheral nervous system lesions were 40.19, respectively; 0.97 and 15.19 and 0.42 per 1000 child population; behavioral disorders - 24.13 and 3.77 per 1000 child population.

In particular, until 2006, ASD was almost undiagnosed in our country. The rate of incidence of ASD in Ukraine has been increasing since 2006. The number of registered ASD patients increased from 662 in 2005 to 7,491 in 2017. According to a large study conducted by the US Centers for Disease Control and Prevention (CDC), the prevalence of autism spectrum disorders in the US was 1 in 88 children. The disease was five times more common in boys (1 in 54) than in girls (1 in 252).

The prevalence of the syndrome (disorder) of attention deficit hyperactivity disorder (ADHD) is from 4 to 12% worldwide, and of minimal brain dysfunction (MBD) - from 2 to 20%, according to other data - from 3-8% to 30 - 50% of the child population. There is a steady trend towards an increase in the number of children with these psychomotor disorders, as well as with mental retardation (MR). Therefore, it can be argued that the available literature does not fully reflect the current situation [4].

This condition is associated with the lack of mass-planned diagnosis of psychomotor disorders at an early age. This problem must be solved, because early diagnosis and medical, in particular, dietary correction can significantly alleviate and sometimes eliminate symptoms. Therefore, the sooner the disease is detected, the causes of its occurrence, as well as measures to reduce the severity of the disease course (and in some cases complete elimination of symptoms) in children become an extremely important task. The exact mechanisms of the effect of genetic factors on damage to the central and peripheral nervous system are investigated with the help of genomic screening, cytogenetic studies, and evaluation of candidate genes [5].

Also, at the stage of primary medical care during routine preventive examinations at 9, 18, 24, and 36 months, screening of children's development is currently required, during which children with developmental delays are identified. Additional special screening is recommended if the child has previously described risk factors for ASD, diseases that have a high level of comorbidity with ASD. Special screening is also recommended for older children if they have disorders of social interaction and communication. The final diagnosis of ASD was carried out by a specialist doctor based on the diagnostic criteria of ICD-10 (WHO, 1992), and since 2018 - ICD-11. However, the connection between the disease and the nutritional background is mostly not considered today.

Diseases of the central nervous system are currently considered as the result of the interaction of genetic and environmental factors. Diseases that disrupt the work of the brain and physical influences, particularly nutrition, are a risk of morbidity. Ensuring the quality of life is an integral part of the treatment program for children with psychomotor disorders. In children with neurological pathology, treatment tactics are aimed not only at alleviating the child's condition and controlling clinical symptoms but also at improving the quality of his life [6].

Adequate nutritional status is the key to the physiological development of the child's body, in particular the central nervous system, in critical periods of ontogenesis (pregnancy and the first years of life). Nutritional deficiency (excess) during these time periods leads to a violation of the formation of cognitive functions, behavioral reactions, and productivity of the child in various spheres of his activity, which will have a negative impact on both childhood and youth years, as well as on all further adult life of the individual, limiting his opportunities for further development [7]. Thus, the development and formation of the central nervous system critically depend on nutritional support that is adequate to the child's needs in one or another period of life. Sufficient intake of proteins in the child's body, and provision of energy substrates, including glucose, fats (long-chain polyunsaturated fatty acids), iron, zinc, copper, iodine, and folic acid, are especially important. Their absence or excess leads to increased metabolic disorders in the CNS, and structural and epigenetic changes in it, which will have a long-term, often lifelong, impact on the regulation of its functions [8-12].
Children with disorders of psychomotor development against the background of existing neurological disorders and various somatic pathologies become especially vulnerable to deficits (surplus) of nutritional support due to their inherent nutritional features, which significantly complicates their rehabilitation [13]. The sensory profile of children with ASD, with its characteristic hypo- and hyper-reactivity to tactile, olfactory, and visual stimuli, causes a strict selectivity in food due to its appearance, consistency, taste, and temperature. [14-16]. In work [17-19] the systematic review of the research about using modern information technologies for the diagnosis of neural diseases. For medical diagnostics, the classification task is very important. In many modern kinds of research, the method of linear regression [20], Nearest neighbor, Nearest Center [21], randomized classification trees [22], and Bayesian classifier [23] are used for solving classification tasks. In work [24,25,26] the systematic review of the research about using modern information technologies for the diagnosis of neural diseases. For medical diagnostics, the classification task is very important. In many modern kinds of research, the method of linear regression [27], Nearest neighbor, Nearest Center, randomized classification trees, and Bayesian classifier [28] are used for solving classification tasks. The recurrent approximation is used for the tasks of neural network synthesis in the medical therapies control [29].

