

DOI: 10.12731/wsd-2017-4-106-121

УДК 616-093/-098

ОЦЕНКА ИНФОРМАТИВНОСТИ И ОТБОР ПРИЗНАКОВ ПРИ ИДЕНТИФИКАЦИИ ОБЪЕКТОВ НА ЦИФРОВЫХ ИЗОБРАЖЕНИЯХ МИКРОСКОПИЧЕСКИХ ПРЕПАРАТОВ, ОКРАШЕННЫХ ПО МЕТОДУ ЦИЛЯ-НИЛЬСЕНА

*Наркевич А.Н., Виноградов К.А., Корецкая Н.М.,
Катаева А.В., Журбенко Е.О.*

***Цель.** Оценка информативности и осуществление отбора признаков для классификации объектов, на цифровых изображениях микроскопических препаратов, окрашенных по методу Циля-Нильсена с помощью методов Шеннона и Кульбака, а также сравнение результатов их применения.*

***Материалы и методы.** Использовались данные о 343 687 объектах, выделенных на цифровых изображениях микроскопических препаратов: 6 708 объектов – кислотоустойчивые микобактерии, 336 979 – иные объекты. Анализ объектов производился по 240 цветовым и морфометрическим признакам, оценка информативности которых осуществлялась с помощью методов Шеннона и Кульбака. Оценка качества отбора признаков осуществлялась с применением наивного Байесовского классификатора.*

***Результаты.** Наибольшую информативность по методу Шеннона имели цветовые признаки объектов и соотношения их размеров, а наибольшая доля верно классифицированных объектов (84,6%) была достигнута при включении в классификационную модель 6 признаков с наибольшей информативностью по методу Шеннона. Наибольшую информативность по методу Кульбака имели радиальные размеры объектов и их соотношения, а наибольшая доля верно классифицированных объектов (78,0%) была достигнута при включении в классификационную модель 14 признаков с наибольшей информативностью по методу Кульбака.*

***Заключение.** Методы Шеннона и Кульбака могут применяться для снижения признакового пространства при классификации объектов. Метод Шеннона позволяет в большей мере сократить количество при-*

знаков, обеспечивая при этом наибольшую долю верной классификации объектов.

Ключевые слова: снижение признакового пространства; информативность признаков; кислотоустойчивые микобактерии; бактериоскопическая диагностика.

ESTIMATING OF INFORMATIVITY AND SELECTION OF FEATURES IN IDENTIFICATION OF OBJECTS IN DIGITAL IMAGES OF MICROSCOPIC PREPARATIONS, STAINED BY THE METHOD OF ZIEHL-NIELSEN

*Narkevich A.N., Vinogradov K.A., Koretskaya N.M.,
Kataeva A.V., Zhurbenko E.O.*

Background. *Selecting features for classifying objects in digital images of microscopic preparations, stained by the method of Ziehl-Nielsen with the use of methods for evaluating the informativeness of the Shannon and Kullback, and a comparison of the results of their application.*

Materials and methods. *Used data on 343 687 the selected objects on digital images of microscopic preparations: 6 708 objects acid-fast bacilli, 336 979 – other objects. The analysis of items was carried out for 240 color and morphometric characteristics, evaluating the information content of which was carried out using the methods of Shannon and Kullback. Quality evaluation of selection traits were carried out using the naive Bayesian classifier.*

Result. *The most informative method Shannon had the color characteristics of objects and the ratio of their sizes, and the highest proportion of correctly classified objects (84,6%) were achieved by inclusion in the classification model 6 characteristics with the highest informativeness according to the method of Shannon. The most informative method of the Kullback had radial sizes of objects and their relationships, and the highest proportion of correctly classified objects (78,0%) was achieved when included in a classification model of the 14 signs from the most informative method of Kullback.*

Conclusion. *Methods of estimating the information content of the Shannon and Kullback, can be used in the reduced feature space for classification problems use the method of Shannon allows to a greater extent to reduce the*

number of signs and thus provides the greatest proportion of correct object classification.

Keywords: *dimension reduction; information content of features; acid-fast bacilli; smear diagnosis.*

Задача отбора признаков или снижения признакового пространства часто возникает при построении и исследовании различных медицинских прогностических моделей [1, 2, 3, 4]. Уменьшение числа исследуемых признаков направлено на исключение их мультиколлинеарности [5, 6], снижение необходимых вычислительных мощностей [7, 8, 9], экономию затрат, требуемых на медицинские исследования для получения дополнительных сведений о пациенте или объекте [10]. С задачей отбора признаков приходится также сталкиваться при создании логистических регрессионных уравнений [5, 6], линейных классификаторов [11] и других математических моделей различных предметных областей [7, 12, 13].

