

<https://doi.org/10.29001/2073-8552-2025-40-1-199-208>
УДК 616.9-036.65-07:004.85

Определение предикторов неблагоприятного исхода в подострый период инфекции SARS-CoV-2 с помощью методов машинного обучения

И.В. Долгалёв¹, Д.А. Вражнов¹, И.В. Толмачев^{1, 2}, Е.Г. Старикова¹,
И.С. Каверина¹, М.В. Завьялова¹

¹ Сибирский государственный медицинский университет Министерства здравоохранения Российской Федерации (СибГМУ Минздрава России),
634050, Российская Федерация, Томск, Московский тракт, 2

² Центральный научно-исследовательский институт организации и информатизации здравоохранения (ЦНИИОИЗ),
127254, Российская Федерация, Москва, ул. Вучетича, 12

Аннотация

Введение. Патологические изменения систем и органов после перенесенного COVID-19 могут приводить к отложенному смертельному исходу. При этом одной из особенно значимых систем-мишеней постковидных изменений является кардиоваскулярная система.

Цель исследования: выявление с помощью методов машинного обучения (МО) показателей, имеющих прогностическую ценность при определении неблагоприятного исхода подострого COVID-19.

Материал и методы. В исследование были включены 212 человек, госпитализированных после перенесенного ранее в тяжелой форме COVID-19. Ретроспективно пациенты были разделены на 2 группы: 140 пациентов, которые были выписаны из стационара с заключением об улучшении состояния, и 72 пациента, умершие в период госпитализации. Всем пациентам проводились общеклинический, биохимический анализы, оценка свертывающей системы крови. Для анализа данных были использованы следующие методы МО: метод опорных векторов, случайный лес, стохастический градиентный бустинг. Валидация полученных моделей производилась методом перекрестной 10-кратной проверки совместно с ROC-AUC анализом (Receiver Operation Characteristics – Area Under Curve).

Результаты. В созданных нами предиктивных моделях предикторами смертельного исхода для методов случайный лес и стохастический градиентный бустинг являлись мочевины и температура тела; для машины опорных векторов – количество эритроцитов, эозинофилов и моноцитов, международное нормализованное отношение (МНО).

Выводы. В проведенном исследовании две предиктивные модели, созданные с помощью методов МО, случайный лес и стохастический градиентный бустинг, показали, что прогностическое значение имеют изменения двух показателей: уровня мочевины и температуры тела. Метод опорных векторов выявил другие предикторы, а именно количество эритроцитов, эозинофилов и моноцитов, МНО. Нами был применен метод голосования, на основе которого в качестве информативных признаков были установлены уровень мочевины и температура тела. Методы МО случайный лес и стохастический градиентный бустинг продемонстрировали схожие результаты, мы не учитывали данные, полученные с помощью метода опорных векторов. Подобный подход выбора предиктивной модели голосованием часто используется при оценке данных методами искусственного интеллекта. Возможно, повышение уровня мочевины являлось пусковым механизмом, ведущим к эндотелииту и последующему инфаркту миокарда, до того, как развилась острая почечная недостаточность.

Ключевые слова:	COVID-19; постковидные изменения; машинное обучение; мочевины; температура тела; количество эритроцитов; количество эозинофилов; количество моноцитов; международное нормализованное отношение; сердечно-сосудистая система.
Финансирование:	исследование проводилось без финансовой поддержки со стороны грантов, общественных, некоммерческих, коммерческих организаций и структур.
Соответствие принципам этики:	ретроспективное исследование включало данные рутинных и лабораторных исследований. Все данные были деперсонализированы перед включением в исследование.
Для цитирования:	Долгалёв И.В., Вражнов Д.А., Толмачев И.В., Старикова Е.Г., Каверина И.С., Завьялова М.В. Определение предикторов неблагоприятного исхода в подострый период инфекции SARS-CoV-2 с помощью методов машинного обучения. <i>Сибирский журнал клинической и экспериментальной медицины</i> . 2025;40(1):199–208. https://doi.org/10.29001/2073-8552-2025-40-1-199-208 .

Старикова Елена Григорьевна, e-mail: elena.g.starikova@yandex.ru.

Determination of predictors of an unfavorable outcome in the subacute period of SARS-CoV-2 infection using machine learning methods

Igor V. Dolgalev¹, Denis A. Vrazhnov¹, Ivan V. Tolmachev^{1, 2},
Elena G. Starikova¹, Irina S. Kaverina¹, Marina V. Zavyalova¹

¹ Siberian State Medical University of the Ministry of Health of the Russian Federation (SSMU),

2, Moskovskiy tract str., Tomsk, 634050, Russian Federation

² Russian Research Institute of Health,

11, Dobrolubova, Moscow, 127254, Russian Federation

Abstract

Background. Pathological changes in systems and organs after COVID-19 can lead to delayed death. One of the most influenced target systems of post-COVID changes is the cardiovascular system.

Aim: To identify, using machine learning methods, indicators that have predictive value in determining the adverse outcome of subacute COVID-19.

Material and Methods. The study included 212 people admitted after previous severe COVID-19. Retrospectively, the patients were divided into 2 groups: 140 patients discharged from the hospital after improvement in their state and 72 patients died during hospitalization. All patients underwent general clinical, biochemical analyses, assessment of blood coagulation system. The following machine learning methods were used for data analysis: support vector machine, random forest, stochastic gradient boosting. Validation of the obtained models was carried out by the method of 10-fold cross-validation in conjunction with ROC–AUC analysis (Receiver Operation Characteristics – Area Under Curve).

Results. In the created models, the predictors of mortality were urea and body temperature for the random forest and stochastic gradient boosting methods, erythrocyte, eosinophil and monocyte counts, and INR (International Normalized Ratio) level for the support vector machine.

Conclusion. In our study, two predictive models created using machine learning methods random forest and stochastic gradient boosting showed that changes in urea level and body temperature had predictive value. The support vector machine revealed other predictors, namely the number of erythrocytes, eosinophils and monocytes, INR. We used the voting method, on the basis of which the urea level and body temperature were established as informative signs. The random forest and stochastic gradient boosting methods showed similar results, we did not take into account the data obtained using the support vector machine. This approach of choosing a predictive model by voting is often used when evaluating data using artificial intelligence methods. It is possible that an increase in urea levels was a trigger leading to endotheliitis and subsequent myocardial infarction, before acute renal failure developed.

Keywords:	COVID-19; post-COVID changes; machine learning; urea; body temperature; erythrocyte count; eosinophil count; monocyte count; International Normalized Ratio; cardiovascular system.
Funding:	the study was carried out without financial support from grants, public, non-profit, commercial organizations and structures.
Compliance with ethical standards:	the study includes retrospective data of routine clinical and laboratory examinations. All data were anonymized prior to the use in the study.
For citation:	Dolgalev I.V., Vrazhnov D.A., Tolmachev I.V., Starikova E.G., Kaverina I.S., Zavyalova M.V. Determination of predictors of an unfavorable outcome in the subacute period of SARS-CoV-2 infection using machine learning methods. Siberian Journal of Clinical and Experimental Medicine. 2025;40(1):199-208. https://doi.org/10.29001/2073-8552-2025-40-1-199-208 .

