IMPROVING THE QUALITY OF INTERVAL PATTERN RECOGNITION BY DIVIDING THE PATIENT SAMPLE IN MEDICAL RESEARCH

Gurami Tsitsiashvili[1], Vera Nevzorova[2], Pavel Dunts[2],
Angelina Talko[2], Marina Osipova[1,3]

•

[1]Institute for Applied Mathematics FEBRAS, Russia
[2]Pacific State Medical University, Russia
[3]Far Eastern Federal University, Russia
guram@iam.dvo.ru, nevzorova@inbox.ru, dpv@bk.ru, talkang92@mail.ru, mao1975@list.ru

Abstract

In this paper, we consider a way to improve the quality of interval pattern recognition by dividing the sample into parts. This method is used to study erythrocyte and platelet germs of haematopoiesis for the ability to positively treat haematological patients from COVID-19. Along with this, this method of improving the quality of interval recognition is used to identify predictors of "difficult" tracheal intubation during artificial ventilation during anaesthetic support of surgical interventions. Despite the rather strong difference between these medical tasks, they are united in this research context by the possibility of improving the quality of recognition through a more specialized study of each sub sample of patients after dividing the sample into parts.

Keywords: interval pattern recognition, big data, dividing sample into parts, Coronavirus infection, haemoblastosis, tumor

Introduction

In this paper, during a computational experiment, the possibilities of the interval pattern recognition method are analysed when dividing the initial sample of patients into parts. Analytically, it is not difficult to prove that such a division of the sample never worsens and can improve the quality of recognition using this method. However, only a computational experiment allows us to assess how much it is possible to improve the quality of recognition in the statistical analysis of research data in medicine.

Arrays of data that you often have to work with (see, for example, the first article) when solving various image recognition tasks have the following properties. On the one hand, these arrays are short samples (Rare events). On the other hand, the presence of a large number of features characterizing individual patients in the sample makes arrays of observations with Big data very inconvenient for using methods of multidimensional statistics. Moreover, the data arrays in question contain such a feature as a negative or positive outcome of the disease. This circumstance required the use of the interval pattern recognition method, in which it is possible to correctly identify all patients with an unfavourable outcome.

In this paper, the interval pattern recognition algorithm is briefly described and its properties are given. Among them, the most significant is the use of dividing the sample into parts in order to increase the recognition coefficient when solving specific medical problems. The study of platelet and erythrocyte germs of hematopoiesis in haematological patients who had Coronavirus Disease 2019 (COVID-19) was chosen as the first of them. The second task is related to the identification of predictors of "difficult" tracheal intubation during artificial ventilation during anesthetic support of surgical interventions.

Despite the rather strong difference between these tasks, they are united in this research context by the possibility of improving the quality of recognition through a more specialized study of each sub sample of patients after dividing the sample into parts.

The aim of the study is to improve the clinical decision-making process based on predictive statistics in some medical scientific research. In accordance with the goal, the tasks are set:
1. To evaluate the role of peripheral blood platelet and erythrocyte germ morphometry data in Covid-19 outcomes in haematological patients.
2. Identification of predictors and risk factors for "difficult" tracheal intubation while ensuring patency of the upper respiratory tract during general anaesthesia.

1. Interval pattern recognition and its properties

Suppose there is a set of observations $\{(x_1, \mathbf{y_1}), \dots, (x_n, \mathbf{y_n})\}$ monitoring the condition of patients after the applied treatment. The x_k element can take the values 0 or 1, let's call it the main feature. The main feature is $x_k = 1$, if the outcome of treatment is negative, $x_k = 0$, if the outcome of treatment is positive. The vector $\mathbf{y_k} = (y_{k1}, \dots, y_{km})$, $k = 1, \dots, n$, is a vector of accompanying features and consists of m components determined by laboratory methods. Let's denote $K = \{k : \ x_k = 1\}$ and put $S = \{1, \dots, n\} \setminus K$. Our task is to recognize the vector of the concomitant feature $\mathbf{y_k}$ of the treatment outcome, i.e. belonging to the index k the set of K or S.