Одной из медицинских проблем, для которой весьма актуален отбор наиболее значимых признаков, – классификация объектов на цифровых изображениях мокроты, окрашенной по методу Циля-Нильсена. Метод Циля-Нильсена относится к специализированным методам окраски биологического материала для поиска кислотоустойчивых микобактерий (КУМ) при бактериоскопической диагностике туберкулеза.

Целью работы является применение и сравнительный анализ методов Шеннона и Кульбака для отбора признаков при классификации объектов на цифровых изображениях микроскопических препаратов, окрашенных по методу Циля-Нильсена.

Материалы и методы исследования

Материалом исследования явились 343 687 объектов, выделенных с использованием одного из разработанных нами алгоритмов, на цифровых изображениях микроскопических препаратов мокроты, окрашенной по методу Циля-Нильсена [14], полученных с помощью тринокулярного микроскопа Микромед 1 вар. 3–20 при увеличении 10×60 с установленной цифровой камерой TopCam UCМOS01300KPA с разрешением 0,3 МР. Каждое изображение имело разрешение 572 x 422 пикселей. Среди выделенных объектов 6 708 объектов являются КУМ, а 336 979 объектов – иные объекты (артефакты, скопления мокроты, части клеток и т.д.).

Анализ данных объектов производился по 240 цветовым и морфометрическим признакам. Данные признаки разделены на 3 группы, кото-

рые подразделяется на подгруппы. Для анализа использована следующая классификация признаков:

1. Основные морфометрические признаки объектов:
 - а) попиксельная площадь объектов (в пикселях);
 - б) размер объектов по оси X (в пикселях);
 - в) размер объектов по оси Y (в пикселях).
2. Радиальные размеры объектов и их соотношения:
 - а) размеры от центра объекта до его края (всего 72 параметра);
 - б) соотношения противоположащих размеров от центра объекта до его края (всего 36 параметров);
 - в) соотношения перпендикулярных размеров от центра объекта до его края (всего 72 параметра);
 - г) соотношения перпендикулярных размеров от одного края объекта до его противоположного края (всего 36 параметров).
3. Цветовые признаки объектов:
 - а) характеристики цвета объектов в цветовой схеме RGB (всего 9 параметров);
 - б) характеристики цвета объектов в цветовой схеме HSV (всего 9 параметров);
 - в) характеристики цвета объектов в оттенках серого (всего 3 параметра).

Основные морфометрические признаки объектов, радиальные размеры объектов и их соотношения измеряются в пикселях. Измерение цветовых признаков осуществлялось в безразмерных единицах, однако интервал возможных значений отличался в зависимости от цветовой схемы. Так, цвета объектов в цветовой схеме RGB и в оттенках серого измерялись в интервале от 0 до 255, параметры, отражающие H в цветовой схеме HSV, измерялись в интервале от 0 до 359, а параметры, характеризующие S и V, измерялись в интервале от 0 до 100.

Оценка информативности признаков изучаемых объектов осуществлялась с помощью методов Шеннона [15, 16] и Кульбака [17, 18, 19, 20, 21].

Информативность по методу Шеннона определялась с помощью следующей формулы:

$$I(x_i) = 1 + \sum_{i=1}^G (P_i \sum_{k=1}^K P_{i,k} * \log_k P_{i,k}) \quad (1)$$

где $I(x_i)$ – информативность признака x_i , G – количество интервалов признака, P_i – вероятность i интервала признака, K – количество классов признака (равно 2 – класс КУМ и класс иных объектов), $P_{i,k}$ – вероятность появления i интервала признака в k классе.

Вероятность i интервала признака (P_i), рассчитывалась по формуле:

$$P_i = \frac{\sum_{k=1}^K m_{i,k}}{N} \quad (2)$$

где $m_{i,k}$ – частота появления i интервала в k классе, K – количество классов признака, N – общее число объектов.

Вероятность появления i интервала признака в k классе ($P_{i,k}$) рассчитывалась по формуле:

$$P_{i,k} = \frac{m_{i,k}}{\sum_{k=1}^K m_{i,k}} \quad (3)$$

где $m_{i,k}$ – частота появления i интервала в k классе, K – количество классов признака.