Введение

В настоящее время известно, что выздоровление после COVID-19 (купирование острых симптомов, отрицательный ПЦР-тест) не является гарантией полного восстановления организма. Патологические изменения систем и органов значительно ухудшают здоровье пациентов и могут приводить к отложенному смертельному исходу. Всемирная организация здравоохранения

Introduction

Currently recovery after COVID-19 (relief of acute symptoms, negative PCR test) is not a guarantee of complete recovery of the body. Pathological changes in systems and organs significantly worsen the health of patients and can lead to delayed death. The World Health Organization has defined the term post-COVID-19 to refer to cases of persistent symptoms after 3 months of SARS-CoV-2 infection

определила термин «пост-COVID-19» для обозначения случаев сохранения симптомов после 3 мес. инфекции SARS-CoV-2, длящихся не менее 2 мес. и не объясняемых каким-либо другим заболеванием¹.

В литературе существует более подробная классификация состояний после перенесенной инфекции SARS-CoV-2. Так, пост-острый COVID-19 определяется как стойкие симптомы и/или отсроченные или долгосрочные осложнения инфекции SARS-CoV-2 в течение 4 нед. после появления симптомов [1]. Далее исследователи делят заболевание на две категории: (1) подострый или продолжающийся симптоматический COVID-19, который включает симптомы и аномалии, присутствующие в течение 4–12 нед. после острого течения COVID-19; (2) хронический или пост-COVID-19 синдром, который включает симптомы и аномалии, сохраняющиеся или присутствующие после 12 нед. от начала острого COVID-19 и не связанные с альтернативными диагнозами [2].

Постковидные изменения могут затронуть различные органы и системы организма человека, но одной из особенно значимых систем-мишеней является кардиоваскулярная система [3, 4].

Вовлечение сердечно-сосудистой системы, подтвержденное методом магнитно-резонансной томографии сердца, наблюдалось у 78% лиц, выздоровевших от острой инфекции COVID-19, независимо от ранее существовавших состояний, тяжести и течения начальных проявлений инфекции или наличия ранее сердечных симптомов [5]. Изменения в сердечно-сосудистой системе после перенесенного COVID-19 включают миокардит, аритмии, перикардит, тромбоэмболии, инфаркт миокарда, внезапную смерть, сердечную недостаточность [6, 7]. При этом даже инфекция SARS-CoV2 средней степени тяжести, не требующая госпитализации, приводила к патологическим изменениям не только в сердечно-сосудистой системе, но и в других органах [8].

В настоящее время в литературе имеется достаточное количество данных, позволяющих строить прогностические модели (ПМ), способные определить, в какой форме будет протекать заболевание у пациента и, следовательно, помочь с дальнейшей маршрутизацией [9]. При этом нет исследований, которые работают на этапе, когда острые симптомы уже купированы, но полученные организмом при заболевании повреждения могут спровоцировать смертельный исход.

Методы машинного обучения (МО) активно применяются для построения ПМ на основе медицинских данных. Это обусловлено рядом факторов. Во-первых, классические статистические методы опираются на строгие математические предположения, например, о виде распределения признака в популяции, пропорциональности рисков, что часто нарушается в медицинских данных. Во-вторых, классические методы математической статистики плохо работают на неоднородных данных (информация представлена категориальными и количественными переменными), которые сильно коррелируют между собой. Недостатком подобных подходов является слабая интерпретируемость получаемых моделей, что является критическим аспектом в области обработки медицинских данных. Интерпретируемость моделей МО достигается

lasting at least 2 months and not explained by any other disease¹. There is a more detailed classification of conditions after a SARS-CoV infection in the literature. Thus, post-acute COVID-19 is defined as persistent symptoms and/or delayed or long-term complications of SARS-CoV-2 infection within 4 weeks after the onset of symptoms [1]. The researchers further divide the disease into two categories: (1) subacute or ongoing symptomatic COVID-19, which includes symptoms and abnormalities, present for 4–12 weeks after the acute course of COVID-19; and (2) chronic or post-COVID-19 syndrome, which includes symptoms and abnormalities that persist or are present after 12 weeks from the onset of acute COVID-19 and are not associated with alternative diagnoses [2].

Post-COVID changes can affect various organs and systems of the human body, but one of the most affected target systems is the cardiovascular system [3, 4].

Involvement of the cardiovascular system, confirmed by magnetic resonance imaging (MRI) of the heart, was observed in 78% of people who recovered from acute COVID-19 infection, regardless of pre-existing conditions, severity and course of initial manifestations of infection or the presence of previous cardiac symptoms [5]. Changes in the cardiovascular system after COVID-19 include myocarditis, arrhythmias, pericarditis, thromboembolism, myocardial infarction, sudden death, and heart failure [6, 7]. At the same time, even a moderate SARS-CoV-2 infection that did not require hospitalization led to pathological changes not only in the cardiovascular system, but also in other organs [8].

Currently, there is a sufficient amount of data in the literature that allows us to build prognostic models (PM) that can determine in what form the disease will progress and, therefore, help with further routing [9]. At the same time, there are no studies that describe the situation when acute symptoms have already been resolved, but the damage received by the body during the disease can provoke a fatal outcome.

Machine learning (ML) methods are actively used to build predictive models based on medical data. This is due to a number of factors. Classical statistical methods rely on strict mathematical assumptions, for example, the type of distribution of a trait in a population, the proportionality of risks, which is often violated in medical data. Classical methods of mathematical statistics do not work well on heterogeneous data (information is represented by categorical and quantitative variables), which strongly correlate with each other. The disadvantage of such approaches is the weak interpretability of the resulting models, which is a critical aspect in the field of medical data processing. Interpretability of ML models is achieved by finding informative predictors. F1-score metrics, specificity/sensitivity/accuracy, ROC-AUC analysis are widely used to assess the quality of models [10].

The purpose of this study was to identify, using ML methods, indicators with predictive value for determining the adverse outcome of subacute COVID-19. In our retrospective study, we analyzed routine clinical indicators obtained from patients during hospitalization. Then predictors of death were identified using ML.

¹ World Health Organization. A clinical case definition of post COVID-19 condition by a Delphi consensus. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1. (10.02.2025).

путем нахождения информативных предикторов. Для оценки качества моделей широко используются метрики F1-score, специфичность / чувствительность / точность, ROC-AUC анализ [10].

