

**ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ  
В ДИАГНОСТИКЕ АНЕМИИ ПО КЛИНИЧЕСКИМ ПОКАЗАТЕЛЯМ**В. В. БОЖЕНКО<sup>1\*</sup>, Н. Ю. ЧЕРНЫШ<sup>2</sup>, Т. М. ТАТАРНИКОВА<sup>1</sup>

<sup>1</sup>Санкт-Петербургский государственный университет аэрокосмического приборостроения,  
Санкт-Петербург, Россия  
\* vibozenko@yandex.ru

<sup>2</sup>Национальный медицинский исследовательский центр им. В. А. Алмазова,  
Санкт-Петербург, Россия

**Аннотация.** Исследован набор медицинских данных, полученных из информационной системы сетевой лаборатории амбулаторного наблюдения, который содержит показатели анализов пациентов с диагнозом „анемия“. Набор содержит показатели общего анализа крови, ретикулоциты, дополнительные биохимические маркеры обмена железа и воспалительного процесса. Разработана программа, позволяющая автоматизировать процесс анализа исследуемого набора согласно предложенному алгоритму обработки с учетом особенностей медицинских данных. Выполнена предварительная подготовка и очистка данных, проведен статистический и факторный анализ. Анализ выделенных групп данных позволил найти некоторые общие показатели для пациентов с анемическим синдромом. С помощью факторного анализа сокращено число переменных, выделены четыре основных фактора (группы исходных признаков), которые необходимы для описания исследуемых данных. Полученные результаты могут быть использованы для предоставления статических отчетов медицинской организации. Также исследуемые данные были подготовлены для возможности использования методов машинного обучения и более глубокого анализа с целью выявления наиболее эффективной диагностики анемии на ранних стадиях.

**Ключевые слова:** анализ медицинских данных, машинное обучение, медицинская статистика, клинические показатели, описательная статистика, факторный анализ

**Ссылка для цитирования:** Боженко В. В., Черныш Н. Ю., Татарникова Т. М. Интеллектуальный анализ данных в диагностике анемии по клиническим показателям // Изв. вузов. Приборостроение. 2024. Т. 67, № 4. С. 321—329. DOI: 10.17586/0021-3454-2024-67-4-321-329.

**DATA MINING IN THE DIAGNOSIS OF ANEMIA BY CLINICAL INDICATORS**V. V. Bozhenko<sup>1\*</sup>, N. Yu. Chernysh<sup>2</sup>, T. M. Tatarnikova<sup>1</sup>

<sup>1</sup>St. Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russia  
vibozenko@yandex.ru

<sup>2</sup>V. A. Almazov National Medical Research Center, St. Petersburg, Russia

**Abstract.** A set of medical data obtained from the information system of a network laboratory for outpatient observation, which contains test indicators of patients diagnosed with anemia, is studied. The set contains indicators of a general blood test, reticulocytes, additional biochemical markers of iron metabolism and the inflammatory process. A program is developed to automate the process of analyzing the test set according to the proposed processing algorithm, taking into account the medical data characteristics. Preliminary preparation and data cleaning are completed, statistical and factor analysis are carried out. Analysis of the selected groups of data makes it possible to find some common indicators for patients with anemic syndrome. Using factor analysis, the number of variables is reduced and four main factors (groups of initial characteristics) necessary to describe the data under study are identified. The results obtained can be used to provide static reports to a medical organization. Also, the studied data are prepared to allow the use of machine learning methods and deeper analysis in order to identify the most effective diagnosis of anemia in the early stages.

**Keywords:** medical data analysis, machine learning, medical statistics, clinical indicators, descriptive statistics, factor analysis

**For citation:** Bozhenko V. V., Chernysh N. Yu., Tatarnikova T. M. Data mining in the diagnosis of anemia by clinical indicators. *Journal of Instrument Engineering*. 2024. Vol. 67, N 4. P. 321—329 (in Russian). DOI: 10.17586/0021-3454-2024-67-4-321-329.