Информативность по методу Кульбака определялась с помощью следующей формулы:

$$I(x_i) = \sum_{i=1}^G [P_{i1} - P_{i2}] * \log_2 \frac{P_{i1}}{P_{i2}} \quad (4)$$

где $I(x_i)$ – информативность признака x_i , G – количество интервалов признака, P_{i1} – вероятность появления i интервала признака в первом классе (в классе объектов КУМ), P_{i2} – вероятность появления i интервала признака во втором классе (классе объектов не КУМ).

Вероятность появления i интервала признака в первом классе (P_{i1}) рассчитывалась по формуле:

$$P_{i1} = \frac{m_{i1}}{\sum_{k=1}^G m_{i1}} \quad (5)$$

где m_{i1} – частота появления i интервала в первом классе (классе объектов КУМ).

Вероятность появления i интервала признака во втором классе (P_{i2}) рассчитывалась по формуле:

$$P_{i2} = \frac{m_{i2}}{\sum_{k=1}^G m_{i2}} \quad (6)$$

где m_{i2} – частота появления i интервала во втором классе (классе объектов не КУМ).

Количество интервалов признаков рассчитывалось с использованием формулы Стерджеса [22]:

$$n = 1 + [\log_2 N] \quad (7)$$

где n – количество интервалов признака, N – общее число объектов.

После определения информативности всех признаков формировался упорядоченный список, в котором на первом месте находился признак с

наибольшей информативностью, а на последнем месте – признак с наименьшей информативностью. Для оценки качества отбора признаков все объекты (343 687) были разделены на обучающее (309 318 объектов) и тестовое множество (34 369 объектов).

Оценка качества отбора признаков производилась с использованием наивного Байесовского классификатора в несколько итераций. На первой итерации для построения модели классификации объектов (КУМ и иные объекты) на обучающем множестве использовался один признак с наибольшей информативностью. На каждой следующей итерации к признакам, на основе которых ранее осуществляется классификация объектов, добавлялся еще один признак, следующий по уменьшению значения информативности. Также на каждой итерации с использованием тестового множества вычислялась доля правильно классифицированных объектов. Обучение моделей наивного Байесовского классификатора выполнялось с использованием KNIME Analytics Platform v.3.1.2.

Результаты исследования и их обсуждение

По результатам наивного байесовского классификатора были получены списки признаков, упорядоченные по убыванию их информативности. На рисунке 1 представлен список из 15 признаков, имеющих наибольшую информативность по методу Шеннона. Приведенные на рисунке 1 данные свидетельствуют о том, что наибольшую информативность по методу Шеннона имеют цветковые признаки классифицируемых объектов и соотношения размеров данных объектов.

Несмотря на то, что признак «Минимальное значение N» имеет большую информативность по сравнению с другими признаками, как видно из данных приведенных на рисунке 2, наибольшая доля верно классифицированных объектов была достигнута на 6 итерации, когда в модель наивного байесовского классификатора были включены следующие 6 признаков: «Минимальное значение N» (информативность по методу Шеннона – 0,9634), «Среднее значение B» (0,7114), «Максимальное значение B» (0,6856), «Соотношение 25 и 43» (0,6833) и «Соотношение 26 и 44» (0,6752) радиальных размеров объектов и «Среднее значение S» (0,6715).

Приведенные выше результаты связаны с тем, что высокое значение информативности того или иного признака не всегда позволяет однозначно классифицировать объекты, а добавление в классификационную

модель признаков, имеющих небольшую информативность приводит к лучшим результатам классификации. Например, включение в классификационную модель только «Минимального значения Н», имеющего максимальную информативность по методу Шеннона (0,9634), позволяет классифицировать объекты с 50,4% долей верной классификации, а добавление такого признака как «Среднее значение В» со сравнительно меньшей информативностью по методу Шеннона (0,7114), позволяет увеличить долю верной классификации до 85,8%.



Рис. 1. Признаки, имеющие наибольшую информативность по методу Шеннона

На рисунке 3 представлен список из 15 признаков, имеющих наибольшую информативность по методу Кульбака. Данные, приведенные на рисунке 3, указывают на то, что наибольшую информативность по методу Кульбака имеют радиальные размеры объектов и их соотношения. Это существенно отличается от списка признаков, имеющих наибольшую информативность по методу Шеннона. Данный факт объясняется тем, что методы Шеннона и Кульбака по-разному определяют информативность. Так, метод Шеннона оперирует средневзвешенным количеством информации, приходящимся на различные интервалы признаков, а метод Кульбака оценивает расхождение между двумя классами.