Цель настоящего исследования: выявление показателей, имеющих прогностическую ценность определения неблагоприятного исхода подострого COVID-19. В проведенном нами ретроспективном исследовании были проанализированы рутинные клинические показатели, полученные у пациентов при госпитализации. Далее с помощью МО выявлены лабораторные показатели, имеющие предиктивную ценность для определения смертельного исхода.

Материал и методы

В исследование были включены 212 человек, госпитализированных в клиники СибГМУ после перенесенного ранее в тяжелой форме COVID-19, согласно маршрутизации, утвержденной Департаментом здравоохранения Томской области. На момент госпитализации все пациенты имели отрицательный COVID-19 ПЦР тест.

Критерии включения: 1) подтвержденный ПЦР COVID-19 в течение 3 мес. до госпитализации; 2) рентгенологические данные, свидетельствующие о перенесенной пневмонии; 3) наличие симптомов, требующих наблюдения в стационаре; 4) возраст старше 50 лет.

Критерии исключения: 1) возраст менее 50 лет.

Ретроспективно пациенты были разделены на 2 группы: 140 пациентов, которые были выписаны из стационара с заключением об улучшении состояния, и 72 пациента, умершие в период госпитализации. Средний возраст пациентов составил 71 год.

Клинические обследования проводили согласно стандартам оказания специализированной медицинской помощи в условиях стационара в соответствии с установленным диагнозом. Всем больным не менее двух раз в сутки выполняли термометрию, измеряли уровни систолического (САД) и диастолического артериального давления (ДАД), сатурацию крови, частоту дыхательных движений (ЧДД), частоту сердечных сокращений (ЧСС) и пульса.

Общий анализ крови включал определение числа эритроцитов, тромбоцитов, лейкоцитов, исследовали лейкоцитарную формулу, скорость оседания эритроцитов (СОЭ), а также показатели гемоглобина и гематокрита. Анализ выполняли на приборе Sysmex XN-1000 (Япония). В биохимическое исследование крови, которое проводили на биохимическом анализаторе Architect c4000 (США, Япония), входило определение содержания мочевины, креатинина, С-реактивного белка (СРБ), аспартатаминотрансферазы (АСТ), аланинаминотрансферазы (АЛТ).

Анализ системы свертывания крови включал определение фибриногена, активированное частичное тромбопластиновое время (АЧТВ) и международное нормализованное отношение (МНО) (Sysmex CS2000i, Япония).

Венозную кровь забирали утром натощак из локтевой вены в количестве 5 мл, стабилизировали КЗЭДТА.

Методы МО являются мощным инструментом анализа медицинских данных с целью построения прогностических моделей. Последовательность алгоритмов, состоящая из предварительной обработки, уменьшение размерности, создание ПМ, ее валидация, выделение информативных признаков образует конвейер МО. Обяза-

Material and Methods

The study included 212 people admitted to SSMU clinics after previous severe COVID-19.

Inclusion criteria: 1) confirmed PCR of COVID-19 within 3 months prior to hospitalization; 2) X-ray data indicating pneumonia; 3) the presence of symptoms requiring hospital monitoring; 4) age over 50 years.

Exclusion criterion: 1) the age less than 50 years old.

Retrospectively, the patients were divided into 2 groups: 140 patients discharged from the hospital with an improvement in their state and 72 patients died during hospitalization. The average age of the patients was 71 years.

Clinical examinations were conducted in accordance with the standards of specialized medical care in a hospital setting and corresponded to the established diagnoses. Thermometry was performed at least twice a day, the levels of systolic and diastolic blood pressure (SBP, DBP), blood saturation, respiratory rate (BR), heart rate (HR) and pulse were measured.

The blood test included the determination of the number of erythrocytes, platelets, leukocytes, the leukocyte formula, erythrocyte sedimentation rate (ESR) was studied, as well as hemoglobin and hematocrit indicators. The analysis was performed on the Sysmex XN-1000 device (Japan). The biochemical blood test included the determination of urea, creatinine, C-reactive protein (CRP), Aspartate aminotransferase (AST), Alanine aminotransferase (ALT) and was performed on an Architect c4000 biochemical analyzer (USA, Japan).

The analysis of the blood coagulation system included the determination of fibrinogen, activated Partial Prothrombin Time (APTT) and INR (Sysmex CS2000i) (Japan).

Venous blood was taken in the morning on an empty stomach from the ulnar vein in an amount of 5 ml, stabilized with K3EDTA.

ML methods are a powerful tool for analyzing medical data in order to build PM. A sequence of algorithms consisting of preprocessing, dimensionality reduction, creation of a PM, its validation, and the allocation of informative features forms the ML pipeline. The mandatory elements of the pipeline are the creation of a PM and its validation, the remaining steps are optional. ML methods generate PM based on examples marked up by experts, which form a training sample. The quality control of the trained model is carried out on a set of examples that did not participate in the training – a test sample.

In this work, the method of filling in gaps and filtering incomplete data is used as a preliminary data processing. A wide range of methods for filling in gaps is provided in the literature, but the choice of the best one depends on the input data. Given the relatively small size of the data set under study, a simple method of filling in gaps was proposed: if there are few gaps (less than 30 percent of the total), zero values are filled in, otherwise the data about the criteria is removed from consideration.

Reduction of input data dimension was not used, because these methods transform the original predictor space and complicate interpretability.

The creation of PM was implemented by such ML methods as the support vector machine, random forest, stochastic gradient boosting. The choice of these ML methods is due to the need to solve the problem of identifying informative predictors (signs). All of the above algorithms have the "built-in" ability to determine the degree of importance

тельными элементами конвейера являются создание ПМ и ее валидация, остальные шаги опциональны. Методы МО генерируют ПМ на основе размеченных экспертами примеров, совокупность которых образует обучающую выборку. Проверку качества обученной модели осуществляли на наборе примеров, не участвовавших в обучении, – на тестовой выборке.

В представленной работе, в качестве предварительной обработки данных использовали метод заполнения пропусков и фильтрация неполных данных. В литературе приводится широкий набор методов заполнения пропусков, однако выбор наилучшего зависит от входных данных. Учитывая относительно небольшой размер исследуемого набора данных, был предложен простой метод заполнения пропусков: если пропусков мало (менее 30% от общего числа), используется заполнение нулевыми значениями, в противном случае данные об исследуемом признаке удаляются из рассмотрения.

Уменьшение размерности входных данных не использовалось, поскольку данные методы преобразуют исходное пространство предикторов и усложняют интерпретируемость.

Создание ПМ было реализовано такими методами МО, как метод опорных векторов, случайный лес, стохастический градиентный бустинг. Выбор данных методов МО обусловлен необходимостью решения задачи выделения информативных предикторов (признаков). Все вышеуказанные алгоритмы обладают «встроенной» возможностью определения степени важности предиктора в процессе создания ПМ.

