Stages of Cluster Analysis in the Diagnosis of Lyme Disease in Children

Vasyl Martsenyuk[1], Svitlana Nykytyuk[2], Yuri Palaniza[3], Oksana Bahrii-Zaiats[2], Sofiia Sverstiuk[4]

[1] *University of Bielsko-Biala, Willowa St. 2, Bielsko-Biala, 43-300, Poland*
[2] *I. Horbachevsky Ternopil National Medical University, 12 Rus'ka St., Ternopil, 46001, Ukraine*
[3] *Ternopil Ivan Puluj National Technical University, Ternopil, Ukraine*
[4] *Ternopil National Pedagogical University, 2 Maxyma Kryvonosa St., Ternopil, 46027, Ukraine*

Abstract

Lyme borreliosis (LB) is the most common vector-mediated disease caused by spirochetes of the Borrelia burgdorfery sensu lato(s.l) complex, which are vectored by Ixodes ticks. The disease tends to be prolonged and chronic. The aim of this study was to develop a multifactorial model for predicting the severe course of Lyme borreliosis in children and to evaluate its effectiveness using Claster analysis and PCA methods. Silhoutte scor method and the Calinski-Horabasz score methods were used for developing mathematical prognosis of severe forms LB. To build a prognostic model of Claster analysis, 143 patients with Lyme disease were examined using multivariate regression analysis who were admitted to the Ternopil Regional Children's Hospital. The model was clustered based on the coefficients. The sum of points from 1 to 10 indicates a mild form of the disease, from 10 to 20 - a severe form of the disease. Therefore, the result is that the Localised form is mild and severe and the Disseminated form is divided into mild and severe.

Keywords

Lyme disease, children, Claster analysis, PCA methods

1. Introduction

Lyme borreliosis (LB) is the most common vector-mediated disease caused by spirochetes of the Borrelia burgdorfery sensu lato(s.l) complex, which are vectored by Ixodes ticks. The pathogen affects the skin, nervous system, musculoskeletal system, heart, and eyes. The disease tends to be prolonged and chronic.

The clinical picture of Lyme disease [1] includes early localized, early disseminated, and late disseminated stages. In the early period, stage I of localized infection is distinguished when the pathogen enters the skin after a tick bite. The early localized stage of the disease begins 3-30 days after a tick bite. Diseases caused by B. burgdorferii sensu stricto are usually inflammatory in nature and more often cause single or multiple EM, arthritis and carditis. In areas with endemic infection, previous subclinical infection with seroconversion is common, and symptoms of seropositive patients may be incidental [2, 3, 4]. Patients with active Lyme disease almost always have objective signs of infection (erythema migrans (ME), facial nerve palsy, arthritis). Nonspecific symptoms usually accompany these specific signs but are almost never the only evidence of Lyme disease [1].

The division into stages is quite arbitrary and largely based on clinical manifestations and time since infection. It should be noted that the disease can gradually move from one stage to another or bypass any of them, as well as appear for the first time at any stage without the presence of the previous one.

Proceedings ITTAP'2023: 3rd International Workshop on Information Technologies: Theoretical and Applied Problems, November 22–24, 2023, Ternopil, Ukraine, Opole, Poland
EMAIL: vmartsenyuk@ath.bielsko.pl (A. 1); androx@tdmu.edu.ua (A. 2); palyanytsa_y@tntu.edu.ua (A. 3); bagrijzayats@tdmu.edu.ua (A. 4); khrystynasofia@gmail.com (A. 5)
ORCID: 0000-0001-5622-1038 (A. 1); 0000-0003-3146-9664 (A. 2); 0000-0002-8710-953X (A. 3); 0000-0002-5533-3561 (A. 4); 0000-0001-5595-4918 (A. 5)

Only children with localized form (erythema migrans) and disseminated form were admitted to the Ternopil children's regional hospital.

According to European authors, LB manifests itself as a skin disease in 80-90 % of patients, while lesions of other organs and systems are reported in about 10-20 % of patients [5-7]. Insufficient consideration of the epidemiological history, hereditary and allergic history leads to misdiagnosis and possible errors in the treatment of the disease. Hematogenous spread of the bacteria occurs within days or weeks after a tick bite; the host's immune response often leads to specific symptoms [1].

The **Aim** of the study was to analyse clinical and immunological cases of the disease, to identify the main markers leading to chronicity of the disease, to optimise the diagnostic search using mathematical analysis, to develop a multifactorial model for predicting the severe course and damage to organs and systems in Lyme borreliosis in children and to evaluate its effectiveness using Claster analysis and PCA methods.

2. Materials and methods

In the research difeerent materials and methods were used, such as general clinical (questionnaire), objective examination, immunological-ELISA (total antibodies of classes M and G to antigens of the Borrelia burgdorfery complex sensu lato(s.l), immunoblot specific antibodies of classes M and G to B. burgdorfery sensu lato (s. l), epidemiological (unified questionnaire), molecular biological, statistical (methods of parametric and non-parametric statistics with the calculation of Student's criteria using the computer programs "Microsoft Office Excel" and "Statistica").

To build a prognostic model of Claster analysis, 143 patients with Lyme disease were examined using multivariate regression analysis who were admitted to the Ternopil Regional Children's Hospital. The study was conducted in the laboratory of the Center for the Study of LB and Other Tick-Borne Infections. 143 children with Lyme disease were examined (aged 13±3 years) from 1 year to 18 years, including 74 boys and 70 girls. Groups of patients: 80 children with erythema migrans, 16 with Lyme arthritis, and 27 with nervous system damage due to Lyme disease and non erythema forms 20 children.

The study participants answered the questions of a single international questionnaire. The detection of Borrelia in ticks was performed by the polymerase chain reaction PCR method [4].

