Using of Data Mining methods to evaluate the myocardial damage in children with juvenile idiopathic arthritis

Illya Chaikovsky
Glushkov Institute of Cybernetics
Kyiv, Ukraine
illya.chaikovsky@gmail.com

Olena Oshlianska
Shupyk National Medical Academy of Postgraduate Education
Kyiv, Ukraine
eoshljanskaja17@gmail.com

Agar Artsymovych
Shupyk National Medical Academy of Postgraduate Education
Kyiv, Ukraine
wintermdp@gmail.com

Olga Kryvova
International Research and Training Center of IT and S
Kyiv, Ukraine
olga.kryvova@gmail.com

Olexander Kovalenko
International Research and Training Center of IT and S
Kyiv, Ukraine
askov49@gmail.com

Leonid Stadniuk
Shupyk National Medical Academy of Postgraduate Education
Kyiv, Ukraine
leonid.stadniuk@gmail.com

***Abstract*—This study is aimed to assess validity the system of comprehensive ECG scoring, using Data Mining. The regression and classification models of complex index of functional state were developed. Cluster analysis and decision trees algorithms (CART) were used for identification subgroup of patients. Predictive indices for diagnosis of the functional state of the cardiovascular system of children with juvenile idiopathic arthritis were selected.**

Keywords—universal ECG scoring, myocardium damage, decision trees, data mining, juvenile idiopathic arthritis.

I. INTRODUCTION

Screening of cardiac function is important to detect pathological conditions in patients with juvenile idiopathic arthritis (JIA), because it is the cardiovascular lesions that take the first place in the structure of mortality of such patients. Early secondary cardiovascular disorders in patients with juvenile idiopathic arthritis are difficult for the clinician to interpret due to the large number of possible factors for their development. In pediatric practice, early diagnosis of these conditions is also complicated by the fact that children are not always able to clearly formulate complaints about their condition. The need for improving instrumental examination of patients led to the improvement of electrocardiogram (ECG) assessment methods.

The fundamental difference of ECG of 4th generation is that the information is processed using specialized software. The analysis uses a "signal-averaged ECG", obtained by processing several electrocardiographic complexes with the exception of atypical ones. The new concept of ECG assessment facilitates decision-making in preventive and monitoring reviews. It includes registering an ECG from the limbs with the help of a portable ECG of the “Cardio-Plus P” software-hardware complex and conducting a multilevel analysis that is performed automatically and instantly with a visual result form. It is necessary to diagnose the problem before clinical signs appear to prevent irreversible disabling changes in cardiovascular system (CVS).

The purpose of this study is to substantiate the applicability of the system of indicators of multilevel ECG analysis for assessing the functional state (FS) of children with JIA.

Data mining techniques have been widely used in healthcare [1], for example, for classification and prediction of heart disease [2-4]. The most popular methods include: Decision Tree, Clustering [2, 3], Naive Bayes, Logistic Regression, Support Vector Machine, Neural Network [4-6]. Recently, a hybrid approach has been used, which uses a combination of the techniques [7-8].

II. METHODS

Modeling was carried out according to the results of clinical and instrumental examination of children (34 people with JIA). There were 12 boys, 22 girls. Average age was 10.1±1.2 years (2-17 years). The duration of the disease ranged from 3 month up to 4 years (1.45±0.51 years). There were no instrumental and laboratory signs of CVS injury (carditis, secondary cardiomyopathy) among observed group. Disease activity was measured according to the Juvenile Arthritis Disease Activity Score (JADAS) [9]. Registration of a 6 channel ECG was carried out for 5 –20 minutes using the “Cardio Plus P” software and hardware device. The ECG indicators and the system of universal ECG scoring were calculated [10].

In addition to the conventional indicators (duration of waves, segments, intervals, complexes and their amplitudes), software-hardware complex “Cardio-Plus P” measured over 180 of additional ECG parameters: standard deviation of normal-to-normal intervals, the measure of overall heart rate variability, the level of overall adaptive potential of autonomic regulation (SDNN); beat-to-beat variability, rapid fluctuations in heart rate, which characterizes parasympathetic regulations (RMSSD); waves, measure of the activity center baroreflex medulla (primarily - sympathetic activity) (LF); absolute spectral power in domain of high frequency waves, represents the measure of nuclei activity of the vagus nerve (parasympathetic activity) (HF); sympathetic and parasympathetic regulation ratio (LF/HF); autonomic balance index to represent the ratio of sympathetic and parasympathetic parts of the autonomic nervous system (IBP); T-wave symmetry etc.