This problem is solved using the following interval recognition procedure. First, the boundaries of the segments to which the components y_{ki}, $k \in K$, of the vectors of accompanying features belong are constructed. For each i, $1 \leq i \leq m$, we calculate

$$z_i = \min_{k \in K} y_{ki}, \ Z_i = \max_{k \in K} y_{ki}, \ i = 1, \dots, m. \tag{1}$$

Then the decisive rule determining whether the index j belongs to the set K is determined by the condition

$$y_{ji} \in [z_i, \ Z_i], \ i = 1, \dots, m. \tag{2}$$

Such a crucial rule always correctly recognizes $j \in K$. Denote by L the set of indices that are correctly recognized and belong to the set of S. Then the quality of interval recognition is determined by the equality $\rho = \dfrac{|K| + |L|}{n}$. Here $|J|$ means the number of elements of the set $J \subseteq \{1, \dots, n\}$. The ρ value characterizes the proportion of correctly recognized objects from the set $\{1, \dots, n\}$.

The interval recognition rule has the following properties, which are easy to establish by mathematical induction.
1. All objects from the K risk zone are recognized correctly. This statement automatically follows from the formulas (1), (2).
2. Increasing the number of components of the vector $\mathbf{y}_k = (y_{k1}, \dots, y_{km})$ does not decrease (and in many cases increases) the quality of interval recognition. Indeed, the addition of a new feature (the component of the vector $\mathbf{y}_k$) does not change the correct recognition of the index $k \in K$, but reduces the possibility of incorrect recognition of the index $k \in L$.
3. Dividing the sample $\{1, \dots, n\}$ into disjoint subsamples (for example, $\{1, \dots, n_*\}$, $\{n_* + 1, \dots n\}$) does not decrease (and in many cases increases) the quality of interval recognition. The proof of this property repeats the proof of the previous property of the interval pattern recognition algorithm.
4. The computational complexity of implementing interval recognition depends linearly on the number of objects n and on the number of m components of the vectors of associated features. This allows you to build fairly fast data processing algorithms. This statement automatically follows from the formulas (1), (2).

2. Identification of predictors and risk factors for "difficult" tracheal intubation with artificial ventilation during general anaesthesia in patients with thyroid diseases

Some types of general anaesthesia involve prosthetics of the respiratory function using artificial respiration devices [1]. During anaesthesia, various methods are used to ensure patency of the upper respiratory tract, but in the vast majority of cases it is tracheal intubation. In this case, breathing is carried out through a special device ? an endothalial tube, which is placed in the trachea and connected to an artificial respiration machine. In some cases, the process of visualizing the trachea is difficult for objective reasons, so examination of the glottis may be partially difficult or not possible. As a result, the endothelial tube cannot be advanced and a critical situation arises: "I cannot intubate, I cannot ventilate."

After the operation, the patient regains independent breathing, wakes up, and the tube is removed. In some cases, the process of placing an endothelial tube into the trachea is fraught with difficulties. They are mainly related to the structural features of the respiratory tract and the presence of pathology.

There are various approaches (scales) for assessing the respiratory tract and identifying risks when placing an endothelial tube. All scales are focused on the assessment of anatomical parameters, but no approach provides an accurate guarantee of prognosis. Moreover, none of the scales assesses the thyroid gland, an organ intimately located near the respiratory tract, which, when enlarged, can lead to changes in anatomy. Accordingly, the development of methods for assessing the respiratory tract and predicting "difficult" ventilation and tracheal intubation (placing a tube in the trachea) is relevant.

A prospective randomized trial was conducted at the Regional Clinical Hospital 2, Vladivostok in 2025. Before the planned elective surgery and anaesthetic care, all patients underwent the required amount of laboratory and instrumental examinations. In the absence of deviations from the reference values, surgical intervention was planned. On the eve of the upcoming surgery, the patients were examined by an anaesthesiologist.

The study group consisted of 35 patients who had scheduled surgery for thyroid disease. The control group consisted of 21 patients who underwent hernioplasty.

Inclusion criteria: age over 18 years; ability to lie on your back; physical status according to the American Society of Anaesthesiologists (ASA) I-III art.; general anaesthesia with tracheal intubation. Criteria for: inability to stand horizontally on the operating table; physical status according to ASA IV-VI; pregnancy; emergency surgical procedures. In the preoperative period, all patients underwent a routine assessment of the respiratory tract in the sitting and supine positions.