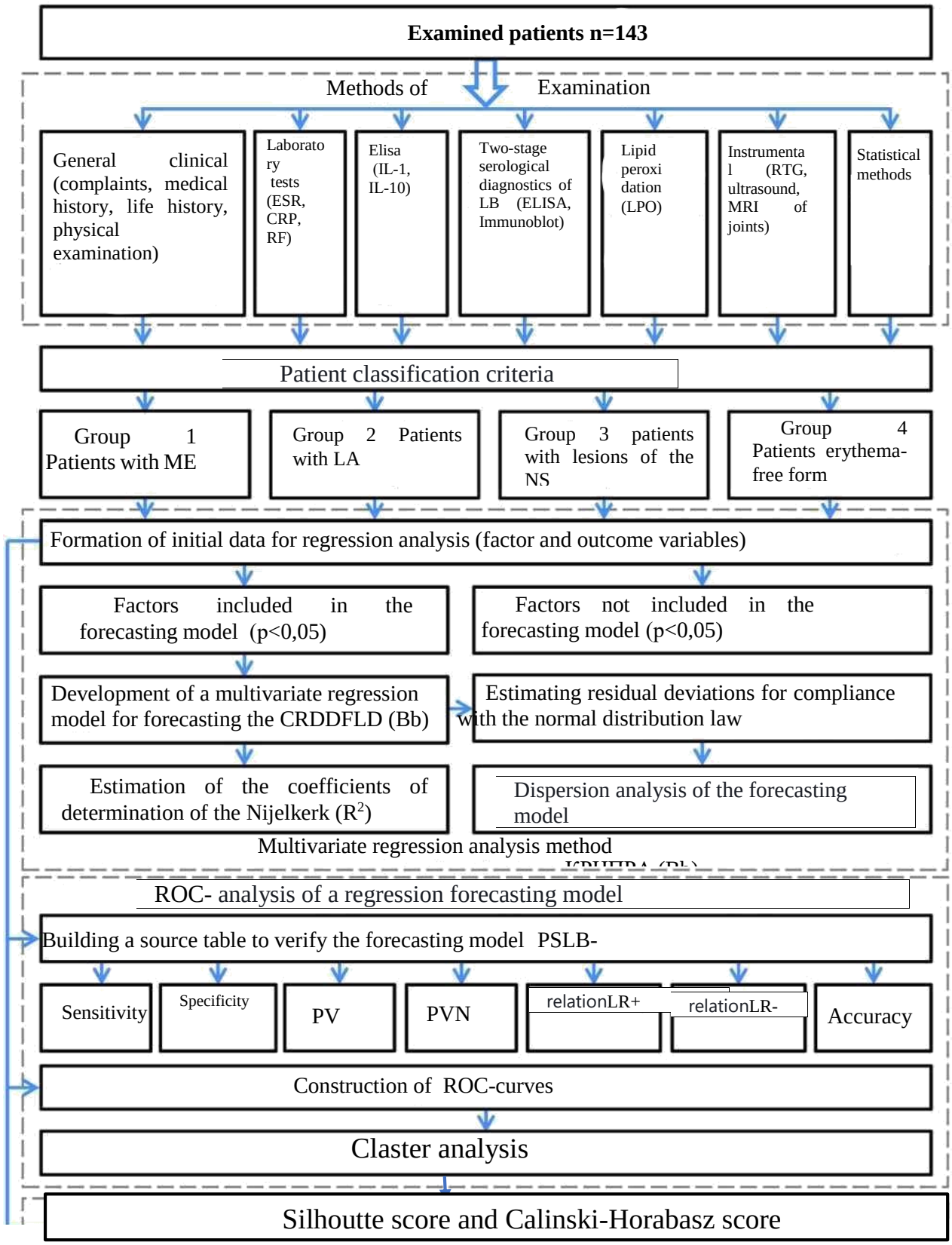


Figure 1: The design of the study

1. Inclusion criteria:

- epidemiological (residence in an endemic area);
- Clinical complaints of patients (erythematous skin lesions, cardiovascular system lesions, Lyme arthritis, clinical signs of nervous system lesions);

- infectious confirmation of the diagnosis: a two-stage study.

The following notation describes the linkages used by the various methods:

- Cluster r is formed from clusters p and q.
- n_r is the number of objects in cluster r.
- x_{ri} is the ith object in cluster r.
- Single linkage, also called nearest neighbor, uses the smallest distance between objects in the two clusters.

$$d(\mathrm{r},\mathrm{s}) = \min(\mathrm{dist}(\mathrm{x_{ri}},\mathrm{x_{sj}})), \mathrm{i} \in (\mathrm{i},\ldots,\mathrm{n}_r), \mathrm{j} \in (1,\ldots,\mathrm{n}_s)$$

- Complete linkage, also called farthest neighbor, uses the largest distance between objects in the two clusters.

$$d(\mathrm{r},\mathrm{s}) = \max\left(\mathrm{dist}\left(\mathrm{x_{ri}},\mathrm{x_{sj}}\right)\right), \mathrm{i} \in (\mathrm{i},\ldots,\mathrm{n}_r), \mathrm{j} \in (1,\ldots,\mathrm{n}_s)$$

- Average linkage uses the average distance between all pairs of objects in any two clusters.

$$d(\mathrm{r},\mathrm{s}) = \frac{1}{\mathrm{n}_r\mathrm{n}_s}\sum_{i=1}^{\mathrm{n}_r}\sum_{j=1}^{\mathrm{n}_s}\mathrm{dist}(\mathrm{x_{ri}},\mathrm{x_{sj}})$$

- Centroid linkage uses the Euclidean distance between the centroids of the two clusters.

$$d(\mathrm{r},\mathrm{s}) = \|\bar{\mathrm{x}}_\mathrm{r} - \bar{\mathrm{x}}_\mathrm{s}\|_2,$$

where

$$\bar{\mathrm{x}}_\mathrm{r} = \frac{1}{\mathrm{n}_r}\sum_{i=1}^{\mathrm{n}_r}\mathrm{x_{ri}}$$

- Median linkage uses the Euclidean distance between weighted centroids of the two clusters.

$$d(\mathrm{r},\mathrm{s}) = \|\tilde{\mathrm{x}}_\mathrm{r} - \tilde{\mathrm{x}}_\mathrm{s}\|_2,$$

d(r,s)=‖‖˜xr−˜xs‖‖2,

where $\tilde{x}_r$ and $\tilde{x}_s$ are weighted centroids for the clusters r and s. If cluster r was created by combining clusters p and q, $\tilde{x}_r$ is defined recursively as

$$\tilde{\mathrm{x}}_\mathrm{r} = \frac{1}{2}\left(\tilde{x}_p + \tilde{x}_q\right)$$

- Ward's linkage uses the incremental sum of squares, that is, the increase in the total within-cluster sum of squares as a result of joining two clusters. The within-cluster sum of squares is defined as the sum of the squares of the distances between all objects in the cluster and the centroid of the cluster. The sum of squares metric is equivalent to the following distance metric $d(r,s)$, which is the formula linkage uses.

$$d(r,s) = \sqrt{\frac{2n_r n_s}{(n_r+n_s)}}\|\bar{x}_r - \bar{x}_s\|_2,$$

where

- o $\|\ \ \|_2$ is the Euclidean distance.
- o $\bar{x}_r$ and $\bar{x}_s$ are the centroids of clusters r and s.
- o n_r and n_s are the number of elements in clusters r and s.

In some references, Ward's linkage does not use the factor of 2 multiplying nrns. The linkage function uses this factor so that the distance between two singleton clusters is the same as the Euclidean distance.

- Weighted average linkage uses a recursive definition for the distance between two clusters. If cluster r was created by combining clusters p and q, the distance between r and another cluster s is defined as the average of the distance between p and s and the distance between q and s.

$$d(r,s) = \frac{(d(p,q) + d(q,s))}{2}$$

A linkage is the distance between two clusters.

