



ПРОГНОЗ СТЕПЕНИ ТЯЖЕСТИ ТЕЧЕНИЯ SARS-COV-2-ИНФЕКЦИИ У ЛИЦ МОЛОДОГО ВОЗРАСТА С ПРИМЕНЕНИЕМ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

К.В. Касьяненко¹, К.В. Козлов¹, К.В. Жданов¹, И.И. Лапиков², В.В. Беликов²

¹ Военно-медицинская академия им. С.М. Кирова, Санкт-Петербург, Россия

² МИРЭА — Российский технологический университет, Москва, Россия

SARS-CoV-2 severity prediction in young adults using artificial intelligence

K.V. Kas'janenko¹, K.V. Kozlov¹, K.V. Zhdanov¹, I.I. Lapikov², V.V. Belikov²

¹ Military Medical Academy named after S.M. Kirov, Saint-Petersburg, Russia

² MIREA — Russian Technological University, Moscow, Russia

Резюме

Цель: построить с использованием методов искусственного интеллекта предсказательную модель тяжелого течения COVID-19 у лиц молодого возраста

Материалы и методы: проанализированы данные 906 историй болезни пациентов в возрасте от 18 до 44 лет с лабораторно верифицированной SARS-CoV-2-инфекцией за период 2020–2021 гг. Оценка лабораторных и инструментальных данных осуществлялась с помощью U-критерия Манна — Уитни с уровнем статистической значимости. Обучение нейросетевой модели проводилось с использованием фреймворка Pytorch.

Результаты. У пациентов с легкой и средней степени тяжести SARS-CoV-2-инфекции периферическая кислородная сатурация, содержание эритроцитов, гемоглобина, общего белка, альбумина, уровень гематокрита, сывороточного железа, трансферрина, а также абсолютное число эозинофилов и лимфоцитов периферической крови были статистически значимо выше, чем у пациентов с тяжелой степенью тяжести заболевания ($p < 0,001$). Значения абсолютного числа нейтрофилов, СОЭ, глюкозы, АЛТ, АСТ, КФК, мочевины, ЛДГ, ферритина, СРБ, фибриногена, D-димера, ЧДД, ЧСС, уровень артериального давления в группе пациентов легкой и средней степени тяжести были статистически значимо ниже, чем в группе тяжелых пациентов ($p < 0,001$). Выделено 11 показателей, являющихся предикторами тяжелого течения (уровень периферической кислородной сатурации, количество эритроцитов, уровень гемоглобина, абсолютные значения эозинофилов, лимфоцитов, абсолютные значения нейтрофилов, уровень ЛДГ, ферритина, С-реактивного белка, D-димера) и их пороговые значения. Разработанная с использованием методов искусственного интеллекта прогностическая модель с высокой чувствительностью и специфичностью способна предсказать развитие тяжелого течения SARS-CoV-2-инфекции у пациентов молодого возраста.

Заключение. Значения лабораторных и инструментальных показателей, полученных у пациентов с SARS-CoV-2-инфекцией различной степени тяжести при поступлении на стационарное лечение, статистически значимо отличаются. Среди них выделены 11 показателей, которые достоверно связаны с развитием тя-

Abstract

Aim: to build a predictive model for severe COVID-19 prediction in young adults using deep learning methods.

Materials and methods: data from 906 medical records of patients aged 18 to 44 years with laboratory-confirmed SARS-CoV-2 infection during 2020–2021 period was analyzed. Evaluation of laboratory and instrumental data was carried out using the Mann-Whitney U-test. The level of statistical significance was $p \leq 0,05$. The neural network was trained using the Pytorch framework.