Primary endpoints: the number of tracheal intubation attempts, optimization of tracheal intubation. Secondary points: ventilation - masked artificial ventilation of the lungs before intubation, laryngoscopy - visualization of the glottis, intubation - "difficult" intubation - placement of a tube into the trachea, classification of laryngeal structures obtained during visualization during direct laryngoscopy (R.S. Cormack and J.R. Lehane modified by T.M. Cook [4]) (see, [5]).

The study group consisted of 35 patients. The influence of the following features on the primary and secondary endpoints was evaluated: body mass index, the volume of the thyroid gland (cm^3), assessment of the respiratory tract by Mallampati, LEMON scale for predicting difficult tracheal intubation (L (Look), E (Evaluate), M (Mallampati), O (Obstruction), N (Neck mobility); M.J. Reed, 2004) [2]), a test with biting the red border of the upper lip, the distance from the skin to the epiglottis (mm), hyomental distance (mm), the distance from the skin to the sublingual bone(mm, not determ ined in all patients), the tongue thickness of the (mm).

Table 1: *RC during surgery and thyroid surgery (RC - recognition coefficient).*

	RC during surgical intervention	Thyroid surgery gland (thyroid gland)
Ventilation	1	1
Intubation	0.971429	0.971429
More than 1 attempt Tracheal Intubation	0.971429	0.971429
Laryngoscopy	1	0.742857
Optimization of execution Tracheal intubation	1	0.714286
Cormack Lehane	1	0.6

Table 1 shows the results of RC calculations for various characteristics without dividing the sample by thyroid volume. It can be seen from it that during surgical intervention the RC is very high, and in some cases the RC for the thyroid gland is lower. To increase thyroid RC, the initial sample was experimentally divided into intervals (above and below a certain value). Some results of the comparative analysis of the calculations performed are given in the Table 2. You can see how the RC is rising.

Table 2: *RC in surgical intervention and on the thyroid gland (RC - recognition coefficient, CR - correctly recognized).*

Laryngoscopy	Thyroid Volume > 40	Thyroid Volume $\leq$40	Throughout range	by intervals
Total	7	28	35	35
Critical value 1 hline Uncritical value 0	2 5	8 20	10 25	10 25
CR with the value 0	5	17	16	5+7=12
RC	1	0.89286	0.74286	0.91429
Intubation	**Thyroid Volume > 40**	**Thyroid Volume $\leq$40**	**Throughout range**	**by intervals**
Total	7	28	35	35
Critical value 1 hline Uncritical value 0	2 5	7 21	9 26	9 26
CR with the value 0	5	17	16	5+17=22
RC	1	0.85714	0.71429	0.88571
Cormack Lehane	**Thyroid Volume > 50**	**Thyroid Volume $\leq$ 50**	**All Over range**	**by Intervals**
Total	7	28	35	35
Critical values 2,3,4	2	11	13	13
Uncritical values 0.1	5	17	22	22
CR with values 0,1	5	8	8	5+87=13
RC	1	0.67857	0.6	0.74286

3. Investigation of platelet and erythrocyte germs of haematologesis in haematological patients with COVID-19

Patients with malignant blood diseases are at higher risk of severe COVID-19 than the general population [3]-[5]. This is due to immunosuppression caused by both the disease itself and specific antitumor therapy. During the ongoing pandemic of coronavirus infection, identifying risk factors for an adverse outcome of the disease is an urgent task for the healthcare system. Monitoring of haematological abnormalities during the course of COVID-19 in patients with haematologic malignancies is critical, allowing assessing the severity of the condition and the risk of adverse events.

A general blood test is available at any medical facility. COVID-19 causes various haematological abnormalities, including anemia and thrombocytopenia. These changes are much more pronounced in patients with severe COVID-19 disease and, thus, can serve as a possible biomarker for those who need hospitalization and treatment in the intensive care unit. Anemia and thrombocytopenia in haematological patients with COVID-19 serve as indicators of severe course and risk of multiple organ failure. Their dynamic evaluation makes it possible to predict outcomes and optimize therapy. Special attention is required for patients with malignant haematological diseases due to the risk of life-threatening complications.