Therefore, the problem of early and reliable detection of possible problems in the nutritional supply of the child and the development of recommendations for its balancing is an important factor in reducing the severity of the course of CNS diseases, rehabilitation, and improving the quality of life and its expected duration.

2. Research methods

The work uses such research methods as statistical collection of information, cluster analysis, and construction of membership functions of a fuzzy logical model. Obtained statistical information regarding the physical condition of children with absent central nervous system (CNS) diseases (hereinafter "Healthy", abbreviated H) (20 children), attention deficit hyperactivity disorder (ADHD) (13 children), mental retardation (MR) (30 children), autism spectrum disorder (ASD) (17 children), and minimal brain dysfunction (MBD) (10 children) were systematized by age (3 years, 4-5 years, 6-7 years) and gender. The study used indicators (Group 1) of physical condition (Weight (kg), Height (cm), Head Circumference (cm), Chest Circumference (cm), Shoulder Circumference (cm), Hip Circumference (cm), Leg Circumference (cm), BMI (body mass index)), as well as subjective indicators: Physical functioning, Social functioning, Emotional functioning, Functioning of schools/kindergartens; calculated indicators: Psychosocial Health Summary Score, Physical Health Summary Score, Total score (0-100%). The calculation was made according to the PedsQLTM 2.0 questionnaire.

Also, in the clustering of the entire sample, the indicators were taken into account: Awareness, Family involvement, Communication Technical skills, Emotional needs, Overall satisfaction. Total score, Physical functioning, Emotional functioning, Social functioning, Mental functioning, communication, Concern, Everyday activity, Relationships in the family, The Parent HRQL Summary Score, The Family Functioning Summary.

Average values (Group 2) of nutrients obtained from food (Water, Proteins, Fats, Carbohydrates, Energy, Total sugar, Saturated fat, Monounsaturated fat, Polyunsaturated fat, Dietary fiber, Na, K Ca Mg, Phosphorus (P), Fe, Cu, Zn, Cl, Mn, Se, I, Retinol, Vitamins D, E, B1, B2, PP, B6, B12, C, Folates, Pantothenic acid, Biotin was calculated using a special of the licensed computer program Dietplan 7, developed by the British company Forestfield Software Limited (Horsham, UK) (user agreement dated 30.07.19).

Clustering (k-means) was carried out in order to identify groups of similarity based on the indicators of Group 1 and Group 2. The number of clusters [24-26] was determined according to the elbow and silhouette methods. Initial value clusters have a large impact on the clustering model, so algorithms are used for different given initial values. The location measure is used to find the closest points in the clusters to the cluster center. By definition different measures of distance can be created for different clusters. The number of clusters (k) is a defined parameter in K-Means clustering. To
determine the optimal value of various methods, in this research we use the following methods: Elbow method; and Silhouette method.

The Elbow method is an empirical method for determining the optimal number of clusters for a data set. In this method, we select a subset of these k values and then apply K-Means clustering using each k value. We find the average distance of each point in the cluster to its center. Then we calculate the sum of squares S of the distances between the centroids. Now to determine the best number of clusters (k), we plot k against their value S. We choose the value of k from which the plot looks like a straight line.