3. Cluster analysis

Table 1

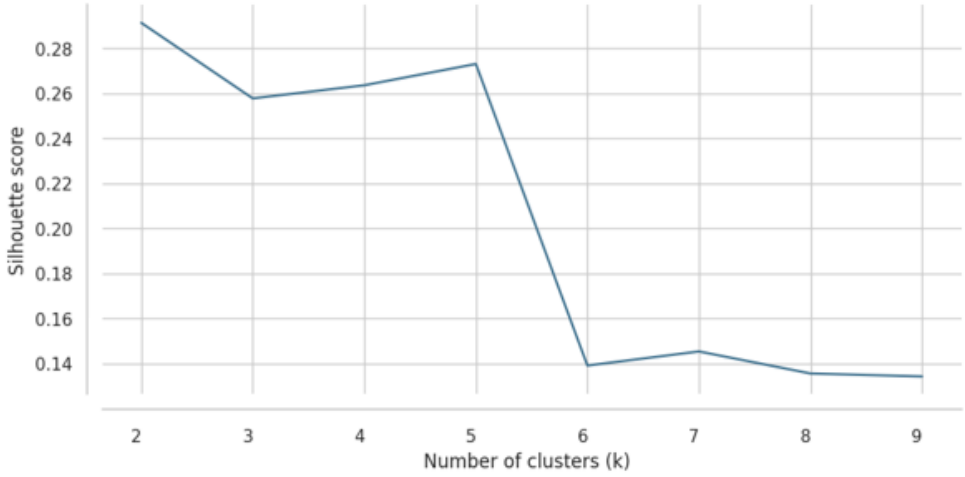
Identification of potential risk factors for localized and disseminated forms of LB

Variable of the model	Name of factor
x_1	Age
x_2	Sex
x_3	Causative agent of infection
x_4	Number of ticks
x_5	Affected system
x_6	IgM (RU/ml)
x_7	IgG (RU/ml)
x_8	Ig G (in dynamics)
x_9	VLsE IgM
x_{10}	P41 IgM
x_{11}	P39 IgM
x_{12}	OspC Ba (Borrelia *afzelii*)
x_{13}	OspC Bb (Borrelia *burgdorferri*)
x_{14}	OspC Bg (Borrelia *garinii*)
x_{15}	IgM
x_{16}	VLsE (Borrelia *afzelii*) IgG
x_{17}	VLsE (Borrelia *burgdorferri*) IgG
x_{18}	VLsE (Borrelia *garinii*) IgG
x_{19}	Lipid Ba (Borrelia *afzelii*) IgG
x_{20}	Lipid Bb (Borrelia *burgdorferri*) IgG
x_{21}	P83 IgG
x_{22}	P41
x_{23}	P39 IgG
x_{24}	OspC (B. *afzelii*) IgG
x_{25}	P58IgG
x_{26}	P21IgG
x_{27}	P20IgG
x_{28}	P19IgG
x_{29}	P18IgG
x_{30}	IgG

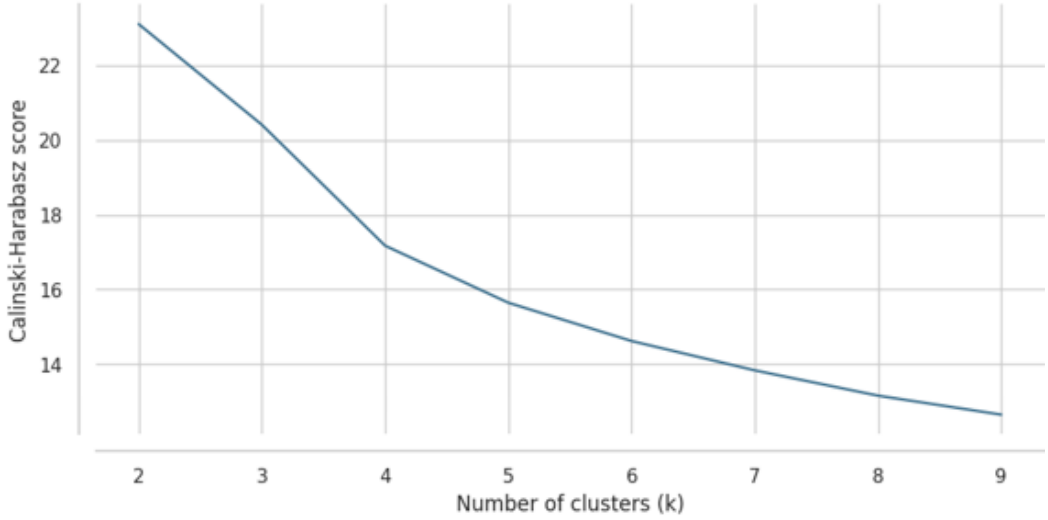
Using multivariate regression analysis, we analysed 28 probable factors for the onset and progression of Lyme borreliosis.

After conducting the classical method of determining the number of clusters in our general sample, the classical approach was to use 2 methods: Silhoutte score and Calinski-Horabasz score.

We can see that the result of the first method (Silhoutte score) tends to be closer to 3 clusters, while the second method (Calinski-Horabasz) fluctuates between three and four clusters, although it is more inclined to four clusters in our overall sample (larger break at four clusters).



a)



b)

Figure 2: Results of cluster analysis using the Silhoutte score method (a) and the Calinski-Horabasz score method (b)

Further cluster analysis is carried out by analysing two principal components. Figure 3 shows a tree dendrogram.