**Введение.** В современном мире ведущие медицинские организации оснащены средствами для сбора и хранения данных в специальных информационных системах. При этом большое количество данных, хранящихся в медицинских базах, требует специальных инструментов для обнаружения полезной информации и выявления важных зависимостей, необходимых для принятия обоснованных решений [1]. Использование искусственного интеллекта в медицине становится возможным благодаря анализу больших объемов данных, одним из источников которых являются лабораторные исследования [2, 3]. Это было бы невозможно без развития технологической базы лабораторной диагностики и применения соответствующих математических инструментов.

В связи со сказанным актуален поиск решений, позволяющих более эффективно использовать базу лабораторных данных для повышения качества медицинской помощи. Как известно, качество лабораторной диагностики подчиняется золотому правилу „правильный тест—правильному пациенту—в правильное время“. Оценка различных клиничко-лабораторных факторов облегчает поиск второго неизвестного в данной формуле — группы пациентов, для которых следует назначить то или иное исследование [4—6].

Общий анализ крови (ОАК) является одним из самых распространенных, широкодоступных и чувствительных методов оценки состояния пациента как по совокупности входящих в него показателей, так и по отдельно взятым параметрам. Важнейшей характеристикой является оценка эритроцитарного ростка цельной крови. Анемический синдром существенно осложняет течение и ухудшает прогноз любого заболевания за счет гипоксии. Выявление распространенности анемического синдрома в амбулаторной практике и его характеристики дают дополнительные основания для клиничко-лабораторной оценки пациентов.

Определение места и значения каждого отдельного параметра крови или совокупности этих параметров, а также объема исследований, входящих в общий анализ крови, позволит оптимизировать финансовые затраты на лабораторные исследования. Разработка алгоритмов позволит оптимизировать количество тестов, тем самым снижая стоимость обследования пациента.

Современная система здравоохранения вынуждена преодолевать проблемы высокой стоимости лабораторных тестов и невозможности проведения высокотехнологичных тестов во всех лабораториях [5]. Таким образом, поиск дополнительного сочетания маркеров в развернутом общем анализе крови и ассоциация с другими лабораторными и клиническими данными является актуальной задачей.

Использование алгоритмов, объединяющих результаты клинической оценки состояния пациента с результатами лабораторных исследований, позволяет компенсировать слабые стороны каждого метода при его изолированном использовании, а также улучшить аналитические характеристики всего алгоритма.

**Цели и задачи исследования.** Целью настоящего исследования является разработка программы персонализированного подхода к диагностике анемии на основе углубленного анализа маркеров гемограммы.

Задачи исследования:

- 1) методами интеллектуального анализа данных провести предварительную обработку больших объемов данных пациентов с диагнозом анемия, наблюдающихся в амбулаторном звене здравоохранения;
- 2) выполнить статистический анализ подготовленных данных;
- 3) подготовить набор данных для дальнейших исследований;
- 4) провести факторный анализ;
- 5) выявить ограничения в использовании данных лабораторных информационных систем в реальной клинической практике.

Для проведения подобных исследований необходимо обрабатывать данные с целью подготовки к математическому моделированию прогностических алгоритмов и применению методов машинного обучения.

**Описание исходных данных.** При анализе больших объемов данных в клинической и лабораторной практике используются результаты цифровизации и, в частности, возможности лабораторных информационных систем (ЛИС). ЛИС накапливает все данные о пациенте, которые служат возможными идентификационными признаками — от персональных данных до результатов лабораторных исследований, полученных напрямую с анализаторов. Такая совокупность сведений предоставляет пользователю — лечащему врачу, руководителю структурного подразделения, организатору здравоохранения — возможность в любой момент времени получить любую информацию, однако только по тем данным, которые были исходно обозначены как необходимые [6].

В клинко-диагностических лабораториях выполняется большой объем исследований [7, 8], при этом для обработки таких массивов данных с целью автоматизированного извлечения полезной информации целесообразно использовать информационные технологии и методы интеллектуального анализа.

В проводимое исследование по оценке данных большого объема включены несколько проектов для выявления связей между клинко-лабораторными показателями. В общей сложности проанализировано 142 000 результатов лабораторных исследований, часть из которых в последующем включена в обработку в соответствии с критериями: возраст старше 18 лет, наличие анемического синдрома, выполнение общего анализа крови на автоматическом 5diff-анализаторе с включением данных о ретикулоцитах, дополнительных биохимических маркерах обмена железа и воспалительного процесса.