The silhouette method is unique in the way it interprets and checks for consistency in data clusters. This method provides a concise graphical representation of how well each object is classified. The size of the silhouette is given by a strictly mathematical formula for each cluster of points using the minimum value of the average distance of a given point in relation to other points of the cluster. Silhouette magnitude is a measure of how similar an object is to its own cluster (cohesion) and different from other clusters (separation). The silhouette ranges from -1 to +1, where a high value indicates that the object matches well with its own cluster and poorly matches neighboring clusters. If one object has a high value, then the clustering configuration is appropriate. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters. The silhouette can be calculated using any distance metric. In the presented research the Euclidian metric is used.

3. Results of data clustering of patients aged 3-7 years according to all indicators of Groups 1 and 2.

The results of clustering the data of the entire sample into 3 clusters according to all indicators, in particular of the 1st and 2nd groups, made it possible to confirm the fact that the composition of nutrients is associated with diseases of the central nervous system. Thus, clusters 0 and 1 included both healthy and sick patients. At that time, only patients with ASD, MR and ADHD were in cluster 2. The value of the centroids is shown in Fig. 1. The significant difference between the content in the diet of such nutrients as Water, Carbohydrates, Energy, Na, K, Ca, P, Cl, Retinol, and Folates gave the impetus for more detailed studies. Since early diagnosis and rehabilitation is the most effective, it was decided to carry out more detailed research with data on children aged 3 years. It is obvious that age differences played a significant role in the clustering of the sample of the 3-7 age group.

![Figure 1: The value of the centroids of clusters 0-2 of all sample participants by all indicators](image-url)
4. Results of clustering of patients in the age category of 3 years.

The results of clustering Fig. 2,3 were obtained in relation to groups of similarity according to the indicators of Group 1 and centroid values (A group of 3-year-old patients according to the indicators of Group 1 was clustered into 5 clusters (0-cluster - 4-cluster). Value centroid indicators for each cluster are presented in Table 2. Both healthy and sick patients were included in clusters 0 and 3. Patients with MR, ASD, and ADHD were included in clusters 1,2,4.

Figure 2: Results of clustering indicators of Group 1 (Emotional functioning – Social functioning)

Figure 3: Results of clustering of indicators of Group 1 (Psychosocial Health Summary Score - Physical Health Summary Score)
The obtained results drew attention to the non-synchronicity of the trends to increase or decrease in the values of individual indicators of Group 1 with the presence (absence) of CNS diseases. Moreover, when comparing the values of such an indicator as Weight in Ukrainian boys corresponds to the norms (Tabl. 1,2), while Height and Head Circumference are deviated in the direction of decrease from the norm. The percentage of head circumference to height corresponds to the norm. At the same time, the breast circumference, as well as the relative index to height, exceeds the normal values [27].

Studies have shown that in patients with MR and ASD (to a lesser extent, ADHD), deviations in many indicators of physical condition are extremely small and extremely large. Studies have shown that in patients with MR and ASD (to a lesser extent, ADHD), the deviations of many indicators of physical condition from the norm are significant, that is, their values are extremely small or extremely large. Thus, according to the Weight indicator m, healthy children have fixed values 13.5<m<15.5 (boys) with an average value of the norm of 14.3; and 13.5<m<17.5 (girls) with an average value of the norm of 13.9. For patients with diseases for boys 13.5<m or 14.5<m<21.5 and for girls – m<13.5. According to the Height indicator h, the corresponding indicators are: healthy (93.5<h<95.5 (boys), 90.5<h<94.5 (girls). For patients with diseases respectively h<93.5, or 95.5<h<105.5 (boys) and 89.5<h or 94.5<h<97.5 (girls) with an average value of the norm of 96.1 (boys) and 95.1 (girls), respectively. A similar situation with regard to the extreme smallest and extreme largest values in children with diagnosed psychomotor disorders is observed according to the Head Circumference, Shoulder Circumference, Hip Circumference, and Leg Circumference indicators. Regarding the BMI of boys, there is a deviation in critically low or critically high values, but for girls, in children with diseases, this indicator is significantly lower than the norm.