Идея метода опорных векторов заключается в поиске гиперплоскости (например, для двумерных данных это прямая, для трехмерных – плоскость и т. д.) разделяющая данные таким образом, чтобы максимизировать суммарное расстояние от векторов из одного класса до нее. Наиболее близкие к разделяющей гиперплоскости называются опорными. Степень близости определяет величину информационного вклада.

Методы случайный лес и стохастический градиентный бустинг используют бинарные деревья для построения ПМ. Различие заключается в том, что стохастический градиентный бустинг последовательно обучает новые деревья для коррекции ошибок предыдущих, а случайный лес параллельно обучает большой массив деревьев и принимает итоговое решение, например, путем голосования большинством. Поскольку в обоих методах используются бинарные деревья, можно в явном виде в узлах ветвления вычислять наиболее информативные предикторы. Кроме того, вышеперечисленные методы хорошо себя зарекомендовали в задачах анализа медицинских данных высокой размерности и малых выборок рассмотрения [11].

Валидация полученных моделей производилась методом перекрестной 10-кратной проверки совместно с ROC-AUC анализом (Receiver Operation Characteristics – Area Under Curve). Суть метода заключается в разбиении набора исходных данных на 10 поднаборов одинакового размера и построения 10 моделей на полученных поднаборах. Для каждой модели строится ROC-AUC график, показывающий зависимость чувствительности от 1-специфичности. Площадь под графиком изменяется в диапазоне [0; 1]. Чем ближе значение площади к 0 или 1, тем меньше ошибок совершает модель. Если площадь близка к 0,5, модель случайно предсказывает результат.

of the predictor in the process of creating a PM. The idea of the SVM is to find a hyperplane (for example, for two-dimensional data it is a straight line, for three-dimensional data it is a plane, etc.) separating the data in such a way as to maximize the total distance from vectors from one class to it. The closest to the separating hyperplane are called reference. The degree of proximity determines the amount of information contribution.

The random forest and stochastic gradient boosting methods use binary trees to construct PM. The difference is that the stochastic gradient boosting consistently trains new trees to correct the errors of previous ones, while the random forest simultaneously trains a large array of trees and makes the final decision, for example, by majority vote. Since both methods use binary trees, it is possible to explicitly calculate the most informative predictors at the branch nodes. In addition, the above methods have proven themselves well in the tasks of analyzing high-dimensional medical data and small samples of examination [11].

Validation of the obtained models was performed by the method of 10-fold cross-validation with ROC-AUC analysis (Receiver Operation Characteristics - Area Under Curve). The content of the method is to divide a set of source data into 10 subsets of the same size and build 10 models on the obtained subsets. For each model, a ROC-AUC graph is constructed showing the dependence of sensitivity on 1-specificity. The area under the graph varies in the range [0; 1]. The closer the area value is to 0 or 1, the fewer errors the model makes. If the area is close to 0.5, the model randomly predicts the result.

Results

In the PM we created, the predictors of mortality were urea and body temperature for random forest and stochastic gradient boosting methods, erythrocyte, eosinophil and monocyte counts, and INR level for support vector machine (Fig. 1).

All presented models passed validation, the ROC-AUC curves were developed, showing that the area under the curve varies in the range from 0.67 to 1. This indicates a high predictive value of the models (Fig. 2).

Discussion

Literature data shows that the severity and outcome of COVID-19 can be determined by the results of laboratory tests in the early stages of a patient's hospitalization

The most noticeable changes in laboratory tests results were observed in samples taken from severely and critically ill patients. Changes in hematological parameters (lymphopenia, leukocytosis with an increase in the number of neutrophils, thrombocytopenia) positively correlated with the severity of the disease. A trend toward increased CRP, ferritin, procalcitonin IL-6, and decreased albumin and/or prealbumin was frequently observed during progression from mild to severe/critical illness and in patients who died. Serial measurements of these markers can be used to predict disease progression, severity, and mortality [9, 12].

Changes in coagulation parameters, including prolonged PT, elevated D-dimer levels, and elevated fibrinogen levels, were common findings in severe disease and non-survivors [13]. Other biomarkers such as lactate dehydrogenase (LDH), creatine kinase (CK), AST and ALT have been associated with severe and critical illness in several studies, and their levels are likely to indicate poor outcome [14].

Результаты

В созданных нами ПМ предикторами смертельного исхода для методов случайный лес и стохастический градиентный бустинг являлись мочевины и температура тела, для метода опорных векторов количество эритроцитов, эозинофилов и моноцитов, уровень МНО (рис. 1).

Все представленные модели прошли валидацию; разработанные ROC-AUC графики показали, что площадь под графиком варьирует от 0,67 до 1. Это говорит о высокой предиктивной ценности моделей (рис. 2).

Обсуждение

Данные литературы свидетельствуют о том, что тяжесть и исход заболевания COVID-19 можно определить по результатам лабораторных тестов на ранних этапах госпитализации пациента.

Наиболее заметные изменения в результатах лабораторных исследований чаще наблюдались в образцах, взятых у тяжелых и критически больных пациентов. Изменение гематологических параметров (лимфопения, лейкоцитоз с увеличением числа нейтрофилов, тромбоцитопения) положительно коррелировало с тяжестью заболевания. Тенденция к повышению СРБ, ферритина, прокальцитонина ИЛ-6 и снижению альбумина и/или преальбумина часто наблюдалась при прогрессировании от легкого до тяжелого / критического состояния, а также у умерших пациентов. Серийные измерения этих маркеров можно использовать для прогнозирования течения заболевания, тяжести и смертности [9, 12].

Изменения параметров свертывания крови, в том числе удлинение протромбинового времени, повышенный уровень D-димера и повышенный уровень фибриногена были частыми «находками» при тяжелом заболевании и у не выживших [13]. Другие биомаркеры, такие как лактатдегидрогеназа (ЛДГ), креатинкиназа (КК), АСТ и АЛТ, в нескольких исследованиях были связаны с тяжелыми и критическими состояниями; и их уровни, вероятно, указывают на неблагоприятный исход [14].

В проведенном нами исследовании две ПМ, созданные с помощью методов МО случайный лес и стохастический градиентный бустинг, показали, что прогностическое значение имеют изменения двух показателей: уровня мочевины и температуры тела. Метод опорных векторов выявил другие предикторы. Поскольку два предшествующих метода (случайный лес и стохастический градиентный бустинг) показали схожие результаты, мы не учитывали данные, полученные с помощью метода опорных векторов. Подобный подход выбора ПМ голосованием часто используется при оценке данных методами искусственного интеллекта [11, 15, 16].

Повышенная температура тела является показателем тяжести иммунновоспалительного ответа при остром COVID-19 и является предиктором тяжелого течения заболевания и смертности [17]. Результаты нашего исследования согласуются с ранее сделанными выводами о том, что повышение температуры тела может быть маркером прогностически неблагоприятного варианта течения подострого COVID-19.