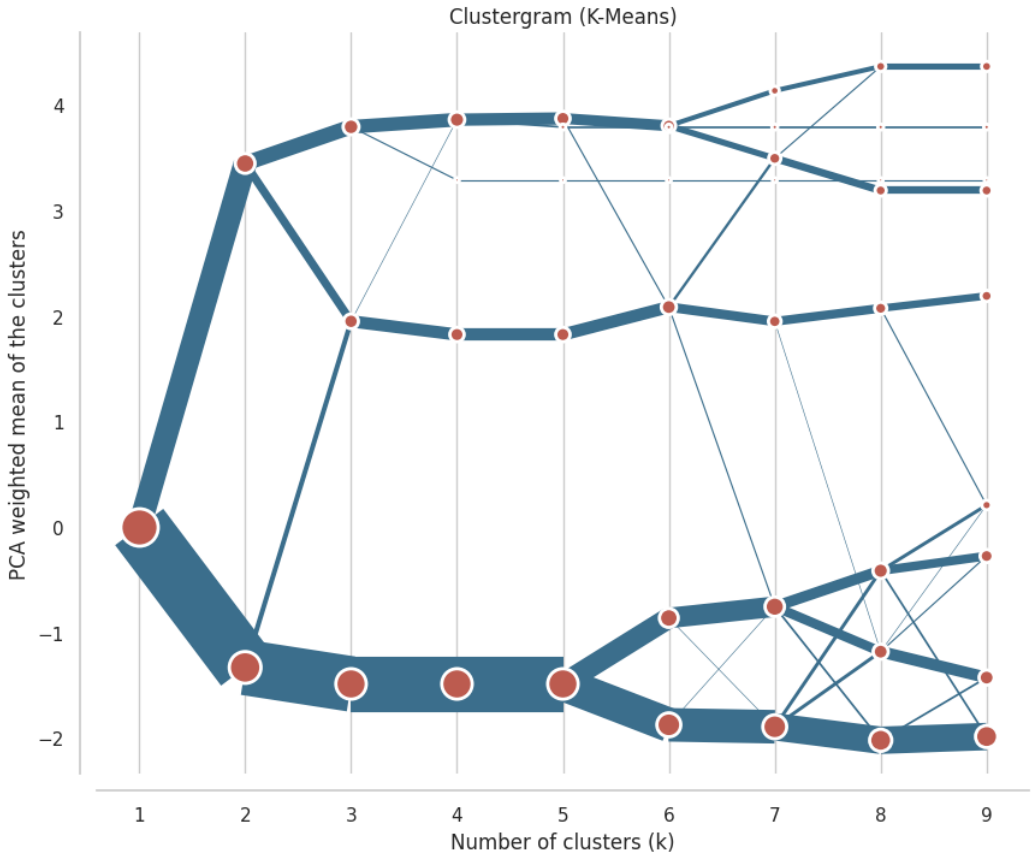


Figure 3: Tree dendrogram

According to the first message (Figure 3), the distance between the centres of the clusters is shown on the -Y axis, and the number of clusters or iteration numbers is shown on the X axis. We find the centre of mass (0 on the Y-axis) of the cluster. We start from this point 0, and start counting the number of clusters from this point.

We find the only centre of mass whose standard deviation to each of the points is maximum.We set up two classes.

Our tree has branched into two branches, in particular, in the second iteration (2 on the x-axis), our data is branched into two branches: one thinner and longer branch branches upwards, and a shorter and thicker branch is placed at the bottom of the figure. The thickness of the branches is proportional to the number of patients in the respective cluster. In particular, we observe that the thickness of the bottom cluster is twice the thickness of the top cluster. In this analysis, we study the number of similar groups.

The next step is to iterate with three clusters. We observe further potential branching of the branches and the tree clusterogram.

Analysing Figure 3 along the vertical positional line numbered 3 on the x-axis, which is parallel to the y-axis.

- Computing linkage (Y) can be slow when y is a vector representation of the distance matrix. For the 'centroid', 'median', and 'ward' methods, linkage checks whether y is a Euclidean distance. Avoid this time-consuming check by passing in X instead of Y.
- The 'centroid' and 'median' methods can produce a cluster tree that is not monotonic. This result occurs when the distance from the union of two clusters, r and s, to a third cluster is less than the distance between r and s. In this case, in a dendrogram drawn with the default orientation, the path from a leaf to the root node takes some downward steps. To avoid this result, use another method. This figure shows a nonmonotonic cluster tree.

To evaluate the significance of the influence of the factor attributes, a stepwise multivariate regression analysis was performed using Statistica 10.0. Initially, a correlation matrix was obtained, in which the absence of pairwise correlation coefficients greater than 0.7 was established. Thus, the absence of multicollinear factors for predicting the severity of LD gives grounds to use all 28 of the

above factors to build a regression model. The next step was to calculate the regression coefficients "b" (Beta), which reflect for each selected factor the relationship of influence on the severity of Lyme borreliosis in the examined patients. The result of obtaining significant factors for this coefficient in multivariate regression analysis in Statistica 10.0 is shown in Figure 4.

N=120	Regression Summary for Dependent Variable: RCDLB_B (1 in 1) R= ,99415581 R?= ,98834577 Adjusted R?= ,98441738 F(30,89)=251,59 p<0,0000 Std.Error of estimate: ,57960					
	b*	Std.Err. of b*	b	Std.Err. of b	t(89)	p-value
Intercept			-1,02657	0,306333	-3,35114	0,001183
Age	0,197909	0,013937	0,20971	0,014768	14,20074	0,000000
Sex	0,087996	0,014586	0,82107	0,136096	6,03303	0,000000
Causative agent of infection	0,158662	0,013582	1,02726	0,087939	11,68146	0,000000
Number of ticks	0,050376	0,015034	0,75351	0,224871	3,35085	0,001184
Affected system	0,202755	0,015146	0,99810	0,074559	13,38676	0,000000
IgM (RU/ml)	0,157043	0,013937	0,08692	0,007714	11,26815	0,000000
IgG(RU/ml)	0,193715	0,014733	0,11686	0,008888	13,14827	0,000000
Ig G (in dynamics)	0,015445	0,014036	0,14966	0,136001	1,10041	0,274122
VLsE IgM	0,007689	0,012840	0,39109	0,653058	0,59885	0,550792
P41 IgM	0,102231	0,018464	1,00921	0,182275	5,53676	0,000000
P39 IgM	0,069012	0,014356	1,21147	0,252008	4,80728	0,000006
OspC Ba (Borrelia afzelii) IgM	0,105884	0,024465	1,26526	0,292349	4,32792	0,000039
OspC Bb (Borrelia burgdorferri)	0,067869	0,021341	1,04603	0,328920	3,18019	0,002026
OspC Bg (Borrelia garinii) IgM	0,085734	0,024711	1,06368	0,306578	3,46951	0,000806
IgM	0,072928	0,021779	0,80751	0,241151	3,34855	0,001192
VLsE (Borrelia afzelii) IgG	0,094616	0,025186	1,17388	0,312471	3,75677	0,000307
VLsE (Borrelia burgdorferri) IgG	0,095609	0,023215	1,04521	0,253783	4,11850	0,000085
VLsE (Borrelia garinii) IgG	0,081841	0,023195	1,03659	0,293781	3,52843	0,000664
Lipid Ba (Borrelia afzelii) IgG	0,024137	0,014567	0,62173	0,375219	1,65698	0,101044
Lipid Bb (Borrelia burgdorferri) IgG	0,019028	0,016189	0,56352	0,479442	1,17536	0,242984
P83 IgG	0,055239	0,017133	0,96970	0,300758	3,22417	0,001767
P41	0,138898	0,020558	1,28517	0,190213	6,75648	0,000000
P39 IgG	0,062252	0,020826	1,15391	0,386031	2,98916	0,003616
OspC (B. afzelii) IgG	0,073155	0,021842	0,72714	0,217107	3,34924	0,001190
P58 IgG	0,028588	0,019602	0,52991	0,363341	1,45844	0,148238
P21 IgG	0,081749	0,014673	1,36760	0,245477	5,57117	0,000000
P20 IgG	0,019821	0,015815	0,71587	0,571205	1,25327	0,213390
P19 IgG	0,040226	0,015905	1,45287	0,574465	2,52907	0,013198
P18 IgG	0,014557	0,017001	0,25555	0,298447	0,85625	0,394158
IgG	0,053191	0,023860	0,58896	0,264191	2,22931	0,028311

Figure 4: The result of obtaining significant factors for predicting CRDDFLB- in multivariate regression analysis in Statistica 10.0