Table 1

<table>
<thead>
<tr>
<th>Indicators of physical development and socio-emotional functioning</th>
<th>cluster no</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (kg)</td>
<td>15.66667</td>
<td>14.5</td>
<td>14.83333</td>
<td>16</td>
<td>14.71429</td>
<td></td>
</tr>
<tr>
<td>Height (cm)</td>
<td>94.66667</td>
<td>93.125</td>
<td>97.83333</td>
<td>94.83333</td>
<td>98.42857</td>
<td></td>
</tr>
<tr>
<td>Head circumference (cm)</td>
<td>49.16667</td>
<td>49.25</td>
<td>49.16667</td>
<td>50.5</td>
<td>49.85714</td>
<td></td>
</tr>
<tr>
<td>Chest circumference (cm)</td>
<td>52.83333</td>
<td>52.5</td>
<td>53.33333</td>
<td>54.33333</td>
<td>52.71429</td>
<td></td>
</tr>
<tr>
<td>Shoulder girth (cm)</td>
<td>17.83333</td>
<td>17.125</td>
<td>17.83333</td>
<td>17.83333</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Hip circumference (cm)</td>
<td>27.66667</td>
<td>25.75</td>
<td>26.5</td>
<td>29</td>
<td>28.28571</td>
<td></td>
</tr>
<tr>
<td>Leg circumference (cm)</td>
<td>22.83333</td>
<td>21</td>
<td>22.33333</td>
<td>23.5</td>
<td>22.85714</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>17.51667</td>
<td>16.75125</td>
<td>15.49667</td>
<td>17.70667</td>
<td>15.15857</td>
<td></td>
</tr>
<tr>
<td>Physical functioning</td>
<td>50.56667</td>
<td>77.425</td>
<td>87.51667</td>
<td>87</td>
<td>74.48571</td>
<td></td>
</tr>
<tr>
<td>Social functioning</td>
<td>45</td>
<td>80</td>
<td>84.16667</td>
<td>52.5</td>
<td>47.14286</td>
<td></td>
</tr>
<tr>
<td>Emotional functioning</td>
<td>40</td>
<td>65.625</td>
<td>57.5</td>
<td>66.66667</td>
<td>64.28571</td>
<td></td>
</tr>
<tr>
<td>Psychosocial Health Summary Score</td>
<td>9</td>
<td>14.2</td>
<td>73.56667</td>
<td>13.65</td>
<td>55.22857</td>
<td></td>
</tr>
<tr>
<td>Physical Health Summary Score</td>
<td>50.56667</td>
<td>77.425</td>
<td>87.51667</td>
<td>87</td>
<td>74.47857</td>
<td></td>
</tr>
</tbody>
</table>

The authors believe that when building a vague model of support for the diagnosis of children with psychomotor disorders, a significant deviation from the norm for at least two indicators of Group 1 should serve as a basis for further examination regarding the indicators of Group 2.
In relation to Group 2 indicators of nutritional support, clustering was carried out into 3 clusters (Fig.4-9). According to Group 2 indicators, all healthy children (except one) fell into one (2nd) cluster.

Figure 4: Setting the number of clusters for indicators of Group 2 by the elbow method

Figure 5: Results of clustering by indicators Mn, Se. Here and in the following figures, the image is used:
- elements of the 0-cluster;
- the centroid of the 0-cluster;
- elements of the 1-cluster;
- the centroid of the 1-cluster;
- element and the centroid of the 2-cluster
Figure 6: Results of clustering by indicators Zn, Cl

Figure 7: Results of clustering by indicators of Fats, Carbohydrates
Figure 8: Results of clustering by indicators Vitamin C, Biotin

Figure 9: Results of clustering by indicators B12, Folate

The obtained centroid values for all indicators of nutritional indicators are presented in Table 2
Table 2.
Values of centroids for all indicators of Group 2