Мочевина – это основной конечный продукт распада белков, являющийся химической формой, в которой азот выводится из организма. В повышенных концентрациях этот метаболит проявляет свойства токсина, оказывающего повреждающее действие на органы и ткани организма. Мочевина оказывает токсическое действие на

In our study two predictive models created using machine learning methods – random forest and stochastic gradient boosting showed that changes in two indicators (urea level and body temperature) have prognostic significance. Support vector machine identified other predictors. Since the two previous methods (random forest and stochastic gradient boosting) showed similar results, we did not take into account the data obtained using support vector machine. A similar approach to selecting a predictive model by voting is often used when assessing data using artificial intelligence methods [11, 15, 16].

Elevated body temperature is an indicator of the severity of the immune-inflammatory response in acute COVID-19 and is a predictor of severe disease and mortality [17]. The results of our study are consistent with previously made conclusions that an increase in body temperature may be a marker of a prognostically unfavorable course of subacute COVID-19.

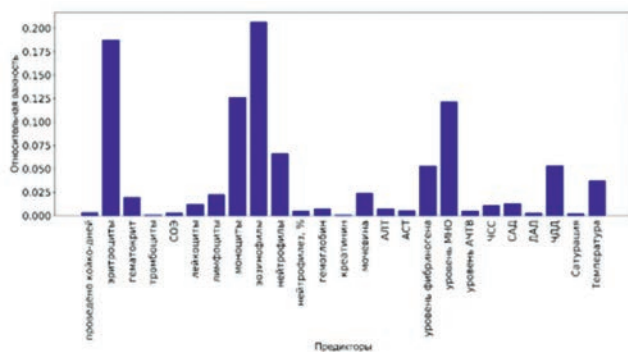
Urea is the main end product of protein breakdown and is the chemical form in which nitrogen is excreted from the body. In elevated concentrations, this metabolite exhibits the properties of a toxin that has a damaging effect on organs and tissues of the body. Urea has toxic effects on the gastrointestinal tract (promoting the breakdown of the epithelial barrier and altering the microbiome), on the kidneys (indirectly promoting renal fibrosis), on adipocytes (inducing insulin resistance), on blood components (inducing carbamylation of erythropoietin) and on the cardiovascular system [18].

High urea levels are observed in patients with kidney disease in the advanced stages of renal failure, which increases the risk of adverse cardiovascular events. This group of patients has a high level of carbamylated low- and high-density lipoproteins (cLDL, cHDL), which is a consequence of uremia and leads to the development of endothelial dysfunction and accelerates the processes of atherogenesis [19].

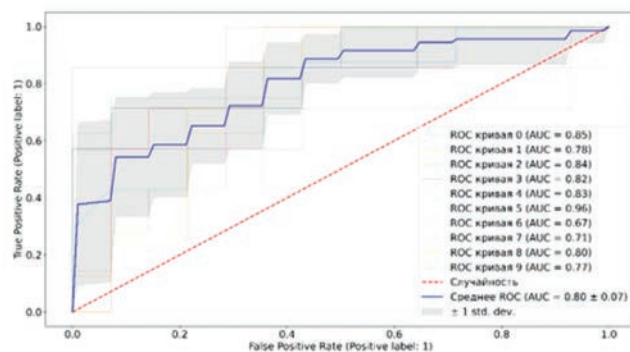
It is known that exposure to urea induces reactive oxygen species in the endothelial cells of the human aorta and arteries, which leads to the activation of pro-inflammatory cytokines and inactivation of the anti-atherosclerotic enzyme PGI₂ synthase, changes in the level of mitochondrial proteins and the expression of inflammatory markers [17].

In an in vitro study, urea caused a decrease in the rate of endothelial cell proliferation after 72 hours of exposure and appeared to be a potential stimulus for endothelial-mesenchymal transition (EndMT). It also induced rearrangement of actin filaments, a significant increase in the expression of matrix metalloproteinases 2 (MMP-2) and changes in the level of other EndMT biomarkers (keratin, fibrillin-2 and collagen IV). After exposure of endothelial cell culture to urea, the expression of dimethylarginine dimethylaminohydrolase (DDAH) and vasorin decreased. Both proteins have been directly associated with cardiovascular disease in in vitro and in vivo studies [20].

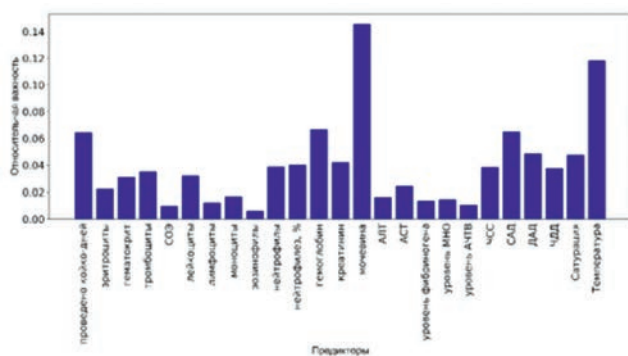
It should be noted that in our study the deceased patients did not have a history of chronic renal failure. Literature data suggests that many hospitalized patients with COVID-19 suffer from kidney damage in the form of proteinuria, hematuria, or acute kidney injury (AKI). AKI is particularly common among severely and critically ill patients with COVID-19 and is a predictor of mortality. The pathophysiology of AKI in COVID-19 is unclear. Early reports of histopathological examination of autopsy kidney tissue



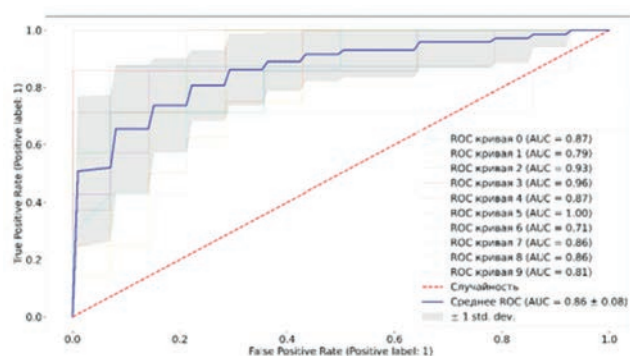
(A)



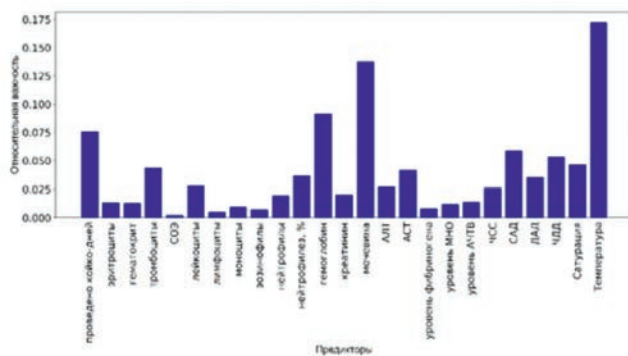
(A)