N=121	Regression Summary for Dependent Variable: RCDLB_B (1 in 1) R= ,99312677 R?= ,98630078 Adjusted R?= ,98305251 F(23,97)=303,64 p<0,0000 Std.Error of estimate: ,61081					
	b*	Std.Err. of b*	b	Std.Err. of b	t(97)	p-value
Intercept			-1,04795	0,306060	-3,42400	0,000906
Age	0,194991	0,014344	0,20935	0,015400	13,59392	0,000000
Sex	0,089866	0,014627	0,84820	0,138063	6,14361	0,000000
Causative agent of infection	0,151474	0,013749	0,99446	0,090268	11,01681	0,000000
Number of ticks	0,045341	0,013872	0,68806	0,210512	3,26854	0,001496
Affected system	0,209334	0,014574	1,03436	0,072014	14,36322	0,000000
IgM (RU/ml)	0,158079	0,013727	0,08879	0,007710	11,51555	0,000000
IgG(RU/ml)	0,195187	0,014583	0,11915	0,008901	13,38490	0,000000
P41 IgM	0,096095	0,018397	0,95447	0,182725	5,22356	0,000001
P39 IgM	0,064102	0,013948	1,14149	0,248385	4,59564	0,000013
OspC Ba (Borrelia afzelii) IgM	0,129186	0,023997	1,53841	0,285765	5,38349	0,000001
OspC Bb (Borrelia burgdorferi)	0,062034	0,019392	0,96976	0,303145	3,19898	0,001864
OspC Bg (Borrelia garinii) IgM	0,085913	0,023024	1,08073	0,289626	3,73149	0,000321
IgM	0,059896	0,021575	0,66361	0,239040	2,77614	0,006602
VLsE (Borrelia afzelii) IgG	0,093759	0,025791	1,15674	0,318188	3,63540	0,000446
VLsE (Borrelia burgdorferi) IgG	0,093741	0,021737	1,02606	0,237926	4,31251	0,000039
VLsE (Borrelia garinii) IgG	0,093045	0,022653	1,17044	0,284964	4,10732	0,000084
P83 IgG	0,064653	0,016276	1,15132	0,289833	3,97235	0,000137
P41	0,139957	0,019134	1,30830	0,178859	7,31468	0,000000
P39 IgG	0,082135	0,016384	1,46262	0,291755	5,01319	0,000002
OspC (B. afzelii) IgG	0,078837	0,020260	0,78818	0,202551	3,89125	0,000183
P21 IgG	0,093464	0,014496	1,58605	0,245987	6,44769	0,000000
P19 IgG	0,058826	0,013941	2,15584	0,510892	4,21977	0,000055
IgG	0,038648	0,021501	0,43372	0,241289	1,79751	0,075365

Figure 5: The result of obtaining significant factors for predicting CRDDFLB- in multivariate regression analysis in Statistica 10.0 without factors IgG

N=121	Regression Summary for Dependent Variable: RCDLB_B (1 in 1) R= ,99289700 R?= ,98584446 Adjusted R?= ,98266668 F(22,98)=310,23 p<0,0000 Std.Error of estimate: ,61772					
	b*	Std.Err. of b*	b	Std.Err. of b	t(98)	p-value
Intercept			-1,07264	0,309212	-3,46895	0,000778
Age	0,196823	0,014470	0,21131	0,015535	13,60250	0,000000
Sex	0,090540	0,014788	0,85457	0,139580	6,12247	0,000000
Causative agent of infection	0,155225	0,013744	1,01908	0,090232	11,29405	0,000000
Number of ticks	0,044496	0,014021	0,67524	0,212772	3,17356	0,002012
Affected system	0,204494	0,014486	1,01044	0,071576	14,11711	0,000000
IgM (RU/ml)	0,158270	0,013882	0,08889	0,007797	11,40076	0,000000
IgG(RU/ml)	0,198349	0,014640	0,12108	0,008936	13,54844	0,000000
P41 IgM	0,090314	0,018318	0,89705	0,181947	4,93028	0,000003
P39 IgM	0,058075	0,013693	1,03417	0,243833	4,24133	0,000050
OspC Ba (Borrelia afzelii) IgM	0,122949	0,024013	1,46414	0,285962	5,12003	0,000002
OspC Bb (Borrelia burgdorferi)	0,063320	0,019598	0,98986	0,306368	3,23094	0,001680
OspC Bg (Borrelia garinii) IgM	0,081810	0,023170	1,02911	0,291461	3,53089	0,000633
IgM	0,071089	0,020891	0,78761	0,231460	3,40280	0,000967
VLsE (Borrelia afzelii) IgG	0,108828	0,024666	1,34266	0,304315	4,41207	0,000026
VLsE (Borrelia burgdorferi) IgG	0,100051	0,021695	1,09512	0,237461	4,61181	0,000012
VLsE (Borrelia garinii) IgG	0,093942	0,022904	1,18172	0,288120	4,10149	0,000085
P83 IgG	0,064251	0,016459	1,14414	0,293086	3,90378	0,000174
P41	0,151632	0,018201	1,41743	0,170144	8,33078	0,000000
P39 IgG	0,085979	0,016427	1,53108	0,292533	5,23387	0,000001
OspC (B. afzelii) IgG	0,078117	0,020485	0,78098	0,204803	3,81330	0,000240
P21 IgG	0,099636	0,014243	1,69079	0,241690	6,99570	0,000000
P19 IgG	0,065098	0,013650	2,38568	0,500232	4,76914	0,000006

Figure 6: The result of obtaining significant factors for the prediction of CRRFLB in multivariate regression analysis in Statistica 10.0 without IgG factor

Based on the results of the multivariate regression analysis of predicting the development of severe Lyme borreliosis, which are shown in Figure 6 and Table 1, we build a mathematical model to determine the coefficient of risk factors of developing LB (CRFDLB):

$$\text{CRFDLB} = \mathbf{X1*0,21131+X2*0,85457+X3*1,01908+X4*0,67524+}$$
$$\mathbf{+X5*1,01044+X6*0,08889+X7*0,12108+X10*0,89705+X11*1,03417+}$$
$$\mathbf{+X12*1,46414+X13*0,98986+X14*1,02911+X15*0,78761+X16*1,34266+}$$
$$\mathbf{+X17*1,09512+X18*1,18172+X21*1,14414+X22*1,41743+X23*1,53108+}$$
$$\mathbf{+X24*0,78098+X26*1,69079+X28*2,38568-1,07264}$$

To evaluate the quality of the regression model, it was necessary to analyse the residual deviations, in particular, to obtain their histogram (Figure 7). As can be seen from the histogram, the residual deviations are distributed symmetrically, approaching the curve of the normal distribution of the residuals, so the statistical hypothesis about their distribution in accordance with the normal distribution law is not rejected.