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Cluster no 0</th>
<th>Cluster no 1</th>
<th>Cluster no 2</th>
<th>Cluster no 3</th>
<th>Cluster no 4</th>
<th>Cluster no 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>1195,3</td>
<td>1107,7</td>
<td>918,3</td>
<td>783,2</td>
<td>730,8</td>
<td>519,4</td>
</tr>
<tr>
<td>Proteins</td>
<td>64,42</td>
<td>103,5</td>
<td>44,36</td>
<td>233,5</td>
<td>573,0</td>
<td>155,1</td>
</tr>
<tr>
<td>Fats</td>
<td>73,66</td>
<td>106,0</td>
<td>48,82</td>
<td>1156,7</td>
<td>3472,0</td>
<td>786,6</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>248,3</td>
<td>590,0</td>
<td>183,0</td>
<td>9,41</td>
<td>70,70</td>
<td>6,34</td>
</tr>
<tr>
<td>Energy</td>
<td>1853,5</td>
<td>3581,61</td>
<td>1300,0</td>
<td>1,22</td>
<td>3,15</td>
<td>1,03</td>
</tr>
<tr>
<td>Total sugar</td>
<td>93,09</td>
<td>151,3</td>
<td>72,98</td>
<td>8,10</td>
<td>18,02</td>
<td>5,62</td>
</tr>
<tr>
<td>Saturated fat</td>
<td>28,13</td>
<td>25,63</td>
<td>21,07</td>
<td>3394,20</td>
<td>2953,0</td>
<td>1532,0</td>
</tr>
<tr>
<td>Monounsaturated fat</td>
<td>25,72</td>
<td>25,80</td>
<td>14,61</td>
<td>3,12</td>
<td>12,15</td>
<td>2,34</td>
</tr>
<tr>
<td>Polyunsaturated fat</td>
<td>16,43</td>
<td>10,00</td>
<td>5,73</td>
<td>38,49</td>
<td>30,00</td>
<td>31,85</td>
</tr>
<tr>
<td>Dietary fiber</td>
<td>15,37</td>
<td>15,70</td>
<td>11,46</td>
<td>111,6</td>
<td>45,30</td>
<td>78,71</td>
</tr>
<tr>
<td>Na</td>
<td>2316,05</td>
<td>2053,30</td>
<td>1054,26</td>
<td>458,8</td>
<td>262,7</td>
<td>573,6</td>
</tr>
</tbody>
</table>

The obtained results make it possible to obtain weighting coefficients, as well as to build dependence functions for a fuzzy logic model [28] and AI applications [29] and support the diagnosis and corrective diet of children with psychomotor disorders.

5. Conclusion

The research approved that somatic health, physical development, and regulatory mechanisms suffer in children with neurological diseases. The data obtained by testing the parents of the children under study served as the initial data. Therefore, the indicators used for the purpose of cluster analysis, which are calculated according to the methods specified above, are largely subjective. In order to obtain more accurate input data, it is necessary to conduct a more thorough explanation of the scale of each of the indicators to the persons who participate in the assessment of the value of the studied characteristics, of all groups of indicators. Also, the authors believe that in order to reduce subjectivity, it is necessary to separately involve objective indicators, namely the chemical composition of blood.

Children with disorders of psychomotor development against the background of existing neurological disorders and various somatic pathologies become especially vulnerable to the deficiency (surplus) of nutritional support due to their inherent nutritional characteristics. The study examined a dataset collected by the authors. The data set was created according to the PedsQLTM 2.0 questionnaire and processed using the Dietplan 7 computer program.

The presented study shows that the k-means clustering method can be an effective tool for better decision-making by doctors in the diagnosis of pediatric neurological diseases. Both considered methods can be used: silhouette and elbow. The study shows that it is advisable to divide the data set of the first group of indicators into five clusters, and according to the group of indicators that characterize the composition of the obtained nutrients - into three clusters according to practical experiments. The next step of the research will be to be involved in the study of blood indicators, as well as the history of parental diseases and the course of pregnancy with the aim of implementing artificial intelligence methods to support the diagnosis of children's neurological diseases.
6. Acknowledgements

The authors would like to thank the Armed Forces of Ukraine for providing security to perform this work. This work has become possible only because of the resilience and courage of the Ukrainian Army.

7. References


[17] Previously Marzena Szkodo MOR, Mica M, Caruso A, Fulceri F, Fazio M, Scattoni M.L. Technologies to support the diagnosis and/or treatment of neurodevelopmental disorders: A