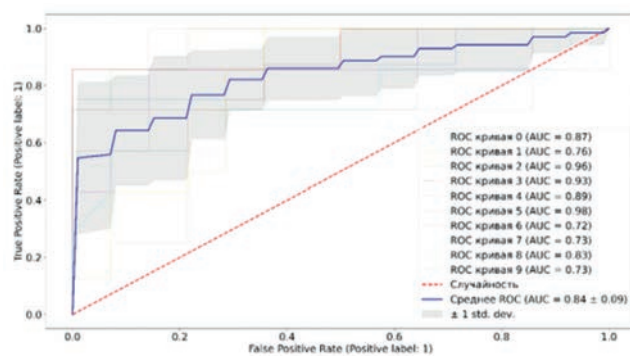
(B)



(B)



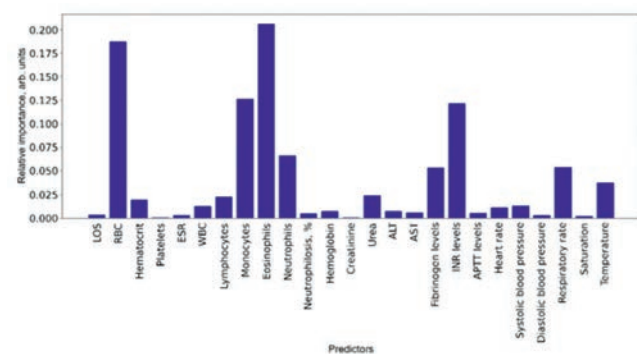
(B)



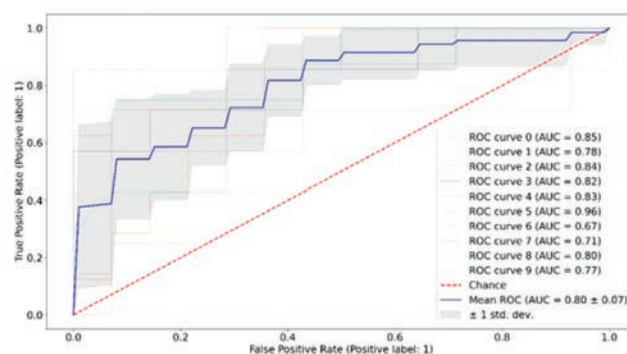
(B)

Рис. 1. Наиболее информативные предикторы риска для методов опорных векторов (а), случайный лес (б), стохастический градиентный бустинг (в)

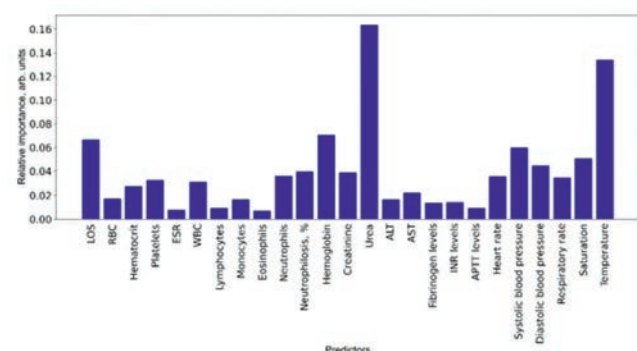
Рис. 2. ROC-AUC анализ для методов опорных векторов (а), случайный лес (б), стохастический градиентный бустинг



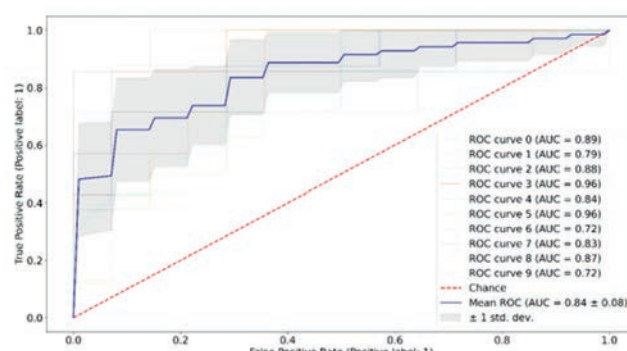
(A)



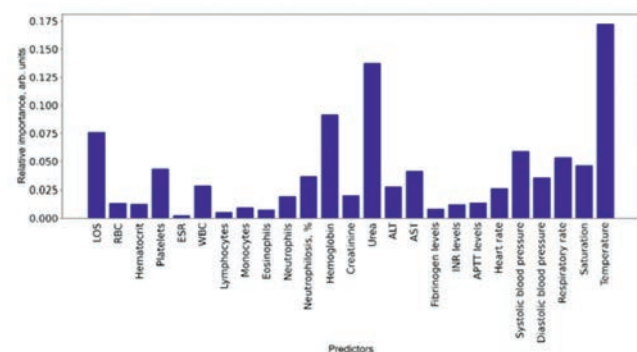
(A)



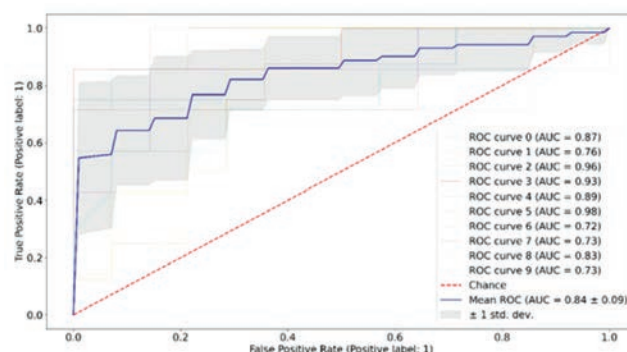
(B)



(B)



(C)



(C)

Fig. 1. Assessment of predictive value of fatal outcome of laboratory and clinical tests during subacute COVID-19 analyzed with support vector machine (a), random forest (b), stochastic gradient boosting (c)

Fig. 2. ROC-AUC analysis for support vector machine (a), random forest (b), stochastic gradient boosting (c)

желудочно-кишечный тракт (способствуя разрушению эпителиального барьера и изменению микробиома), почки (косвенно способствуя почечному фиброзу), адипоциты (индуцируя резистентность к инсулину), компоненты крови (вызывая карбамиллирование эритропоэтина) и сердечно-сосудистую систему [18].

Высокие уровни мочевины наблюдаются у пациентов с заболеваниями почек на стадии развития почечной недостаточности, что повышает риск неблагоприятных сердечно-сосудистых событий. У этой группы больных отмечается высокий уровень карбамиллированных липопротеинов низкой и высокой плотности (цЛПНП, цЛПВП), что является следствием уремии, приводит к развитию эндотелиальной дисфункции и ускоряет процессы атерогенеза [19].