В исследовании использовались данные, полученные из ЛИС сетевой лаборатории амбулаторного наблюдения. Была проведена выгрузка данных по общему анализу крови. В качестве дополнительных данных использованы результаты биохимического исследования для дифференциальной диагностики анемии. Таким образом, одновременно оценивались показатели обмена железа, маркеры воспаления, ретикулоцитарные показатели, необходимые для определения патогенеза анемии с целью выявления прогностически значимых сочетаний.

**Описание этапов исследования.** При разработке программ для интеллектуального анализа необходимо учитывать особенности медицинских данных. Обработка огромного количества данных сопряжена с большим количеством неточностей, несоответствий, что осложняет получение достоверных результатов. Необходимо грамотно обрабатывать возможные асимметричность, пропуски значений, выбросы, шумы в данных [9, 10].

Процесс обработки медицинских данных обычно включает несколько этапов:

- 1) импорт данных из ЛИС,
- 2) анализ предметной области, постановка целей и задач,
- 3) первичная обработка данных для устранения аномалий,
- 4) статистический анализ, который может включать разделение на группы, для оценки параметров различных групп пациентов,
- 5) применение методов интеллектуального анализа данных, включая факторный анализ, методы машинного обучения для кластеризации пациентов на группы,
- 6) сохранение результатов исследования.

Стоит отметить, что для получения корректных результатов исследования необходима совместная работа медицинского специалиста и аналитика, который использует программные инструменты и информационные технологии для обработки больших массивов данных. Именно такой тандем позволяет добиться наиболее эффективных результатов анализа.

С учетом результатов анализа предметной области была написана программа для автоматической обработки загруженных из системы данных, которая позволяет получить

первичную информацию, выявить аномалии, получить сводную информацию, обнаружить пропуски, провести статистический анализ, и является полезным инструментом извлечения информации из имеющихся данных. С помощью этой программы были проанализированы исходные данные пациентов.

Необходимо подчеркнуть, что программа разрабатывалась на основе выделенных этапов с учетом требований медицинского учреждения к результатам.

**Предварительная обработка данных.** Важность предварительной обработки очевидна, так как ввод данных в систему может сопровождаться ошибками ввода информации, которые затем повлекут за собой выдвижение неверных гипотез и выводов, что недопустимо в медицинских исследованиях [10].

В ходе предварительного анализа были удалены выбросы — данные, которые отличаются от общего распределения. Для этого использован метод интерквартильных расстояний, где выбросами считаются значения, превышающие 1,5 интерквартильного расстояния [11]. Такое удаление выбросов привело к лучшему соответствию набора данных требованию нормальности, что позволило применить статистический анализ.

На рис. 1 представлено распределение параметра МСН (среднее содержание гемоглобина в эритроците) до (а) и после (б) удаления выбросов ( $N$  — количество наблюдений, попадающих в определенные интервалы значений параметра МСН. Данные разбиваются на интервалы, которые соответствуют основаниям столбцов гистограммы).