After determining the average values of the standard ECG intervals, “Cardio-Plus P” additionally automatically calculated integral indicators: K1 ((PQ+QTc)/RR), K2 ((QTc+0.5QR)/(PQ+QTc)); Macruz index etc.

After evaluating the changes detected, patients were divided into groups, depending on the degree of activity of the inflammatory process. Cumulative indexes ECG were

978-1-7281-9713-5/20/$31.00 ©2020 IEEE

considered. These cumulative indexes were constructed by aggregation set of ECG indicators:

- Cardiovascular regulation index (CVSI). This index is formed from 27 indicators of heart rate variability (HVR). These are indicators of temporal, spectral geometric analysis, as well as measures of nonlinear analysis of the dynamics of the heart rhythm.
- Current control of myocardium state index (CMSI). 28 ECG indicators are aggregated into this index.
- Myocardial reserves index (MRI) - 124 amplitude-time parameters (amplitude of a wave, segment or interval) of ECG.
- In-depth ECG analysis index (VCGI) is a set of 7 vector cardiographic parameters.
- Myocardial state index (MSI) as average of indices CMSI, MRI and VCGI.
- Functional state index of CVS (FSI) was calculated by aggregating CVSI and MSI indexes.

In order to develop an interpretable classifier for assessing myocardial damage, we proposed a novel method that combines clustering (K-means, EM algorithm) and decision tree method (Classification And Regression Trees, CART algorithm.

The method includes the following stages:

1) Clustering on these indexes for the separation patients into typological groups of the FS.
2) Development of regression models for these indexes using the decision tree method.
3) Creating the classification model for FSI with the best combination of ECG indicators
4) Analysis of the predictive capabilities of the set FS predictors to determine the degree of disease activity.

III. RESULTS

Analysis of clinical manifestations of the disease showed that in the group of children examined, complaints that were caused by disorders of the CVS were almost not observed, only general clinical complaints were noted.

Analysis of laboratory examination data of this group of patients with JIA revealed the following changes. Anemia in the general blood test was observed in 4 children, which could lead to circulatory hypoxia and contribute to the development of secondary metabolic disorders in the myocardium. The erythrocyte sedimentation rate (ESR) at the time of the observation was mostly within the normal range (up to 15 mm/h), but in some children (up to 10%) it reached 40 mm/h. Significant differences from normal values in the biochemical parameters of the examined patients were also not observed (creatinecinase, lactate dehydrogenase, transaminase activity did not differ from normal values), the C-reactive protein (CRP) level averaged 12.92 ± 4.98 ng/l.

The JADAS-27 (includes 4 measures: physician global assessment of disease activity, parent/patient global assessment of well-being, active joint count, and ESR or CRP) was 9.5 ± 2.04. Immunological parameters in the examined patients with JIA also corresponded to the activity of the inflammatory process, serum immunoglobulins level was within the reference age values, antinuclear antibodies to native and denatured DNA and antiphospholipid antibodies were not found in the diagnostic-significant titers.

According to the standard 12-channel ECG, no issues was observed in only 3 (9%) patients. However, the changes detected were minimal, generally of a functional nature and were not diagnostically significant for the establishment of certain cardiac pathology. Thus, sinus arrhythmia was observed in 42% of patients, deviation of the electrical axis of the heart in 11.5% of children, incomplete right bundle branch block in 15% of patients (which mostly corresponded to age characteristics), shortening of pQ was found in 2 children (6%), in the same amount there was a violation of intra-ventricular conduction. Moderate myocardial exchange changes were determined during the standard ECG assessment in 61.5% of the patients examined. Such ECG data did not allow the doctor to ascertain or predict any abnormalities on the part of the CVS and suggest their correction or prevention, even combining with laboratory data and activity score.

According to cardiac ultrasound in the examined children, almost no changes were detected, except for small congenital anomalies, such as: additional chords of the left ventricle, mitral valve prolapse 1st degree etc. No child had an organic pathology of the myocardium or coronary vessels (no fibrosis, endothelial damage, hypertrophy of left ventricular, amyloid accumulation). According to ECG with the help of “Cardio Plus P” evaluation, a large number of additional parameters have been identified in children with JIA, which allows to characterize in more detail the functional status of the myocardium and to evaluate the impact of various regulatory factors (table 1).