Известно, что воздействие мочевины индуцирует выработку активных форм кислорода в эндотелиальных клетках аорты и артерий человека, что приводит к активации провоспалительных цитокинов и инактивации антиатеросклеротического фермента PGI₂-синтазы, изменению уровня митохондриальных белков и экспрессии воспалительных маркеров [17].

В исследовании *in vitro* мочевина вызывала снижение скорости пролиферации эндотелиальных клеток через 72 ч воздействия и, по-видимому, являлась потенциальным стимулом эндотелиально-мезенхимального перехода (EndMT), а также индуцировала перестройку актиновых филаментов, значительное увеличение экспрессии матриксных металлопротеиназ 2 (MMP-2) и изменение уровня других биомаркеров EndMT (кератин, фибриллин-2 и коллаген IV). После воздействия на культуру эндотелиальных клеток мочевины снижалась экспрессия диметиларгининдиметиламиногидролазы (DDAH) и вазоприна. Оба белка были напрямую связаны с сердечно-сосудистыми заболеваниями в исследованиях *in vitro* и *in vivo* [20].

Следует отметить, что погибшие пациенты не имели в анамнезе хронической почечной недостаточности. Данные литературы свидетельствуют о том, что многие госпитализированные пациенты с COVID-19 страдают поражением почек в виде протеинурии, гематурии или острой почечной недостаточности (ОПН). ОПН особенно распространена среди тяжелых и критически больных пациентов с COVID-19 и является предиктором смертности. Патофизиология ОПН при COVID-19 не ясна. Ранние отчеты о гистопатологическом исследовании аутопсийной ткани почки показывают вирусные частицы SARS-CoV-2 в клетках почечных канальцев и подоцитах, что свидетельствует о прямой вирусной инфекции, а также об остром некрозе канальцев; также возможны рабдомиолиз и гломерулопатия [21]. Вероятно, повышение уровня мочевины являлось пусковым механизмом, ведущим к эндотелииту и последующему инфаркту миокарда до того, как развилась ОПН.

Заключение

Использование трех методов МО с последующим голосованием позволило выявить два предиктора неблагоприятного исхода при подостром течении COVID-19. Мы впервые показали, что повышение уровня мочевины имеет неблагоприятное прогностическое значение; при этом причиной смерти пациентов явился инфаркт миокарда.

show SARS-CoV-2 viral particles in renal tubular cells and podocytes, suggesting direct viral infection, as well as acute tubular necrosis, rhabdomyolysis, and glomerulopathy may also occur [21]. It is possible that elevated urea levels were the trigger leading to endothelitis and subsequent myocardial infarction before AKI developed.

Conclusion

The use of three ML methods with subsequent voting allowed identifying two predictors of an unfavorable outcome in post-acute COVID-19. We have shown for the first time that an increase in urea levels has an unfavorable prognostic value, while the cause of death of patients is myocardial infarction.

Литература / References

1. Nalbandian A., Sehgal K., Gupta A., Madhavan M.V., McGroder C., Stevens J.S. et al. Post-acute COVID-19 syndrome. *Nat. Med.* 2021;27(4):601–615. <https://doi.org/10.1038/s41591-021-01283-z>
2. Shah W., Hillman T., Playford E.D., Hishmeh L. Managing the long term effects of COVID-19: Summary of NICE, SIGN, and RCGP rapid guideline. *Brit. Med. J.* 2021;372:136–139. <https://doi.org/10.1136/bmj.n136>
3. Maestre-Muñoz M.M., Arias Á., Mata-Vázquez E., Martín-Toledano M., López-Larramona G., Ruiz-Chicote A.M. et al. Long-term outcomes of patients with coronavirus disease 2019 at one year after hospital discharge. *J. Clin. Med.* 2021;10(13):2945–2949. <https://doi.org/10.12775/JEHS.2022.12.04.006>
4. Xie Y., Xu E., Bowe B., Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. *Nat. Med.* 2022;28(3):583–590. <https://doi.org/10.1038/s41591-022-01689-3>
5. Puntmann V.O., Carerj M.L., Wieters I. Outcomes of cardiovascular magnetic resonance imaging in patients recently recovered from coronavirus disease 2019 (COVID-19). *JAMA Cardiol.* 2020;5:1265–1273. <https://doi.org/10.1001/jamacardio.2020.3557>
6. Stokes E., Zambrano L., Anderson K., Marder E., Raz K., El Burai F. et al. Coronavirus disease 2019 case surveillance – United States, January 22–May 30, 2020. *MMWR Morb. Mortal. Wkly Rep.* 2020;69:759. <https://doi.org/10.15585/mmwr.mm6924e2>
7. John K.J., Mishra A.K., Ramasamy C., George A.A., Selvaraj V., Lal A. Heart failure in COVID-19 patients: critical care experience. *World J. Virol.* 2022;11:1–19. <https://doi.org/10.5501/wjv.v11.i1.1>
8. Mehandru S., Merad M. Pathological sequelae of long-haul COVID. *Nat. Immunol.* 2022;23(2):94–202. <https://doi.org/10.1038/s41590-021-01104-y>
9. Ji M., Yuan L., Shen W., Lv J., Li Y., Li M. et al. Characteristics of disease progress in patients with coronavirus disease 2019 in Wuhan, China. *Epidemiology and Infection.* 2020;148:94–97. <https://doi.org/10.1017/S0950268820000977>
10. Collett D. Modelling survival data in medical research (4th ed.). New York: Chapman and Hall/CRC. 2023:556. <https://doi.org/10.1201/9781003282525>
11. Vrazhnov D., Mankova A., Stupak E., Kistenev Y., Shkurinov A., Cherkasova O. Discovering glioma tissue through its biomarkers' detection in blood by raman spectroscopy and machine learning. *Pharmaceutics.* 2023;15(1):203–215. <https://doi.org/10.3390/pharmaceutics15010203>
12. Gong J., Dong H., Xia Q.S., Huang Z.Y., Wang D.K., Zhao Y. et al. Correlation analysis between disease severity and inflammation-related parameters in patients with COVID-19: a retrospective study. *BMC Infect. Dis.* 2020;20(1):963. <https://doi.org/10.1186/s12879-020-05681-5>
13. Liao D., Zhou F., Luo L., Xu M., Wang H., Xia J. et al. Haematological characteristics and risk factors in the classification and prognosis evaluation of COVID-19: a retrospective cohort study. *Lancet Haematol.* 2020;7(9):e671–e678. [https://doi.org/10.1016/S2352-3026\(20\)30217-9](https://doi.org/10.1016/S2352-3026(20)30217-9)
14. Bloom P.P., Meyerowitz E.A., Reinus Z., Daidone M., Gustafson J., Kim A.Y. Liver Biochemistries in Hospitalized Patients With COVID-19. *Hepatology.* 2021;73(3):890–900. <https://doi.org/10.1002/hep.31326>
15. Bashir S., Almazroi A., Ashfaq A., Almazroi A., Khan F. A knowledge-based clinical decision support system utilizing an intelligent ensemble voting scheme for improved cardiovascular disease prediction. *IEEE Access.* 2021;9:130805–130822. <https://doi.org/10.1109/ACCESS.2021.3110604>