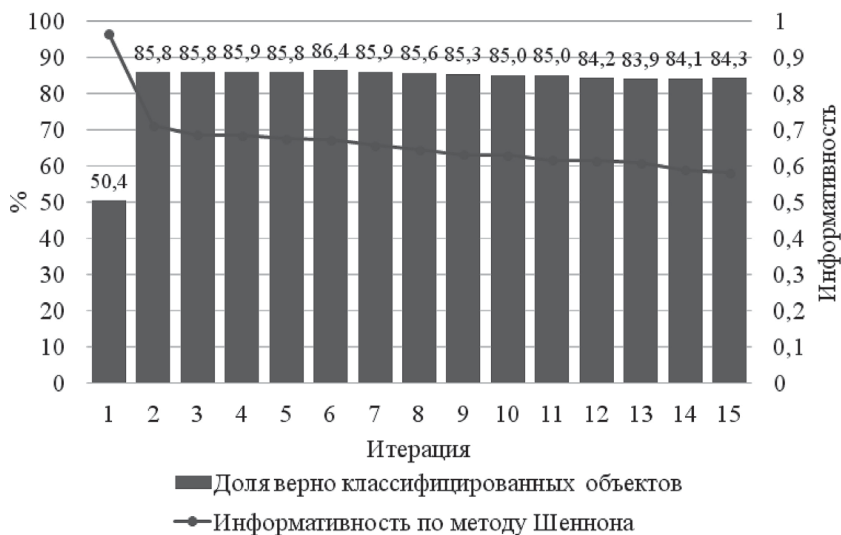


Рис. 2. Информативность признаков по методу Шеннона, добавляемых на каждой итерации, и изменение доли верно классифицированных объектов

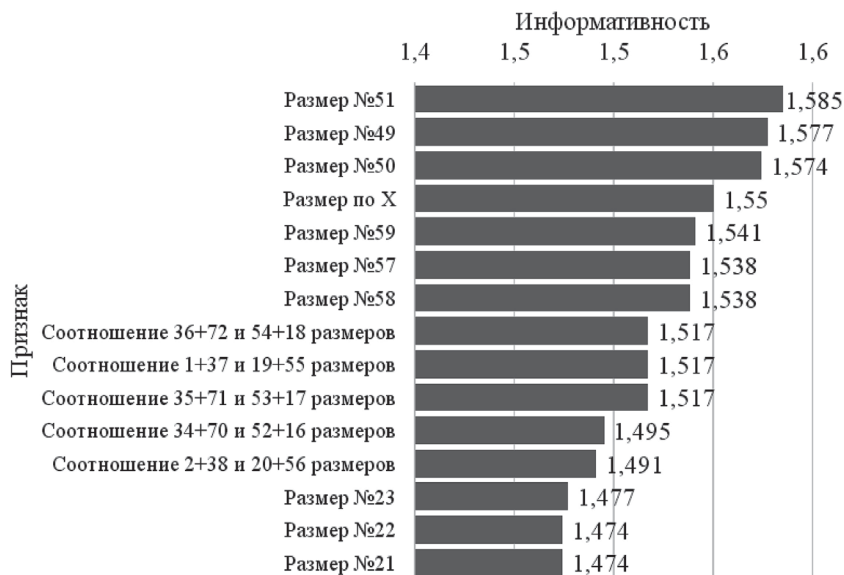


Рис. 3. Признаки, имеющие наибольшую информативность по методу Кульбака

Анализ признаков, имеющих наибольшую информативность по методу Кульбака, показывает, что среди них отсутствуют признаки с существенно большей информативностью по сравнению с остальными. При этом включение в классификационную модель признака, имеющего наибольшую информативность, позволило получить долю верной классификации – 74,2%. Однако дальнейшее включение признаков на последующих итерациях не дало существенного увеличения доли верно классифицированных объектов (до 75,6% на 13-ой итерации). Лишь добавление 14 признаков, имеющих наибольшую информативность по методу Кульбака, позволило добиться максимальной доли верно классифицированных объектов – 78,0%.