The method allows to detect cardiac arrhythmias one third more often than the standard 12-channel ECG. Comprehensive assessment of regulation showed its significant decrease in patients with JIA in 65% of cases: 68.47±2.6, mostly due to vegetative imbalance and lack of reserves. It is crucial to evaluate the prognosis of a further course of the disease by detecting early the metabolic changes and their associated functional myocardial disorders. The evaluation of ECG recorded by “Cardio Plus P” allowed us to conduct a thorough and comprehensive study of metabolic processes at different loci of the cardiac muscle.

The data obtained (table 2) showed the presence of significant changes in the myocardium in the majority of patients with JIA children. Thus, the operative control of the myocardial status was reduced in 59% of patients. Because the integral indicators of the ST-T form have been altered in all leads, it can be concluded that the metabolic disorders of the cardiac muscle were diffuse. However, the T-wave morphology (116.47±18.22) was generally normal in most patients (82%), moderate changes were observed in 6%, profound in 12%.

TABLE I. INDICATORS OF HEART RHYTHM VARIABILITY ACCORDING TO ECG EVALUATION USING "CARDIO PLUS P" IN PATIENTS WITH JIA

Name, units	Values		
	Children with JIA, M±m	*Normal ranges*	*Deviations from the norm, n (%)*
Immediate control of the regulation	74.88±3.07	75-100	14 (41%)
Heart rate per min.	83.88±3.75	60-100	4 (12%)
SDNN, ms	46.18±5.34	≥39	14 (41%)
RMSSD, ms	45±6.5	≥30	12 (35%)
Stress – index, s^{-2}	268.58±53.79	≤120	24 (71%)
Triangular index	12.17±1.78	≥9	12 (35%)
PNN50, %	21.18±4.9	≥9	12 (35%)
Condition of regulation reserves	63±3.03	75-100	28 (82%)
Autonomic balance 1 (LF/HF)	1,2±0,26	1-3	26 (76%)
Autonomic balance 2 (BPI), s-1	338,82±80	100-350	18 (53%)
Sympathetic and parasympathetic regulation ratio;	1.2±0.26	1-3	26 (76%)
Activity of subcortical centers, %	2.65±0.59	3	24 (71%)
Integral indicators	69.23±3.5	75-100	18 (53%)

T-wave symmetry by the correlation of max derivatives was decreased in 36-75% patients depending on lead, while 100% of patients had their T-wave symmetry by correlation of squares of triangles decreased in all leads except I. T-wave displays the final phase of ventricular myocardial repolarization, and its decrease may display cardiomyopathy.

ECG-waves amplitude indexes were decreased in all leads, more frequently in I and AvL leads (which are corresponding to the front wall and the left anterolateral wall of the heart respectively), but the most severe damage was caused to the right side wall of the heart, which is presented by AvR lead. The absence of an ST segment shift 0.08 s after point J in all recorded leads reflected that the detected changes in the myocardium of patients with JIA were not due to hypoxia.

The myocardial index of stationarity in most patients was normal (99%), less than a third had lower values. HF-QRS changes were detected in 100% of patients with a mean of 0.68±0.06, with a significant deviation from the norm in most cases. The decrease in the index of ECG phase's ratio reflected the poor performance of the diastole of the heart. Decreases of R- and T-wave amplitudes were marked in the first lead (300.94±55.99) for R-wave, 35.82±47.47 for T-wave). The golden section was deviated from the norm in 32 cases (94%).

Strong negative correlations of some of the following parameters with JADAS were found: with T-wave symmetry by correlation of squares of triangles in lead I (r= - 0.89); with T-wave symmetry by the correlation of max derivatives, lead II (r= - 0.58); with T-wave symmetry by the correlation of max derivatives, lead III (r= - 0.59); with T-wave symmetry by the correlation of max derivatives, lead AvL (r= - 0.59); with T-wave symmetry by the correlation of max derivatives, lead AvF (r= - 0.55); T - wave morphology (r = - 0.64).