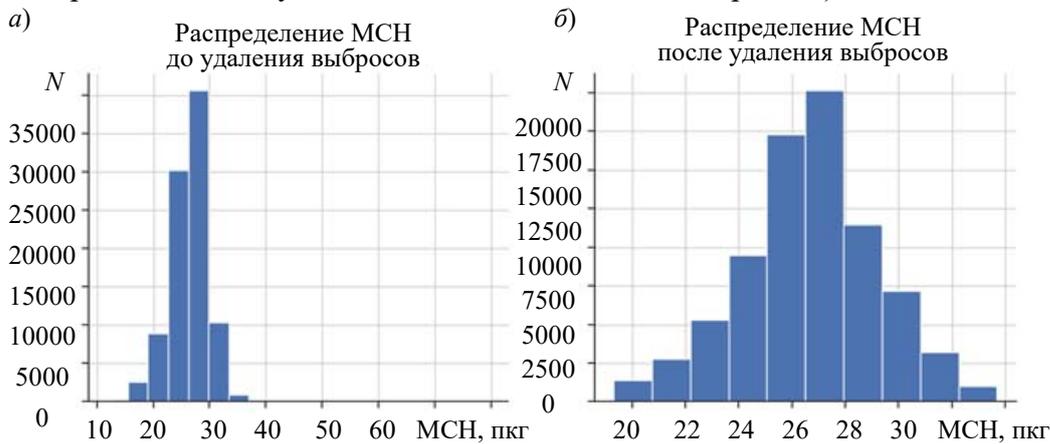


Рис. 1

Также в ходе предварительного анализа было обнаружено, что в наборе данных полностью заполнены только показатели общего анализа крови, пол, возраст и код пациента. Более специфичные показатели (ретикулоциты, железообмен, ферритин и т.д. — эти показатели могут быть важны для принятия решений врачом) указаны только для 1 % пациентов. Заполнение этих пропусков средним или медианным значением может исказить итоговые результаты. Представляет интерес заполнение таких пропусков с помощью методов машинного обучения, однако классические методы предсказания числовых значений (линейная регрессия, регрессия с помощью дерева решений, случайного леса) не позволяют достигнуть нужной точности, в связи с чем целесообразно применять методы глубокого обучения [12, 13]. В настоящей работе пропущенные значения не заполнялись, а анализ специализированных показателей проводился по тому количеству данных, что имелись в системе.

**Статистический анализ.** Для анализа количественных данных ОАК были построены диаграммы распределения, чтобы оценить нормальность распределения величин и выделить основные диапазоны значений показателей. Также были построены графики типа boxplot, которые позволяют наглядно описать характеристики величин и являются незаменимым инструментом в описательной статистике [14].

Стоит отметить, что программа позволяет автоматизировать процессы построения графиков, выводить описательные статистики по каждому показателю исходной таблицы, а также выделять некоторые особенности, например, предупреждать об имеющихся выбросах или пропусках в данных, упрощая исследование для конечного пользователя. Однако медицинскому специалисту необходимо оценивать полученные результаты, которые в дальнейшем могут быть применены в ежегодных статистических отчетах.

В ходе работы была построена специальная система фильтров, которая позволяет выделить различные группы пациентов из исходного набора и проанализировать результаты для каждой из этих групп. При обновлении выгрузки и добавлении новых данных с помощью разработанной программы можно легко проводить анализ новых пациентов, разделенных по заранее определенным группам, и выгружать необходимые результаты.

Первоначальная выгрузка из ЛИС сетевой лаборатории составила 142 000 гематологических исследований с результатами, демонстрирующими уровень гемоглобина (Hb) ниже референсного интервала — менее 120 г/л для женщин и менее 130 г/л — для мужчин — в выборке ОАК (рис. 2). Красными линиями показаны нижние границы референсного интервала. В целом по совокупности измерений уровень гемоглобина равен 109,7 г/л (103,0—119,0 г/л).

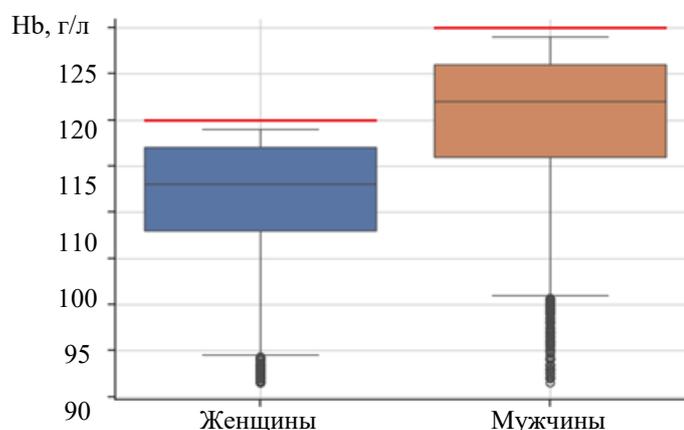


Рис. 2

На рис. 3 представлено распределение значения среднего объема эритроцитов (MCV) у пациентов с анемией. Референсный интервал (красные линии) — 80—95 фл. На рис. 4 показано распределение значения среднего содержания гемоглобина в эритроцитах у пациентов с анемией в выборке ОАК сетевой лаборатории. Референсный интервал (красные линии) — 27—32 пкг.