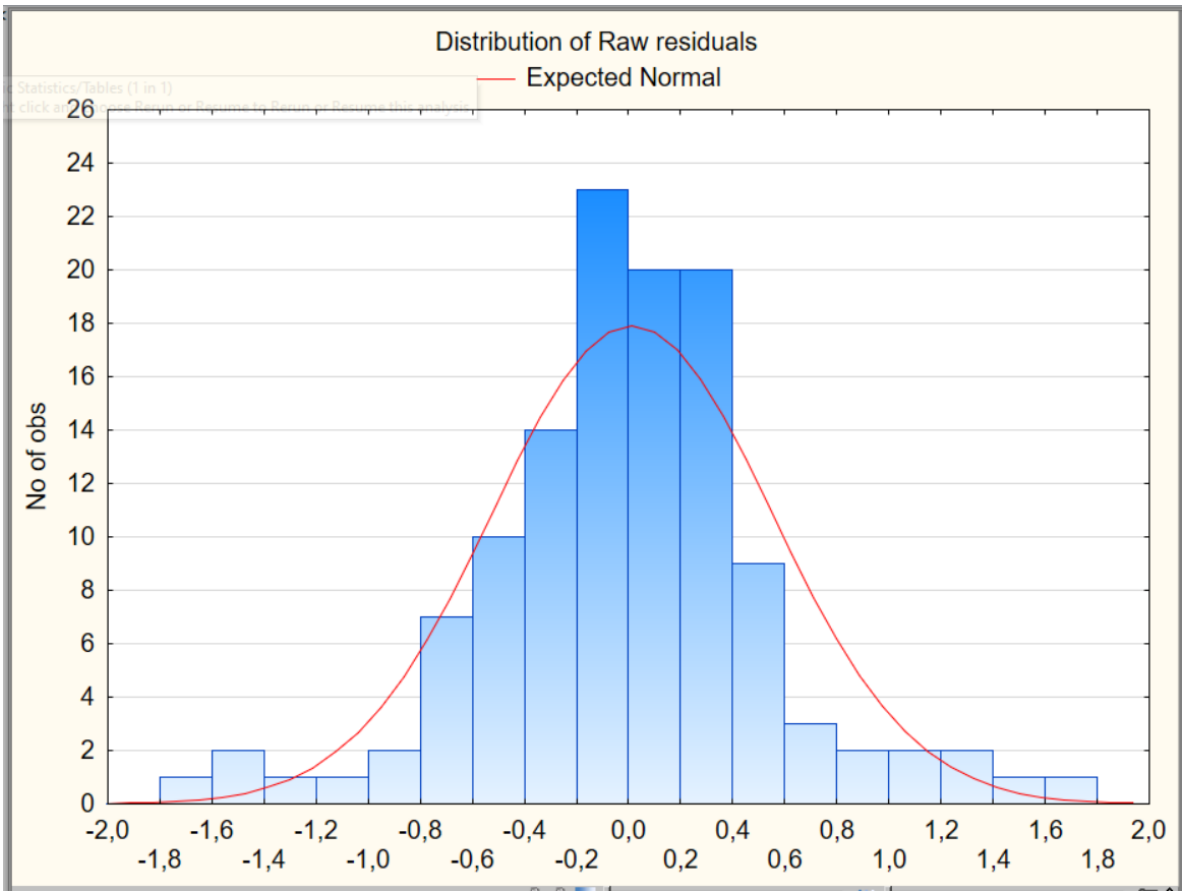


Figure 7: Histogram of residual deviations of the multivariate regression model for predicting the CRFDLB

In order to further confirm the residual deviations from the normal distribution law, a normal-probability graph was constructed (Figure 8). Analysing its data, we note the absence of systematic deviations from the normal probability line. This allows us to conclude that the residual deviations are distributed according to the normal distribution law.

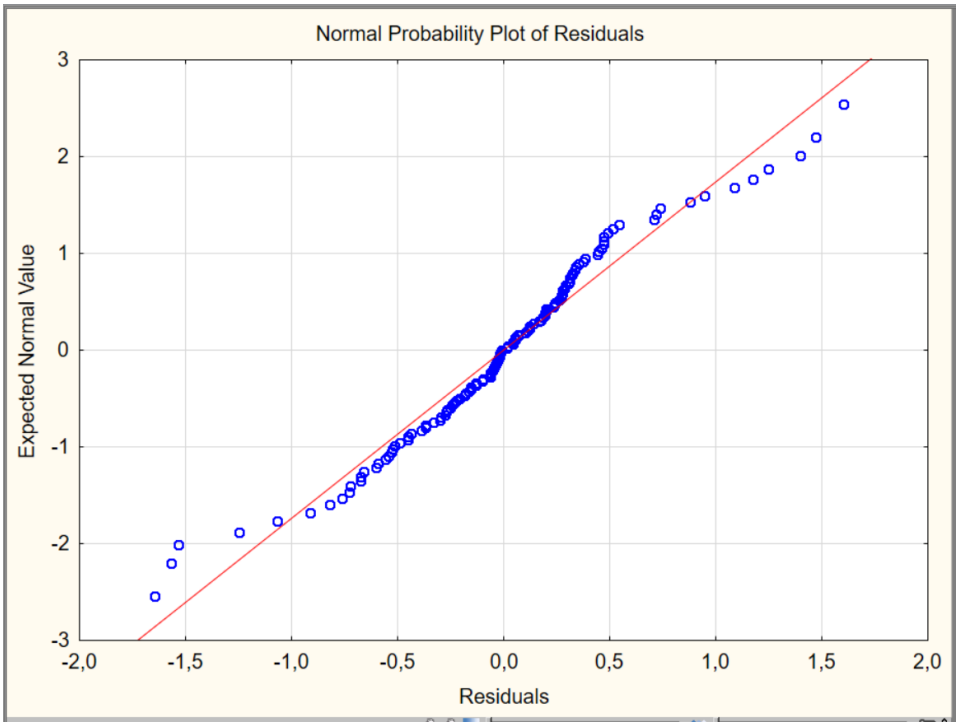


Figure 8: Normal probability plot of residual deviations of the multivariate regression model for predicting the CRFDLB

To check the dependence of the residual deviations on the predicted values, we construct a scatter plot (Figure 9).

Based on the results obtained, we note that the residuals relative to the predicted values are scattered randomly, which indicates that there is no dependence on the predicted values of the CRRFLB. The histogram and the normal probability plot confirm that the residual deviations follow the normal distribution law. Thus, the obtained model for predicting the risk of thrombosis is qualitative and adequate.