TABLE II. INDICATORS OF MYOCARDIUM STATE ACCORDING TO ECG EVALUATION USING "CARDIO PLUS P" IN PATIENTS WITH JIA

Name, units	Values		
	Children with JIA, M±m	*Normal ranges*	*Deviations from the norm, n (%)*
Immediate control of condition of myocardium	51.65±4.8	75-100	32 (94%)
Integral indicator of form ST-T, lead I	60.41±6.81	75-100	20 (59%)
T-wave/R-wave ratio, lead I	0.67±0.11	0.143-0.333	27 (80%)
Condition of myocardium reserves	62.65±2.09	75-100	32 (94%)
Amplitude-areas index (lead I)	50.94±3.16	75-100	32 (94%)
Amplitude-areas index (lead II)	58,12±5,34	75-100	26 (76%)
Amplitude-areas index (lead III)	61,47±5,07	75-10	24 (71%)
Amplitude-areas index (lead AvL)	41,59±4,18	75-100	32 (94%)
Amplitude-areas index (lead AvR)	39±9,83	75-100	26 (76%)
Amplitude-areas index (lead AvF)	66,59±5,21	75-100	22 (65%)
Macruz index P/(PQ-P)	1.39±0.76	1.1-1.6	32 (94%)
Advanced ECG analysis	63.71±2.78	75-100	28 (82%)
Condition of myocardium reserves	62.65±2.09	75-100	32 (94%)
Amplitude-areas index (lead I)	50.94±3.16	75-100	32 (94%)
Amplitude-areas index (lead II)	58,12±5,34	75-100	26 (76%)
Amplitude-areas index (lead III)	61,47±5,07	75-10	24 (71%)

Summarizing the above, it should be noted that the complex indicator of the condition of the myocardium in the examined patients at JIA was 57.35±3.09, only 2 of observed patients (6%) had normal values. in 10 cases (30%) significant violation occurred. the rest of patients with JIA - revealed minor violations. The results of clustering and Classification And Regression algorithm (CART) were as follows.

A. Clustering solution

Using EM and K-means clustering methods with 10-fold cross-validation, we automatically obtained the optimal distribution of children into two typological classes according to 6 variables - FS indicators. Fig. 1 shows a plot of the average's variables in the each cluster.

Cluster 1 (20 children) - low values of all variables; composite index of functional state: $FSI_1 = 61±3.21$.

This subgroup consisted mostly of children, who had prolonged inflammatory process, unfavorable course of disease (severe eye lesions. sacroileitis etc.) or suffered from systemic variant of JIA.

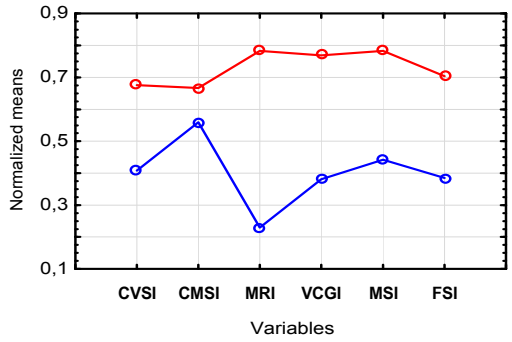

Fig. 1. Graf of means for variables

Cluster 2 (14 people) - higher grades for all complex indicators: FSI2 = 77.29±0.64. These children had mostly had mild disease course and low laboratory markers of inflammation.

B. Selection the best predictors FS index

We tested ECG indicator sets to determine the optimal models for each of the six indices (CVSI, CMSI, MRI, VCGI, MSI, and FSI) by using Classification And Regression Trees (CART) algorithm. The structure of the resulting binary regression trees was analyzed. In order to enhance generalization of decision tree we used pruning with combination of 10-folds cross-validation error rate estimation. The best combination of ECG indicators for predicting the FSI value was selected by training process. Table 3 represents these predictors with high-ranking importance scores for regression model of FSI (R=0.884; training error 18.8%).

TABLE III. THE BEST COMBINATION OF ECG INDICATORS

Predictors FSI	**Predictor importance**	
	Variable rank	***Importance***
Heart rhythm disorders	100	1.00
T - wave amplitude (lead II)	92	0.92
Integral indicator of the form of STT (lead II)	77	0.77
QRS - alpha angle	75	0.75
T - wave symmetry ratio	31	0.31

C. Classificational model for disease activity

These 4 predictors of FSI were tested as attributes for the classification model for disease activity. CART algorithm [2] is a classification algorithm for building a binary tree based on Gini index as splitting criterion. The Gini index (GI) is an impurity-based criterion that measures the divergence between the probability distributions of the target attributes values [8].