16. Fathima M.D., Samuel S., Natchadalingam R., Kaveri V. Majority voting ensembled feature selection and customized deep neural network for the enhanced clinical decision support system. *International Journal of Computers and Applications*. 2022;44(10):991–1001. <https://doi.org/10.1080/1206212X.2022.2069643>
17. Tharakan S., Nomoto K., Miyashita S., Ishikawa K. Body temperature correlates with mortality in COVID-19 patients. *Crit. Care*. 2020;24:298. <https://doi.org/10.1186/s13054-020-03045-8>
18. Lau W.L., Vaziri N.D. Urea, a true uremic toxin: the empire strikes back. *Clin. Sci. (Lond)*. 2017;131(1):3–12. <https://doi.org/10.1042/CS20160203>
19. d'Apolito M., Colia A.L., Manca E., Pettoello-Mantovani M., Sacco M., Maffione A.B., Brownlee M., Giardino I. Urea memory: transient cell exposure to urea causes persistent mitochondrial ROS production and endothelial dysfunction. *Toxins (Basel)*. 2018;10(10):410. <https://doi.org/10.3390/toxins10100410>
20. Colombo G., Altomare A., Astori E., Landoni L., Garavaglia M.L., Rossi R. et al. Effects of Physiological and Pathological Urea Concentrations on Human Microvascular Endothelial Cells. *Int. J. Mol. Sci.* 2022;24(1):691. <https://doi.org/10.3390/ijms24010691>
21. Ertuğlu L.A., Kanbay A., Afşar B., Elşürer Afşar R., Kanbay M. COVID-19 and acute kidney injury. *Tuberk. Toraks*. 2020;68(4):407–418. <https://doi.org/10.5578/tt.70010>

Информация о вкладе авторов

Долгалева И.В. – планирование исследования, клинические исследования, подготовка рукописи.
Вражнов Д.А. – анализ данных, подготовка рукописи.
Толмачев И.В. – написание текста, поиск литературы.
Старикова Е.Г. – написание текста, составление списка литературы, поиск литературы.
Каверина И.С. – сбор данных, поиск литературы.
Завьялова М.В. – концепция исследования, рецензирование рукописи.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Information on author contributions

Dolgaleva I.V. – research design, clinical studies, manuscript preparation.
Vrazhnov D.A. – data analysis, manuscript preparation.
Tolmachev I.V. – writing the text, literature search.
Starikova E.G. – writing the text, compilation of a literature list, literature search.
Kaverina I.S. – data acquisition, literature search.
Zavyalova M.V. – research concept, manuscript review.

Conflict of interest: the authors do not declare a conflict of interest.

Сведения об авторах

Долгалёв Игорь Владимирович, д-р мед. наук, профессор, заведующий кафедрой факультетской терапии с курсом клинической фармакологии, СибГМУ, Томск, Россия, <http://orcid.org/0000-0003-2658-0181>.

E-mail: dolgalev.iv@ssmu.ru.

Вражнов Денис Александрович, научный сотрудник Научно-технологического центра «Цифровая медицина и киберфизика», СибГМУ, Томск, <http://orcid.org/0000-0002-6915-6156>.

E-mail: vrazhnov.da@ssmu.ru.

Толмачев Иван Владиславович, канд. мед. наук, ведущий научный сотрудник, руководитель Научно-технологического центра «Цифровая медицина и киберфизика», СибГМУ, Томск, Россия; ведущий научный сотрудник, отдел научных основ организации здравоохранения, ЦНИИО-ИЗ, Москва, Россия, <http://orcid.org/0000-0002-2888-5539>.

E-mail: tolmachev.iv@ssmu.ru.

Старикова Елена Григорьевна, д-р мед. наук, ведущий научный сотрудник, Научно-технологический центр «Цифровая медицина и киберфизика», СибГМУ, Томск, Россия, <http://orcid.org/0000-0002-8899-0795>.

E-mail: elena.g.starikova@yandex.ru.

Каверина Ирина Сергеевна, научный сотрудник, Научно-технологический центр «Цифровая медицина и киберфизика», СибГМУ, Томск, Россия, <http://orcid.org/0000-0001-9748-482X>.

E-mail: kaverina.is@ssmu.ru.

Завьялова Марина Викторовна, д-р мед. наук, профессор, заведующий кафедрой патологической анатомии СибГМУ, Томск, Россия, <http://orcid.org/0000-0001-9429-9813>.

E-mail: zavyalova.mv@ssmu.ru.

 **Старикова Елена Григорьевна**,
e-mail: elena.g.starikova@yandex.ru.

Information about the authors

Igor V. Dolgalev, Dr. Sci. Med., Professor, Head of the Department of Faculty Therapy with a course in Clinical Pharmacology of SSMU, Tomsk, <http://orcid.org/0000-0003-2658-0181>.

E-mail: dolgalev.iv@ssmu.ru.

Denis A. Vrazhnov, Research Scientist, Scientific and Technological Center "Digital Medicine and Cyberphysics" of SSMU, Tomsk, <http://orcid.org/0000-0002-6915-6156>.

E-mail: vrazhnov.da@ssmu.ru.

Ivan V. Tolmachev, Cand. Sci. (Med.), Leading Research Scientist, Head of Scientific and Technological Center "Digital medicine and cyberphysics" of SSMU, Tomsk; Senior Research Fellow, Division of Health Sciences, Russian Research Institute of Health, Moscow, <http://orcid.org/0000-0002-2888-5539>.

E-mail: tolmachev.iv@ssmu.ru.

Elena G. Starikova, Dr. Sci. (Med.), Leading Research Scientist, Scientific and Technological Center "Digital Medicine and Cyberphysics", SSMU, Tomsk, <http://orcid.org/0000-0002-8899-0795>.

E-mail: elena.g.starikova@yandex.ru.

Irina S. Kaverina, Research Scientist, Scientific and Technological Center "Digital medicine and cyberphysics" of SSMU, Tomsk, Russia, <http://orcid.org/0000-0001-9748-482X>.

E-mail: kaverina.is@ssmu.ru.

Marina V. Zavyalova, Dr. Sci. (Med.), Professor, Head of the Department of Pathological Anatomy, SSMU, Tomsk, <http://orcid.org/0000-0001-9429-9813>.

E-mail: zavyalova.mv@ssmu.ru.

 **Elena G. Starikova**, e-mail: elena.g.starikova@yandex.ru.

Received 28.05.2024;
review received 09.12.2024;
accepted for publication 10.02.2025.

Поступила 28.05.2024;
рецензия получена 09.12.2024;
принята к публикации 10.02.2025.