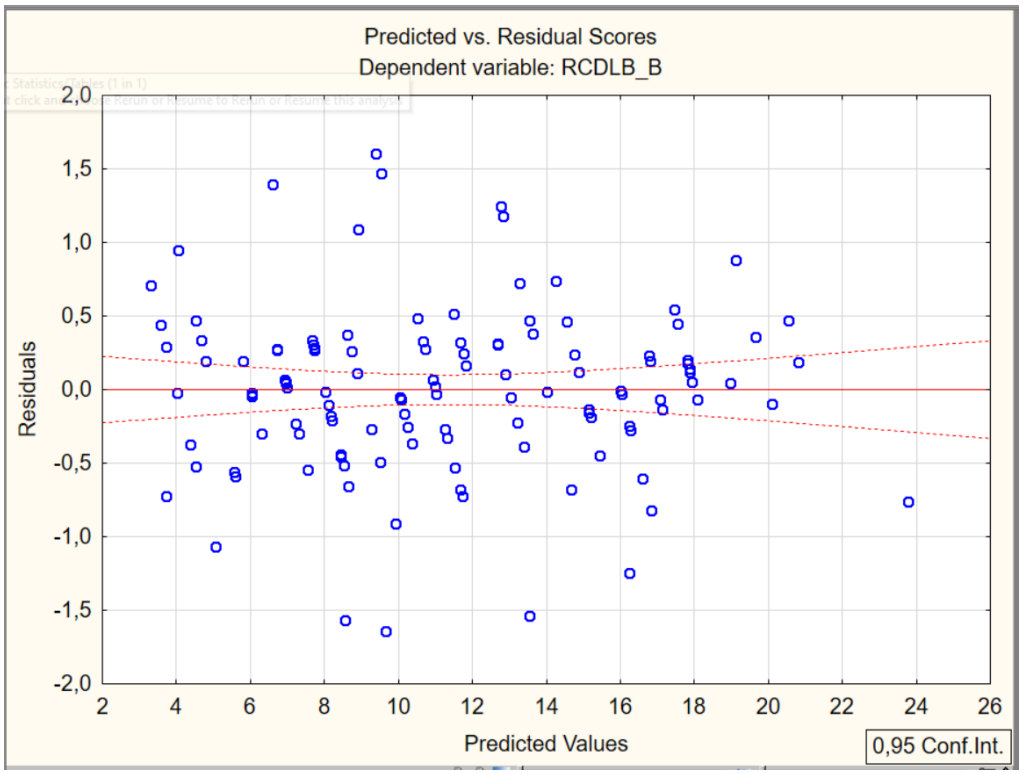


Figure 9: Scatterplot of residual deviations of the multivariate regression model for predicting CRFDLB

The next step was to evaluate the overall goodness of fit of the model, for which we performed an ANOVA analysis (Figure 10). Analysing the data obtained, we can conclude that the model for predicting the CRRFLB is highly satisfactory in general using ANOVA analysis, since the significance level is $p<0.001$, and the model itself will work better than a simple forecast using average values.

Effect	Analysis of Variance; DV: RCDLB_B (1 in 1)				
	Sums of Squares	df	Mean Squares	F	p-value
Regress.	2604,291	22	118,3769	310,2311	0,00
Residual	37,394	98	0,3816		
Total	2641,686				

Figure 10: Analysis of the coefficient of determination of the multivariate regression model for predicting the CRRFLB

To further evaluate the quality of the mathematical model of the CRRFLB, we analysed the coefficient of determination of the Neijelkerk (R^2), which shows what part of the factors is taken into account in the forecast. It is considered a universal measure of the relationship between one random variable and others. The coefficient of determination varies from 0 to 1. The more its value approaches "1", the better the multivariate regression model is. In the proposed mathematical model of the CRRFLB, the coefficient of determination is R2=0.9858 (in Statistica 10.0 R?= ,98584446 (Fig. 10)). Thus, in our case, 98.58% of the factors are taken into account in the CRRFLB prediction model. The coefficient of determination indicates the extent to which the observations confirm the mathematical model.

4. Discussion

Nonspecific symptoms, such as arthralgias, myalgias, fatigue, headaches, irritability, stiff neck muscles, and paresthesias, often last for a long time. Systemic symptoms, including myalgia and arthralgia, can accompany EM, especially in Bb and Bg infections [2].

LA is manifested by fever, persistent monoarthritis, and synovitis. Children with joint involvement caused by Lyme disease have more frequent knee involvement, pain, myalgia, and lower peripheral leukocyte counts; they are less likely to have fever compared to children with septic arthritis [7]. Serologic tests (ELISA and immunoblot) are the gold standard for verifying the diagnosis even in the absence of an epidemiologic history. The main immunodominant proteins are OspC, VlSE, OspA, BmpA, p66, P83/100. Thus, innate and adaptive forms of immunity are mobilized to fight the infection. Most often, specific IgM antibodies in the immunoblot are detected against antigens P18, OspC, P39, P41 from B. afzelii strains; P39, p 41, P 66, P83 from B. garinii strains; OspC, OspA from B. burgdorferi sensu stricto strains. Small amounts of Ig M to flagellin P41 and the membrane protein OspC are detected in the first days of the disease. Their titers increase over 4-6 weeks, and in untreated patients longer.

During the period of generalization of the infectious process, IgG against a number of proteins, such as P39, P58, appear. At the late stage of the disease, a wide range of antibodies to borrelia proteins P83/100, P58, P43, P39, P30, P21 Osp17, P14 appear [8, 9,10].

Genetic testing for HLA antigen B27 is essential in the differential diagnosis of arthritis [11]. The presence of HLA-B27 is associated with certain autoimmune and immune-mediated diseases, including ankylosing spondylitis, which causes inflammation of the spinal bones, and reactive arthritis. In three patients with arthritis, we discovered a gene HLA-B27.

Additional B. burgdorferi epitopes may be involved in the development of antibiotic-resistant Lyme arthritis. OspA163-175 remains the only known recognized epitope of BB and related diseases [12].

Patients with antibiotic-resistant arthritis usually have certain HLA-DRB1 molecules that bind the B. burgdorferi epitope to the outer surface (OspA163-175), and the cellular and humoral immune response to OspA is greater than in patients with antibiotic-responsive arthritis [13].

A number of scientific works have used the method of mathematical forecasting to assess the course of diseases [14,15,16]. To develop the model, we conducted a retrospective analysis of clinical and laboratory data from a cohort of pediatric patients diagnosed with Lyme borreliosis. We then developed a scoring system based on these factors and evaluated its performance using ROC analysis [17]. The addition of the group of patients with erythema-free form in the cluster analysis resulted in a division into four clusters.

Initial factors are after constructing the correlation matrix without taking into account the number of bites (X3), Lipid Bb (Borrelia *burgdorferri*) (X18), P39(IgG) (X21), and P20(IgG) (X25), there were no multicollinear factors, as there were no pairwise correlation coefficients greater than 0.7. All of the above 24 factors were used to build a multivariate regression model.

The works [18, 19, 20, 21] consider approaches to the development of medical sensor monitoring tools, the main component of which is data transmission through electronic communication channels and networks [22-24]. Thanks to such approaches, appropriate sensors were used to register the indicators, which are listed in Table 1.

5. Conclusions

1. PCA and Claster metods should be used in diagnostic of Lyme disease

2. The model was clustered based on the coefficients. The sum of points from 1 to 10 indicates a mild form of the disease, from 10 to 20 - a severe form of the disease. Therefore, the result is that the Localised form is mild and severe and the Disseminated form is divided into mild and severe.

References

[1] Centers for Disease Control and Prevention. Lyme disease (Borrelia burgdorferi): 2017 case definition. URL: https://wwwn.cdc.gov/nndss/conditions/ lyme-disease/case-definition/2017.