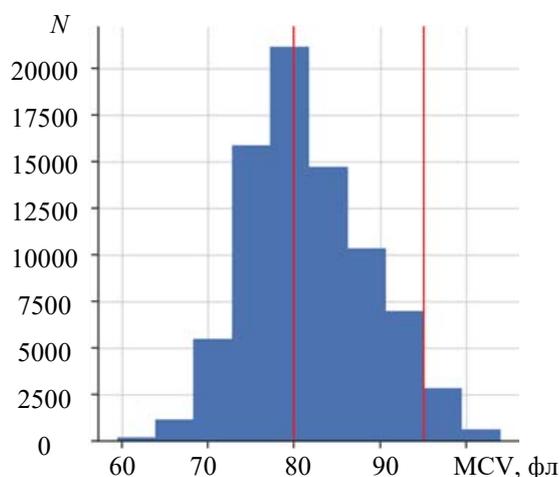


Рис. 3

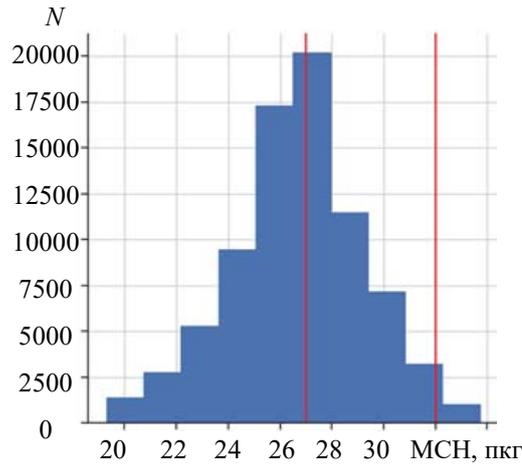


Рис. 4

Из представленных графиков видно, что более половины результатов измерения MCV и MCH находятся в пределах референсных интервалов или смещены в сторону снижения показателей, это в сочетании со снижением гемоглобина может указывать на множество анемических состояний, включая наиболее распространенные железодефицитную анемию (ЖДА) и анемию хронических заболеваний (АХЗ), что отмечается у 75 % пациентов, имеющих MCH ниже 28,5 и MCV ниже 86.

Данные были разделены на две группы: группа, где заполнены все параметры (столбцы таблицы), в которую входят пациенты, сдавшие дополнительные анализы, и группа, где заполнены только общие показатели крови. В ходе проверки выявлено, что для пациентов первой группы средние значения показателей MCH, MCHC, гемоглобина, эритроцитов ниже. С целью проверки был проведен дисперсионный анализ (ANOVA) [14] для сравнения средних значений выборок, уровень значимости был выбран 0,05.

**Методы интеллектуального анализа данных.** В зависимости от поставленных задач в медицине применяются методы прогнозирования течения болезни или действия препаратов, анализа выживаемости, классификации и выявления диагноза, кластеризации пациентов на различные группы, а также описательные методы для нахождения скрытых зависимостей между показателями и выявления влияния различных факторов на болезнь [15].

Методы преобразования данных включают методы уменьшения размерности для извлечения важных признаков, так как иногда нет необходимости использовать все имеющиеся признаки. Сжатие данных часто необходимо, чтобы достичь более простых моделей и сократить время выполнения алгоритмов машинного обучения, при этом не потеряв точность результата [16, 17].

В работе также был проведен факторный анализ, который позволяет выделить группы взаимосвязанных признаков, которые описывают набор данных. Оценка сравнительной значимости главных компонент производится по их собственным значениям, которые представляют собой дисперсию стандартизированных исходных данных по этой компоненте, т.е. „объясняемую“ ей часть общей дисперсии объектов. Эти значения обычно сопровождаются соответствующими им процентами полной дисперсии и процентами накопленной дисперсии.

На рис. 5, а представлен график „каменистой осыпи“ (scree plot), который позволяет определить число значимых факторов  $f$ , необходимых в исследовательском факторном анализе (т.е. количество главных компонент). График представляет распределение собственных значений (собственное значение показывает, какая доля общей изменчивости приходится на компоненту) факторов ( $E$ , Eigenvalue) в порядке убывания их величины. На рис. 5, б показан график объясненной дисперсии (синяя кривая — доля дисперсии, зеленая —

суммарная „объясняемая“ дисперсия) — видно, что наибольшая доля  $p$  объясненной дисперсии приходится на четыре фактора.