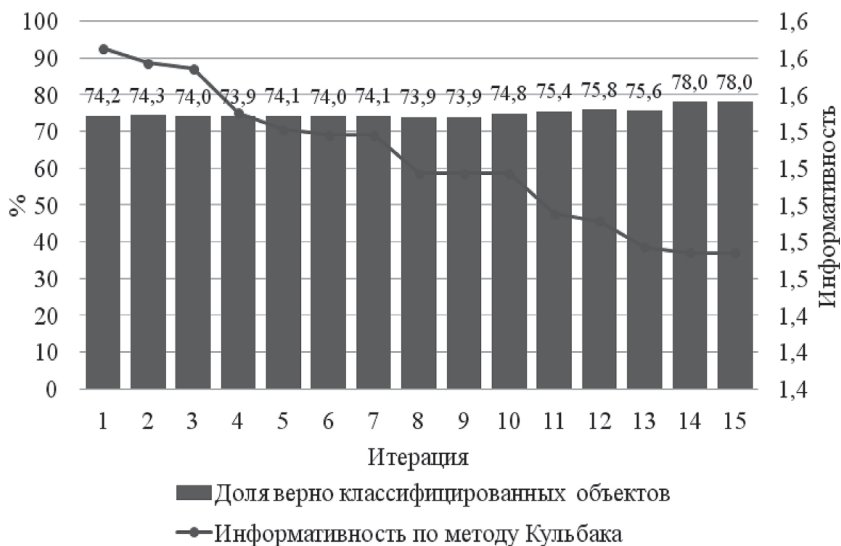


Рис. 4. Информативность признаков по методу Кульбака, добавляемых на каждой итерации, и изменение доли верно классифицированных объектов

Несмотря на достижение максимальной доли верной классификации объектов (78,0%) на 14-ой итерации обучения наивного Байесовского классификатора с включением признаков, имеющих наибольшую информативность по методу Кульбака, данный показатель был меньше, чем максимальная доля верной классификации (86,4%) на 6-ой итерации обучения классификационной модели с включением признаков, имеющих наибольшую информативность по методу Шеннона.

Заключение

Таким образом, вычислительные эксперименты показали, что методы оценки информативности Шеннона и Кульбака могут применяться для снижения признакового пространства при построении и исследовании медицинских прогностических моделей и в частности при классификации объектов, выделенных на цифровых изображениях микроскопических препаратов мокроты, окрашенной по методу Циля-Нильсена. При этом, использование метода Шеннона позволяет в большей мере сократить количество признаков, обеспечивая при этом значительную долю верной классификации объектов.

Информация о конфликте интересов. Конфликт интересов отсутствует.

Информация о спонсорстве. Финансовая поддержка данного исследования не осуществлялась.

Благодарности. Отсутствуют.

Список литературы

1. Kaplan A., Lock E.F. Prediction with dimension reduction of multiple molecular data sources for patient survival // *Cancer informatics*. 2017. № 16. e1176935117718517.
2. Automated seizure detection using limited-channel EEG and non-linear dimension reduction / Birjandtalab J., Baran Pouyan M., Cogan D., Nourani M., Harvey J. // *Computers in biology and medicine*. 2017. № 82, pp. 49–58.
3. An estimating equation approach to dimension reduction for longitudinal data / Xu K., Guo W., Xiong M., Zhu L., Jin L. // *Biometrika*. 2016. № 1, pp. 189–203.
4. Лосева Е.Д., Сергиенко Р.Б. О методах отбора информативных признаков с привлечением самоорганизующихся нейросетевых классификаторов и их ансамблей // *Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева*. 2016. № 4. С. 878–882.
5. Леонтьева Л.Н. Последовательный выбор признаков при восстановлении регрессии // *Машинное обучение и анализ данных*. 2012. № 3. С. 335–346.
6. Попков А.В., Настенко Е.А. Многокритериальный алгоритм пошагового отбора признаков при восстановлении регрессии // *Инновационные технологии в науке и образовании*. 2015. № 1. С. 267–268.
7. Гулин В.В. Методы снижения размерности признакового описания документов в задаче классификации текстов // *Вестник МЭИ*. 2013. № 2. С. 115–121.
8. Мясников Е.В. Анализ методов снижения размерности в задаче представления коллекций цифровых изображений // *Компьютерная оптика*. 2008. № 3. С. 296–301.