Data from medical records of 44 patients with blood tumor diseases who were treated for COVID-19 coronavirus infection in the infectious diseases department of Regional Clinical Hospital 2 in Vladivostok in the period from 2020-2022. The interval pattern recognition method using a zero-one matrix was used to process medical information. This made it possible to highlight several groups patients and compare the recognition quality between the groups.
Group 1 (characteristics of platelet parameters): PLT(10 in 9/L) - total platelet count, MPV(fL) - average platelet volume, PDW(fL) - the width of platelet distribution, PCT - thrombocrit.
Group 2 (characteristics of erythrocyte parameters): RBC - total number of red blood cells, HGB (g/L) - hemoglobin level, MCV (fL) - the average volume of a red blood cell, MCH (pg) - the hemoglobin content in the erythrocyte.

Table 3 shows the recognition results for different patient groups from the initial sample, which included 44 patients. Table 4 shows the recognition results for a reduced sample of 29 patients. The division of this sample into sub samples (see Table 4) significantly affects the recognition coefficient. Moreover, depending on the age category, the quality of recognition for the 1st group of signs (platelets) increases with increasing age of the examined, and the quality of recognition for the 2nd group of signs (erythrocytes) decreases.

Based on the results obtained (in a small sample), patients in the range of 61-69 years have a risk of death in the presence of anemia. In patients over 70 years of age, thrombocytopenia and a decrease in platelet parameters are of prognostic importance and can be regarded as risk factors for an unfavorable outcome.

Table 3: *Recognition results for different patient groups from the original sample (MT – malignant tumors, M – men, W – women).*

patients	all	W	M	from 70 years old	61 – 69 years old	up to 60 years old	MT
Total	44	23	21	10	19	15	36
The dead	16	8	8	6	6	4	15
Survivors	28	15	13	4	13	11	21
CR survivors (platelets)	8	7	8	4	9	5	5
RC(platelets)	0.54	0.65	0.76	1	0.79	0.6	0.56
CR survivors (red blood cells)	8	4	6	2	10	11	7
RC (red blood cells)	0.54	0.52	0.67	0.8	0.84	1	0.61

Table 4: *Results recognition results for different groups of a reduced sample of 29 patients.*

patients	from 70 years old	61 – 69 years old	from 61 years old (by whole group)	from 61 years old (by subgroups)
Total	10	19	29	29
Deceased	6	6	12	12
Survivors	4	13	17	17
CR survivors (platelets)	4	9	8	4+9=13
RC (platelets)	1	0.79	0.69	0.86
CR survivors (red blood cells)	2	10	4	2+10=12
RC(erythrocytes)	0.8	0.84	0.55	0.83

The method proposed in the work, despite the small sample size, allows for a fairly contrasting comparison of the state of platelet and erythrocyte hematologesis germs at risk of fatal events in haematological patients who have suffered COVID-19.

4. Conclusion

Due to its high speed, the interval pattern recognition method proved to be convenient for processing large data and showed its effectiveness and practical significance in the two medical studies presented.

The research was carried out within the state assignment for IAM FEB RAS N 075-00459-25-00.

References

[1] Cook, T. M. (2000). A new practical classification of laryngeal view. *Anaesthesia*, 55: 275–279.

[2] Dunts, P. V. et al. (2023). Modifed Delphi analysis of the guidelines of the All-Russian public organization "Federation of Anesthesiologists and Reanimatologists" "The use of non-invasive lung ventilation". *Annals of Critical Care*, 4: 99–111.(In Russian). DOI: 10.21320/1818-474X-2023-4-99-111.

[3] Pinana, J. L. et al. (2020) Risk factors and outcome of COVID-19 in patients with haematological malignancies. *Experimental haematology and oncology*, 9: 21–25. DOI: 10.1186/s40164-020-00177-z

[4] Gossell-Williams M. Faculty Opinions recommendation of Case Fatality Rate of Cancer Patients with COVID-19 in a New York Hospital System. Faculty Opinions. Post-Publication Peer Review of the Biomedical Literature, 2020.

[5] Kaprin, A. D. et al. (2020). The impact of the COVID-19 pandemic on cancer practice. *Siberian Journal of Oncology*, 19(3): 5–22 (In Russian). DOI: 10.21294/1814-4861-2020-19-3-5-22.