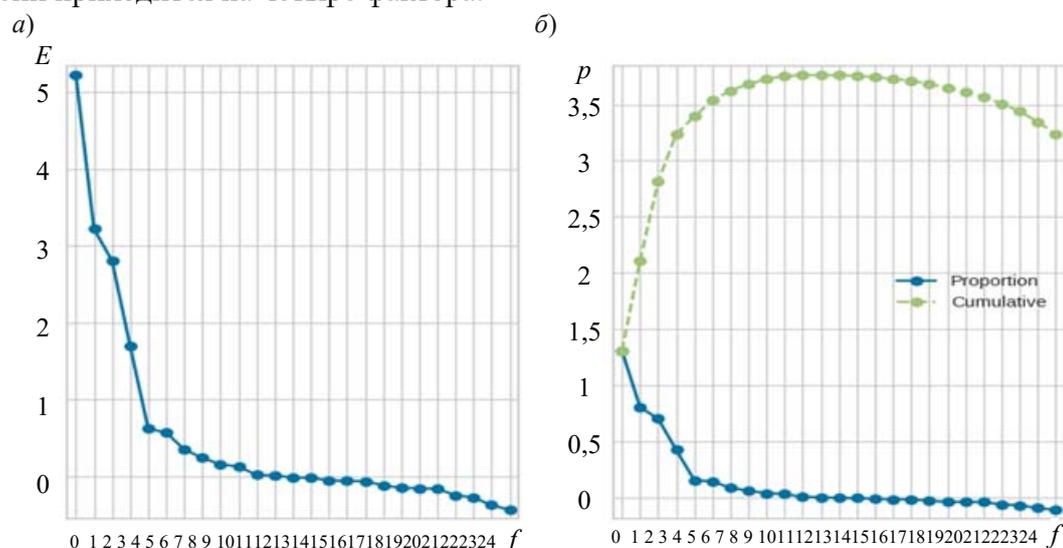


Рис. 5

В результате были выделены четыре группы признаков, которые объединяют некоторые параметры. Таким образом, можно строить модель кластеризации, основанную на факторах, полученных с помощью метода главных компонент, и необязательно использовать все показатели.

**Заключение.** Таким образом, первичный анализ показывает, что есть ряд ограничений на использование полного набора данных пациентов с диагнозом „анемия“ в связи с пропусками данных. Кроме того, использование методов интеллектуального анализа ограничивает отсутствие:

1) информации об основном заболевании или причинах выполнения исследования ОАК;

2) сведений о проводимой терапии;

3) дополнительных лабораторных исследований, необходимых для идентификации и дифференциальной диагностики причин анемии — не определены ферритин, трансферрин, железо сыворотки, насыщение трансферрина железом, общая железосвязывающая способность сыворотки, ретикулоциты. Наиболее критичным в дифференциальной диагностике анемии хронических заболеваний и железодефицитной анемии являлось отсутствие данных о ферритине.

Возможности автоматизированного гематологического анализа позволяют искать новые клинически значимые маркеры для дифференциальной диагностики анемии. Исследование показало, что традиционный морфологический анализ с проведением ручной микроскопии должен быть расширен расчетными показателями автоматизированной гемограммы. При этом полученные результаты показали, что объем обследований, выполняемых у амбулаторных пациентов, недостаточен для определения характера анемии. Определены ограничения в использовании больших лабораторных данных, накапливающихся в реальной клинической практике.

Следует отметить, что интеллектуальный анализ и машинное обучение активно применяются в медицине, но по большей степени — при работе с изображениями или в размеченных наборах данных для решения задачи классификации с целью постановки диагноза. Наборы данных, позволяющие диагностировать анемию, отсутствуют, поэтому в настоящей работе был подготовлен набор реальных данных для специалистов, которые смогут применять методы интеллектуального анализа и машинного обучения в диагностике анемии.