[2] M. Chowaniec, A. Starba, P. Wiland, Erythema nodosum - review of the literature. Reumatologia, 54(2), 2016, pp. 79-82. doi: 10.5114/reum.2016.60217.

[3] G. Stanek, G. Wormser, J. Gray, F. Strle, Lyme borreliosis, Lancet, 379 (2012) 461–473. doi: 10.1016/S0140-66736(11)60103-7.

[4] S. Nykytyuk, S. Klymnyuk, O. Marchuk, V. Panichev, I. Klishch, Experience of PCR research in Lyme borreliosis in children of Ternopil region, in: Family Medicine & Primary Care Review, 24 (2022) 334–335. doi:10.5114/fmpcr.2022.120857.

[5] S. Nykytyuk, S. Klymniuk, D. Pyvovarchuk, A. Sverstyuk, Multifactorial model for predicting the severe course and damage to organs and systems in Lyme borreliosis in children, Modern pediatrics, 2 (2023) 6-17.

[6] S. Nykytyuk, S. Levenets, M. Horishnyi, I. Horishnyi, Awareness of Lyme disease among vocational school students and children (Ternopil Region, Western Ukraine), Georgian Medical News, 24(2022) 67–71. PMID:36780626.

[7] Borrelia burgdorferi peptidoglycan is a persistent antigen in patients with Lyme arthritis, in: Proceedings of the National Academy of Sciences of the United States of America, volume 7(2), 2019, pp. e13498-e13507. doi:10.1073/pnas.1904170116.

[8] G. Norman, J. Antig, G. Bigaignon, B. Hogrefe, Serodiagnosis of Lyme borreliosis by Borrelia burgdorferi sensu stricto, B. garinii, and B. afzelii western blots (immunoblots), Journal of Clinical Microbiology, 34 (1996) e1732–e1738.

[9] M. Lager, B. Ram, P. Wilhelmsson, D. Nyman, Serological diagnostics of Lyme borreliosis: comparison of assays in twelve clinical laboratories in Northern Europe, ·European Journal of Clinical Microbiology & Infectious Diseases, volume 38(2019) e1933-e1945.

[10] G. Norman, J. Antig, G. Bigaignon, W. Hogrefe, Serodiagnosis of Lyme borreliosis by Borrelia burgdorferi sensu stricto, B. garinii, and B. afzelii western blots (immunoblots), Journal of Clinical Microbiology, volume 34 (1996) 1732–1738.

[11] S. Nimmrich, I. Becker, G. Horneff, Intraarticular corticosteroids in refractory childhood Lyme arthritis. Rheumatology International, 34 (2014) e987-e994.

[12] Biologic Markers of Antibiotic-Refractory Lyme Arthritis in Human: A Systematic Review, Infectious Diseases & Therapy, 8(2019) e5-e22. doi:10.1007/s40121-018-0223-0.

[13] S. Nykytyuk, O. Boyarchuk, S. Klymnyuk, S. Levenets, The Jarisch-Herxheimer reaction associated with doxycycline in a patient with Lyme arthritis, *Reumatologia*, 58 (2020) 335–338. doi:10.5114/reum.2020.99143.

[14] Musiienko V., Marushchak M., Sverstuik A., Filipyuk A., Krynytska I. Prediction Factors for The Risk Of Hypothyroidism Development In Type 2 Diabetic Patients. Pharmacology On Line. 3 (2021) 585-594.

[15] Musiienko V, Sverstiuk A, Lepyavko A, Mazur L, Danchak S, Lisnianska N. Prediction factors for the risk of diffuse non-toxic goiter development in type 2 diabetic patients. Pol Merkur Lekarski. 2022;4. 19;50(296) 94-98. PMID: 35436270.

[16] Chukur O, Pasyechko N, Bob A, Sverstiuk A. Prediction of climacteric syndrome development in perimenopausal women with hypothyroidism. Prz Menopauzalny. 2022 :12;21(4):236-241. doi:10.5114/pm.2022.123522.

[17] Gruzieva T.S., Lekhan V.M., Ognev V.A. and others, Biostatistics: a textbook, Vinnytsia: Nova Knyha, 2020, 384 p.

[18] V. Martsenyuk, A. Sverstiuk, A. Klos-Witkowska, N. Kozodii, O. Bagriy-Zayats, I. Zubenko, Numerical Analysis of Results Simulation of Cyber-Physical Biosensor Systems, in: 1st International Workshop Information-Communication Techologies&Embedded Systems, 14-15 November, Mykolaiv, volume 1, 2019, pp. 149–164.

[19] V. Martsenyuk, A. Sverstiuk, O. Bahrii-Zaiats, A. Klos-Witkowska, Qualitative and Quantitative Comparative Analysis of Results of Numerical Simulation of Cyber-Physical Biosensor Systems, in: CEUR Workshop Proceedings, 2022, 3309, pp. 134–149.

[20] V. Martsenyuk, A. Klos-Witkowska, A. Sverstiuk, O. Bagrii-Zayats, M. Bernas, On modelling predator-prey cellular automaton with help of lattice differential equations with time dilay: in Advances in biotechnology. 18th International Multidisciplinary Scientific GeoConference SGEM 2018. NANO, BIO, GREEN and SPACE TECHNOLOGIES FOR A SUSTAINABLE FUTURE, 2th-8th of July. – 2018. Volume 18. pp.407-414.

[21] S.V. Romaniv, Y.B. Palaniza, D.V. Vakulenko, I.Y. Galaychuk, The method of using fractal analysis for metastatic nodules diagnostics on computer tomographic images of lungs. Horizons in Cancer Research, March 30, 2023, v. 85, pp. 231–247.

[22] L. Khvostivska, M. Khvostivskyy, V. Dunetc, I. Dediv. Mathematical and Algorithmic Support of Detection Useful Radiosignals in Telecommunication Networks. Proceedings of the 2nd International Workshop on Information Technologies: Theoretical and Applied Problems (ITTAP 2022). Ternopil, Ukraine, November 22-24, 2022. pp.314-318. ISSN 1613-0073.

[23] L. Khvostivska, M. Khvostivskyi, I. Dediv, V. Yatskiv, Y. Palaniza. Method, Algorithm and Computer Tool for Synphase Detection of Radio Signals in Telecommunication Networks with Noises. Proceedings of the 1st International Workshop on Computer Information Technologies in Industry 4.0 (CITI 2023). CEUR Workshop Proceedings. Ternopil, Ukraine, June 14-16, 2023. pp.173-180. ISSN 1613-0073.

[24] M. Khvostivskyy, H. Osukhivska, L. Khvostivska, T. Lobur, D. Velychko, S. Lupenko, T. Hovorushchenko, Mathematical modelling of daily computer network traffic. ITTAP 2021. CEUR Workshop Proceedings. Ternopil, Ukraine, November 16-18, 2021. Vol. 3039. pp.107-111.