9. Liu Y., Chiaromonte F., Li B. Structured Ordinary Least Squares: A Sufficient Dimension Reduction approach for regressions with partitioned predictors and heterogeneous units // *Biometrics*. 2017. № 2, pp. 529–539.
10. Грицинская В.Л., Москаленко О.Л. Использование компьютерных технологий при проведении диспансеризации детского населения Республики Тыва // *В мире научных открытий*. 2017. № 2. С. 158–167.
11. Гайдель А.В., Крашенинников В.Р. Отбор признаков для задачи диагностики остеопороза по рентгеновским изображениям шейки бедра // *Компьютерная оптика*. 2016. № 6. С. 939–946.
12. Supervised nonlinear dimension reduction of functional magnetic resonance imaging data using Sliced Inverse Regression / Tu Y., Tan A., Fu Z., Sam Hung Y., Hu L., Zhang Zh. // *Conference proceedings: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Milan, 2015, pp. 2641–2645.
13. Ivan G., Grolmusz V. On dimension reduction of clustering results in structural bioinformatics // *Biochimica et biophysica acta*. 2014. № 12, pp. 2277–2283.
14. Наркевич А.Н. Алгоритмы сегментации цифровых микроскопических изображений мокроты, окрашенной по методу Циля-Нильсена // *World Science: Proceedings of articles the international scientific conference*. Киров: MCNIP LLC, 2017. С. 431–436.
15. Шеннон К. Работы по теории информации и кибернетике. М.: Иностранная литература, 1963. 832 с.
16. Быкова В.В., Катаева А.В. Методы и средства анализа информативности признаков при обработке медицинских данных // *Программные продукты и системы*. 2016. № 2. С. 172–178.
17. Кульбак С. Теория информации и статистика. М.: Наука, 1967. 408 с.
18. Львович И.Я., Гладских Н.А., Казьмина И.В. Алгоритмическое обеспечение вероятностной оценки рецидива инсульта с использованием методики формирования словаря информативных признаков на основе критерия Кульбака // *Вестник Воронежского государственного технического университета*. 2009. № 9. С. 56–57.
19. Математическое и алгоритмическое обеспечение расчета медико-социальных признаков заболеваний тканей пародонта у взрослого населения / Шлыкова Е.А., Есауленко И.Э., Косолапов В.П., Гладских Н.А. // *Системный анализ и управление в биомедицинских системах*. 2014. № 4. С. 947–951.
20. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. Новосибирск: Издательство ИМ СО РАН, 1999. 270 с.

21. Прикладная статистика: Классификация и снижение размерности: справочник / Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. М.: Финансы и статистика, 1989. 250 с.
22. Sturges H. The choice of a class-interval. *Journal of the American Statistical Association*. 1926. № 21, pp. 65–66.

References

1. Kaplan A., Lock E.F. Prediction with dimension reduction of multiple molecular data sources for patient survival. *Cancer informatics*, 2017, № 16, e1176935117718517.
2. Birjandtalab J., Baran Pouyan M., Cogan D., Nourani M., Harvey J. Automated seizure detection using limited-channel EEG and non-linear dimension reduction. *Computers in biology and medicine*, 2017, № 82, pp. 49–58.
3. Xu K., Guo W., Xiong M., Zhu L., Jin L. An estimating equation approach to dimension reduction for longitudinal data. *Biometrika*, 2016, № 1, pp. 189–203.
4. Loseva E.D., Sergienko R.B. O metodakh otbora informativnykh priznakov s privilecheniem samoorganizuyushchikhsya neyrosetevykh klassifikatorov i ikh ansambley [On the methods of selection of informative signs with the involvement of a self-organizing neural network classifiers and their ensembles]. *Vestnik Sibirskogo gosudarstvennogo aerokosmicheskogo universiteta im. akademika M.F. Reshetneva* [Bulletin of Siberian state aerospace University. academician M. F. Reshetnev], 2016, № 4, pp. 878–882.
5. Leont'eva L.N. Posledovatel'nyy vybor priznakov pri vosstanovlenii regressii [Sequential feature selection when restoring a regression]. *Mashinnoe obuchenie i analiz dannykh* [Machine learning and data analysis], 2012, № 3, pp. 335–346.
6. Popkov A.V., Nastenko E.A. Mnogokriterial'nyy algoritm poshagovogo otbora priznakov pri vosstanovlenii regressii [Multicriteria algorithm step-by-step selecting features when restoring a regression]. *Innovatsionnye tekhnologii v nauke i obrazovanii* [Innovative technologies in science and education], 2015, № 1, pp. 267–268.
7. Gulin V.V. Metody snizheniya razmernosti priznakovogo opisaniya dokumentov v zadache klassifikatsii tekstov [Methods of reducing the dimension of feature vectors describing the documents in the task of text classification]. *Vestnik MEI* [Bulletin of MEI], 2013, № 2, pp. 115–121.
8. Myasnikov E.V. Analiz metodov snizheniya razmernosti v zadache predstavleniya kolleksiy tsifrovyykh izobrazheniy [Analysis of methods for dimensionality reduction in the representation of collections of digital images]. *Komp'yuternaya optika* [Computer optics], 2008, № 3, pp. 296–301.