Разработанная программа позволяет автоматизировать процесс получения статистических данных, помогает принимать решения, осуществлять поиск структурированных данных по запросу, находить скрытые логические и статистические закономерности, а также может использоваться для развернутой оценки изменений, характерных для анемических состояний.

## СПИСОК ЛИТЕРАТУРЫ

1. Huang M.-J., Sung H.-S., Hsieh T.-J., Wu M.-C., Chung S.-H. Applying data-mining techniques for discovering association rules // *Soft Computing – A Fusion of Foundations, Methodologies and Applications*. 2020. Vol. 24. P. 8069—8075.
2. Цветкова Л. А., Черченко О. В. Технология больших данных в медицине и здравоохранении России и мира // *Врач и информационные технологии*. 2016. № 3. С. 60—73.
3. Карнаухов Н. С., Ильюхин Р. Г. Возможности технологий „Big Data“ в медицине // *Врач и информационные технологии*. 2019. № 1. С. 59—63.
4. Explore the Top 10 Cardiac Diagnosis Trends in 2023 [Электронный ресурс]: <<https://www.startus-insights.com/innovators-guide/cardiac-diagnosis-trends/>>. (дата обращения 15.01.2024)
5. Chang A., Cowling K. Past, present, and future of global health financing: A review of development assistance, government, out-of-pocket, and other private spending on health for 195 countries, 1995-2050 // *Lancet*. 2019. Vol. 393, N 10187. P. 2233—2260.
6. Овчинникова М. А., Жиленкова Ю. И., Черныш Н. Ю. Применение big data в лабораторной медицине // *Российский журнал персонализированной медицины*. 2023. Т. 3, № 4. С. 77—87.
7. Wang L. Big data analytics in medical engineering and healthcare: methods, advances and challenges // *J. of Medical Engineering & Technology*. 2020. Vol. 44, N 6. P. 267—283.
8. Razzak M. I., Imran M., Xu G. Big data analytics for preventive medicine // *Neural Computing & Applications*. 2020. Vol. 32. P. 4417—4451.
9. Bozhenko V. V. Data preprocessing in machine learning // *Bulletin of the UNESCO department „Distance education in engineering“ of the SUAI: Collection of the papers*. St. Petersburg, 2023. P. 52—55.
10. Bozhenko V. V., Tatarnikova T. M. Application of Data Preprocessing in Medical Research // *Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*. St. Petersburg, 2023. P. 1—4.
11. Баврина А. П. Современные правила использования методов описательной статистики в медико-биологических исследованиях // *Медицинский альманах*. 2020. С. 95—104.
12. Makridakis S., Spiliotis E., Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward // *PloS one*. 2018. Vol. 13. P. 1—20.
13. Попова И. А. Исследование алгоритмов машинного обучения для предварительной обработки данных в задачах регрессии // *Искусственный интеллект в автоматизированных системах управления и обработки данных: Сб. ст. Всерос. научн. конф. Т. 1. М., 2022. С. 237—242.*
14. Гудинова Ж. В., Демакова Л. В. Технологии статистического анализа медицинских данных: первичный анализ данных, сравнение групп // *Фундаментальная и клиническая медицина*. 2023. Т. 8, № 1. С. 119—131.
15. Комков А. А., Мазаев В. П., Рязанова С. В., Самочатов Д. Н., Базаева Е. В. Основные направления развития искусственного интеллекта в медицине // *Научное обозрение. Медицинские науки*. 2020. № 5. С. 33—40.
16. Taherdoost H., Sahibuddin S., Jalaliyoon N. Exploratory Factor Analysis; Concepts and Theory // *Advances in applied and pure mathematics*. 2022. Vol. 27. P. 375—382.
17. Покидышева Л. И., Мадаминова М. О., Ладнюк В. В. Исследование и разработка алгоритмов и программных средств сокращения размерности многомерных данных // *Международ. науч.-исслед. журн.* 2017. № 5–3(59). С. 94—97.