Since our data set was relatively small, we applied cross-validation (10 –fold) to improve the overall model performance and to optimize model complexity without over-fitting. As a result, an interpretable classification tree was obtained for 4 classes. Disease activity score had 4 grades: 0, 1, 2, 3 according to patient's JADAS score.

The classification tree for disease activity is shown in Fig. 2.

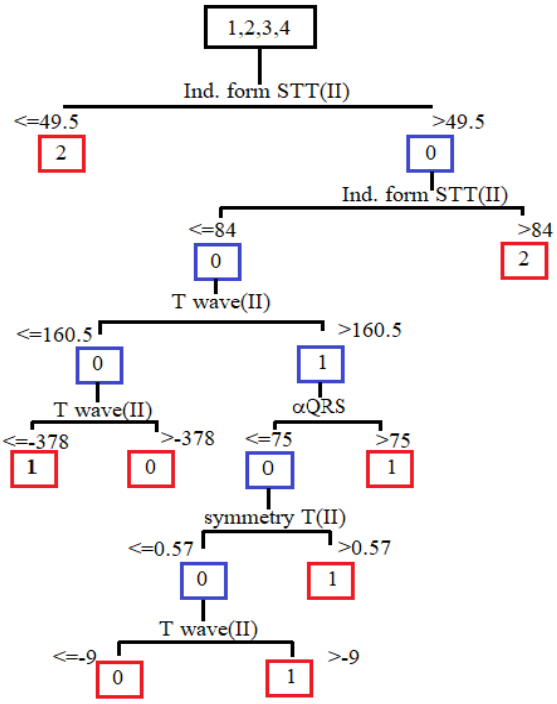

Fig. 2. Tree graf for disease activity

Overall accuracy of classification is 88%. It should be noted that the first two grades of disease activity (the initial stages of juvenile idiopathic arthritis) are recognized by 4 ECG indicators with 100% accuracy.

We calculated such performance measures for the disease activity as sensitivity, specificity and accuracy for 3 classes [2]:

$$Sensitivity=TP/(TP+FN) \quad (1)$$

$$Specificity=TN/(FP+TN) \quad (2)$$

$$Predictive\ Accuracy=(TP+TN)/total \quad (3)$$

$$total=(TP+TN+FP+FN),$$

where - TP, TN, FP, and FN are the numbers of true positives, true negatives, false positives, and false negatives, respectively.

For multiclass case, these measures can be obtained from the confusion matrix by comparing numbers of instances for each class in the matrix against instances of all the other classes. In Table IV presents performance measure for 3 classes of JVA activity.

TABLE IV. CLASSIFICATION RESULTS FOR 3 CLASSES OF DIASEASE ACTIVITY USING THE 4 PREDICTORS OF FS

Measures (%)	**Disease activity**		
	0	***1***	***2***
Sensitivity	84,6	100	66,7
Specificity	100	87,5	100
Predictive Accuracy	92	92	56

IV. DISCUSSION

Inflammation is the universal pathophysiologic regulation mechanism of reaction of living creatures, the precise prediction of which is giving us a possibility to interrupt the process early. During the inflammatory process of any localization, a few messengers of inflammation is being delivered to the internal environment of the body, such as blood, lymph etc. These messengers lead to the deceleration of metabolic processes, including lipid metabolism, electrolyte balance, which directly affect the myocardial function.

Usage of the proposed ECG method can help physician to catch the initial inflammatory process on the outpatient stage.

Obviously, the detection of changes in integral indicator of form ST-T points at severe inflammation process in children with JIA. In case of minor or insignificant changes of ST-T form. T - wave amplitude and QRS - alpha angle must be considered.

Considering the results (table 3), even the slightest heart rhythm disorders such as sinus arrhythmia can be the screening marker that reflects the presence of severe inflammatory process in children with JIA. In such cases it is rational to perform the multilevel ECG analysis for the further evaluation of functional state of CVS.

The advantage of the method is the possibility to evaluate the CVS lesion level and the disease activity level at the same time even before the clinical manifestation of inflammation's sequences.

Thus, as primary evaluation of results of ECG performed by "Cardio Plus P", internal indicator of form of ST-T is making possible the distinguishing a group of children with ECG-changes of T-wave, which reflects the activity of the inflammatory process the most.