9. Liu Y., Chiaromonte F., Li B. Structured Ordinary Least Squares: A Sufficient Dimension Reduction approach for regressions with partitioned predictors and heterogeneous units. *Biometrics*, 2017, № 2, pp. 529–539.
10. Gritsinskaya V.L., Moskalenko O.L. Ispol'zovanie komp'yuternykh tekhnologiy pri provedenii dispanserizatsii detskogo naseleniya Respubliki Tyva [The use of computer technologies at carrying out of prophylactic medical examination of children of the Republic of Tuva] // *V mire nauchnykh otkrytiy* [In the world of scientific discoveries]. 2017. № 2, pp. 158–167.
11. Gaydel' A.V., Krashennnikov V.R. Otbor priznakov dlya zadachi diagnostiki osteoporoza po rentgenovskim izobrazheniyam sheyki bedra [The selection of features to the diagnosis of osteoporosis at x-ray images of the femoral neck]. *Komp'yuternaya optika* [Computer optics], 2016, № 6, pp. 939–946.
12. Tu Y., Tan A., Fu Z., Sam Hung Y., Hu L., Zhang Zh. Supervised nonlinear dimension reduction of functional magnetic resonance imaging data using Sliced Inverse Regression. *Conference proceedings: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Milan, 2015, pp. 2641–2645.
13. Ivan G., Grolmusz V. On dimension reduction of clustering results in structural bioinformatics. *Biochimica et biophysica acta*, 2014, № 12, pp. 2277–2283.
14. Narkevich A.N. Algoritmy segmentatsii tsifrovyykh mikroskopicheskikh izobrazheniy mokroty, okrashennoy po metodu Tsilya-Nil'sena [Algorithms for segmenting digital microscopic images of sputum stained by the method of Ziehl-Nielsen]. *World Science: Proceedings of articles the international scientific conference*, Kirov: MCNIP LLC, 2017, pp. 431–436.
15. Shannon K. *Raboty po teorii informatsii i kibernetike* [Works on information theory and Cybernetics], Moscow: Inostrannaya literatura, 1963, 832 p.
16. Bykova V.V., Kataeva A.V. Metody i sredstva analiza informativnosti priznakov pri obrabotke meditsinskikh dannykh [Methods and means of analysis of the informative signs in the processing of medical data]. *Programmnye produkty i sistemy* [Software products and systems], 2016, № 2, pp. 172–178.
17. Kul'bak S. *Teoriya informatsii i statistika* [Information theory and statistics], Moscow: Nauka, 1967, 408 p.
18. L'vovich I.Ya., Gladskikh N.A., Kaz'mina I.V. Algoritmicheskoe obespechenie veroyatnostnoy otsenki retsidiva insul'ta s ispol'zovaniem metodiki formirovaniya slovary informativnykh priznakov na osnove kriteriya Kul'baka [Algorithmic support for probabilistic assessment of recurrence of stroke with the use of the method of forming a vocabulary of informative

- features based on the Kullback criterion]. *Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta* [Vestnik Voronezh state technical University], 2009, № 9, pp. 56–57.
19. Shlykova E.A., Esaulenko I.E., Kosolapov V.P., Gladskikh N.A. Matematicheskoe i algoritmicheskoe obespechenie rascheta mediko-sotsial'nykh priznakov zabolevaniy tkaney parodonta u vzroslogo naseleniya [Mathematical and algorithmic support of calculation of health and social signs of periodontal diseases among the adult population]. *Sistemnyy analiz i upravlenie v biomeditsinskikh sistemakh* [System analysis and management in biomedical systems], 2014, № 4, pp. 947–951.
20. Zagoruyko N.G. *Prikladnye metody analiza dannykh i znaniy* [Applied methods of data and knowledge analysis], Novosibirsk: Izdatel'stvo IM SO RAN, 1999, 270 p.
21. Ayvazyan S.A., Bukhshtaber V.M., Enyukov I.S., Meshalkin L.D. *Prikladnaya statistika: Klassifikatsiya i snizhenie razmernosti: spravochnik* [Applied statistics: Classification and reduction of dimensionality: Handbook], Moscow: Finansy i statistika, 1989, 250 p.
22. Sturges H. The choice of a class-interval. *Journal of the American Statistical Association*, 1926, № 21, pp. 65–66.