**Сведения об авторах****Виктория Вячеславовна Боженко**

— Санкт-Петербургский государственный университет аэрокосмического приборостроения, кафедра прикладной информатики; старший преподаватель; E-mail: vibozenko@yandex.ru

**Наталья Юрьевна Черныш**

— канд. мед. наук; Национальный медицинский исследовательский центр им. В. А. Алмазова, кафедра лабораторной медицины с клиникой; доцент; E-mail: nycher@mail.ru

**Татьяна Михайловна Татарникова**

— д-р техн. наук, профессор; Санкт-Петербургский государственный университет аэрокосмического приборостроения, кафедра прикладной информатики; профессор; E-mail: tm-tatarn@yandex.ru

Поступила в редакцию 04.12.23; одобрена после рецензирования 07.12.23; принята к публикации 08.02.24.

## REFERENCES

1. Huang M.-J., Sung H.-S., Hsieh T.-J., Wu M.-C., Chung S.-H. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 2020, vol. 24, pp. 8069–8075.
2. Tsvetkova L.A., Cherchenko O.V. *Medical Doctor and IT*, 2016, no. 3, pp. 60–73. (in Russ.)
3. Karnaukhov N.S., Ilyukhin R.G. *Medical Doctor and IT*, 2019, no. 1, pp. 59–63. (in Russ.)
4. *Explore the Top 10 Cardiac Diagnosis Trends in 2023*, <https://www.startus-insights.com/innovators-guide/cardiac-diagnosis-trends/>.
5. Chang A., Cowling K. *Lancet*, 2019, no. 10187(393), pp. 2233–2260.
6. Ovchinnikova M.A., Zhilenkova Yu.I., Chernysh N.Yu. *Russian Journal for Personalized Medicine*, 2023, no. 4(3), pp. 77–87. (in Russ.)
7. Wang L. *Journal of Medical Engineering & Technology*, 2020, no. 6(44), pp. 267–283.
8. Razzak M.I., Imran M., Xu G. *Neural Computing & Applications*, 2020, vol. 32, pp. 4417–4451.
9. Bozhenko V.V. *Bulletin of the UNESCO department „Distance education in engineering“ of the SUAI*, Collection of the papers. St. Petersburg, 2023, pp. 52–55.
10. Bozhenko V.V., Tatarnikova T.M. *Wave Electronics and its Application in Information and Telecommunication Systems (WECINF)*, St. Petersburg, 2023, pp. 1–4.
11. Bavrina A.P. *Meditsinskiy al'manakh*, 2020, pp. 95–104. (in Russ.)
12. Makridakis S., Spiliotis E., Assimakopoulos V. *PloS one*, 2018, vol. 13, pp. 1–20.
13. Popova I.A. *Iskusstvennyy intellekt v avtomatizirovannykh sistemakh upravleniya i obrabotki dannykh* (Artificial Intelligence in Management, Control, and Data Processing Systems), Collection of articles of the All-Russian Scientific Conference, vol. 1, Moscow, 2022, pp. 237–242. (in Russ.)
14. Gudina Zh.V., Demakova L.V. *Fundamental and Clinical Medicine*, 2023, no. 1(8), pp. 119–131. (in Russ.)
15. Komkov A.A., Mazaev V.P., Ryazanova S.V., Samochatov D.N., Bazaeva E.V. *Science review. Medical Sciences*, 2020, no. 5, pp. 33–40. (in Russ.)
16. Taherdoost H., Sahibuddin S., Jalaliyoon N. *Advances in applied and pure mathematics*, 2022, vol. 27, pp. 375–382.
17. Pokidysheva L.I., Madaminova M.O., Ladnyuk V.V. *International Research Journal*, 2017, no. 5-3(59), pp. 94–97. (in Russ.)

## Data on authors

**Viktoriya V. Bozhenko**

— St. Petersburg State University of Aerospace Instrumentation, Department of Applied Informatics; Senior Lecturer; E-mail: vibozhenko@yandex.ru

**Natalia Yu. Chernysh**

— PhD; V. A. Almazov National Medical Research Center, Department of Laboratory Medicine with Clinic; Associate Professor; E-mail: nycher@mail.ru

**Tatiana M. Tatarnikova**

— Dr. Sci., Professor; St. Petersburg State University of Aerospace Instrumentation, Department of Applied Informatics; Professor; E-mail: tm-tatarn@yandex.ru

Received 04.12.23; approved after reviewing 08.12.23; accepted for publication 08.02.24.