Since the persistence of inflammation is the most unfavorable factor in the development of CVS lesions, early detection of such changes will help to provide prevention measures in time.

The methods of mathematical processing are an important part of the evaluation of many physiological and pathological processes in the human body. One of such methods is the multilevel ECG analysis based on creating classification and cluster models [8].

V. CONCLUSION

CVS lesions in patiens with JIA are common. Depending on disease activity, these lesions may graduate from mild to severe. The core problem to be solved is to diagnose or predict CVS injuries in early period, which is impossible by using only existing methods like complete blood count and CRP level, cardiac ultrasound or standart ECG.

The use of hardware-software complex "Cardio-Plus P" makes it possible to found the changes in CVS and to detect latent heart rhythm disturbances in children with JIA three times more frequently by evaluating complex indicators.

The informative nature of the system of ECG indices for the complex evaluation of FS of CVS of children was investigated. Using the methods of data analysis, models (regression and classification trees) of interconnection of complex indicators with parameters of multichannel ECG parameters were constructed, which made it possible to determine the ECG predictors for different types of children's groups according to the FS of their cardiovascular system.

Optimal discriminator values for splits are derived from the data during analysis by testing all possible discriminators and are not settled by the investigators before. The sensitivity, specificity, efficiency, positive, and negative predictive values for this classification procedure were calculated.

REFERENCES

[1] I. Yoo et all, "Data mining in healthcare and biomedicine: a survey of the literature," J. Med. Syst., 2012, vol. 36(4), pp. 2431-2448.

[2] Breiman L., Friedman J., Olshen R., Stone C. Classification and Regression Trees. Belmont: Wadsworth International Group. 1984, 259 p.

[3] M. Shouman, T. Turner, R. Stocker, "Integrating clustering with different data mining techniques in the diagnosis of heart disease," J. Comput. Sci. E ng., 2013, vol. 20, No. 1, pp. 1-10.

[4] I. Kurt, M. Ture, A.Turhan, "Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease," Journal of Expert Systems with Application, 2008, vol. 34, No.1, pp. 366-374.

[5] V. Dominic, D. Gupta , S. Khare, "An effective performance analysis of machine learning techniques for cardiovascular disease", Applied Medical Informatics, 2015, vol. 36, No. 1, pp: 23-32.

[6] M S Amin, Y.K. Chiam, Identification of significant features and data mining techniques in predicting heart disease", Telematics and Informatics vol. 36, 2019, data, pp. 82-93.

[7] M.Tarawneh, O. Embarak., "Hybrid Approach for Heart Disease Prediction Using Data Mining Techniques," International Conference on Emerging Internetworking, Data & Web Technologies, 2019, pp.447-454.

[8] Y. Khan, U. Qamar, M. Asad, and B. Zeb, "Applying feature selection and weight optimization techniques to enhance artificial neural network for heart disease diagnosis," Proceedings of SAI Intelligent Systems Conference, 2019, pp. 340-351.

[9] A. Consolaro. et al. "Development and validation of a composite disease activity score for juvenile idiopathic arthritis," Arthritis & Rheumatism, 2009, vol. 61, pp. 658-666.

[10] I.Chaikovsky, V. Batushkin, W. Visnevsky, "ECG Universal score system: new instrument for electrocardiogram analysis," ICCIIDT'16. London, UK, 2016, pp. 254-268.

[11] O. Oshlianska, I. Chaykovskyy, et all, "A Evaluation of the cardiovascular system status in children with rheumatic diseases by means of the hardware-software complex «Cardio-plus P», "Sovremennaya pedyatryya, " 2017, 8(88), pp. 59-67.

[12] B.Koca, S. Sahin, A. et al. "Cardiac involvement in juvenile idiopathic arthritis," Rheumatol. Int., 2017, vol. 37, pp. 137–142.

[13] J.Elizabeth et all, "Cardiovascular risk in juvenile idiopathic arthritis," Rheumatology, 2013, vol.52 (7), pp. 1163–1171.

[14] C.M. Duffy "Health outcomes in pediatric rheumatic diseases," Current opinion in rheumatology, 2004, vol. 16, pp.102–108.

[15] I. Chaikovsky, O.Krivova, A. Kazmirchyk at all, "Assessment of the post-traumatic damage of myocardium in patients with combat trauma using a data mining analysis of an electrocardiogram," 2019 Signal Processing Symposium (SPSympo), 2019, pp. 34-38.