ДАННЫЕ ОБ АВТОРАХ

Наркевич Артем Николаевич, руководитель научно-исследовательской лаборатории медицинской кибернетики и управления в здравоохранении; кандидат медицинских наук
Красноярский государственный медицинский университет им. проф. В.Ф. Войно-Ясенецкого
ул. Партизана Железняка, 1, Красноярск, 660022, Российская Федерация
narkevichart@gmail.com

Виноградов Константин Анатольевич, заведующий кафедрой медицинской кибернетики и информатики, доктор медицинских наук, профессор
Красноярский государственный медицинский университет им. проф. В.Ф. Войно-Ясенецкого
ул. Партизана Железняка, 1, Красноярск, 660022, Российская Федерация
vinogardov16@yande.ru

Корецкая Наталия Михайловна, врач-фтизиатр, доктор медицинских наук, профессор
Медико-санитарная часть № 24
ул. Академгородок, д. 56а, стр. 1, Красноярск, 660036, Российская Федерация
koretskaya.nm@mail.ru

Катаева Алина Владимировна, аспирант кафедры высшей и прикладной математики
Институт математики и фундаментальной информатики Сибирского федерального университета
пр. Свободный, д.79, Красноярск, 660041, Российская Федерация
kataeva_av@mail.ru

Журбенко Екатерина Олеговна, студент 6 курса медико-психолого-фармацевтического факультета
Красноярский государственный медицинский университет им. проф. В.Ф. Войно-Ясенецкого
ул. Партизана Железняка, 1, Красноярск, 660022, Российская Федерация
vinogradova71@bk.ru

DATA ABOUT THE AUTHORS

Narkevich Artem Nikolaevich; Head of the Research Laboratory of Medical Cybernetics and Management in Healthcare, Candidate of Medical Sciences
Krasnoyarsk State Medical University
1, Partizana Zheleznyaka Str., Krasnoyarsk, 660022, Russian Federation
narkevichart@gmail.com
SPIN-code: 9030-1493
ORCID: 0000-0002-1489-5058
ResearcherID: H-5830-2012
Scopus Author ID: 55810287600

Vinogradov Konstantin Anatol'evich, Head of the Department of Medical Cybernetics and Informatics; Doctor of Medical Sciences; Professor
Krasnoyarsk State Medical University
1, Partizana Zheleznyaka Str., Krasnoyarsk, 660022, Russian Federation

vinogradov16@yandex.ru
SPIN-code: 6924-0110
ORCID: 0000-0001-6224-5618
ResearcherID: M-2823-2014
Scopus Author ID: 57193429569

Koretskaya Nataliya Mikhaylovna, Doctor-Phthisiatrician, Doctor of Medical Sciences, Professor
Medical-Sanitary Department № 24
56a, Akademgorodok, Krasnoyarsk, 660036, Russian Federation
koretskaya.nm@mail.ru
SPIN-code: 8566-5263
ORCID: 0000-0002-2332-912X
ResearcherID: M-5450-2014
Scopus Author ID: 7801569495

Kataeva Alina Vladimirovna, Postgraduate at the Department of Higher and Applied Mathematics of Institute of Mathematics and Fundamental Informatics
Siberian Federal University
79, Svobodnyy str., Krasnoyarsk, 660041, Russian Federation
alisite@mail.ru
SPIN-code: 8492-0222
ORCID: 0000-0002-4840-3314
Scopus Author ID: 57194972802

Zhurbenko Ekaterina Olegovna, Student of the 6th Course of Medical-Psychological-Pharmaceutical Faculty
Krasnoyarsk State Medical University
1, Partizana Zheleznaya Str., Krasnoyarsk, 660022, Russian Federation
vinogradova71@bk.ru
ORCID: 0000-0002-8262-5199
ResearcherID: P-2634-2017