Results: in patients with mild to moderate SARS-CoV-2 infection, peripheral oxygen saturation, erythrocytes, hemoglobin, total protein, albumin, hematocrit, serum iron, transferrin, and absolute peripheral blood eosinophil and lymphocyte counts were significantly higher than in patients with severe COVID-19 ($p < 0,001$). The values of the absolute number of neutrophils, ESR, glucose, ALT, AST, CPK, urea, LDH, ferritin, CRP, fibrinogen, D-dimer, respiration rate, heart rate, blood pressure in the group of patients with mild and moderate severity were statistically significantly lower than in the group of severe patients ($p < 0,001$). Eleven indicators were identified as predictors of severe COVID-19 (peripheral oxygen level, peripheral blood erythrocyte count, hemoglobin level, absolute eosinophil count, absolute lymphocyte count, absolute neutrophil count, LDH, ferritin, C-reactive protein, D-dimer levels) and their threshold values. A model intended to predict COVID-19 severity in young adults was built.

Conclusion. The values of laboratory and instrumental indicators obtained in patients with SARS-CoV-2 infection upon admission significantly differ. Among them eleven indicators were significantly associated with the development of a severe COVID-19. A predictive model based on artificial intelligence method with high accuracy predicts the likelihood of severe SARS-CoV-2 course development in young adults.

желого течения. На основе методов искусственного интеллекта построена прогностическая модель глубокого обучения, которая с высокой точностью предсказывает развитие тяжелого течения SARS-CoV-2-инфекции у лиц молодого возраста на этапе госпитализации.

Ключевые слова: COVID-19, SARS-CoV-2, новая коронавирусная инфекция, искусственный интеллект.

Введение

Активное развитие теоретических положений искусственного интеллекта как науки и технологий создания интеллектуальных машин произошло в последней четверти XX в. и было обусловлено новым этапом развития вычислительных устройств [1]. В области здравоохранения принципы искусственного интеллекта используются при построении экспертных систем, основанных на опыте высококвалифицированных медицинских специалистов, который формализуется в виде математического алгоритма с последующей его реализацией в виде программного продукта [2]. Одной из первых экспертных систем стала MYCIN, которая оптимизировала подбор антибактериального препарата для лечения пациентов с полирезистентной микрофлорой [3]. При этом большой объем правил, необходимых для корректной работы этой системы, а также сложности при ее интеграции в работу клиник значительно ограничили применение готового продукта. Наличие ограничений при применении экспертных систем и сложность формализации отдельных экспертных знаний способствовали развитию альтернативных методов поддержки принятия решений, среди которых особое место занимают методы машинного обучения, спецификой работы которых является возможность определения закономерностей на основе правил, автоматически формирующихся в процессе моделирования из набора обучающих данных. Еще одним преимуществом машинного обучения стала возможность дообучения готовых моделей для интерпретации сложных клинических случаев [4]. Среди недостатков систем поддержки принятия решений на основе машинного обучения можно отметить зависимость качества предсказания от объема и качества используемой для их разработки базы обучающих признаков.

К наиболее широко используемым в медицине методам машинного обучения относят системы глубокого обучения, искусственные нейронные сети и классические методы дискриминантного и регрессионного анализа [5]. В настоящее время в области инфекционных болезней разработаны 20 систем, предназначенных для работы с данными амбулаторных и стационарных пациентов, однако данных об опыте их применения в реальной клинической практике нет [6–10].

Key words: COVID-19, SARS-CoV-2, novel coronavirus disease, artificial intelligence.

Пандемия SARS-CoV-2-инфекции бросила вызов не только медицинскому сообществу, но и специалистам в области анализа данных. Первые работы, основанные на искусственном интеллекте, были направлены на отслеживание потенциальных источников заболевания [11, 12]. Эффективность указанных моделей привела к расширению ряда задач по разработке систем, предназначенных как для подбора оптимальных терапевтических схем лечения пациентов с SARS-CoV-2-инфекцией в условиях стационара, так и для автоматизированного анализа результатов компьютерной томографии грудной клетки [12]. Еще одним востребованным направлением стало решение задач по предсказанию развития неблагоприятного исхода и/или перевода пациентов в отделения интенсивной терапии [13, 14]. Среди наиболее часто применяемых для построения моделей показателей можно выделить результаты лабораторных анализов пациентов. По мнению исследователей, это позволяет максимально объективизировать входные параметры, исключая возможное искажение по причине субъективной оценки жалоб и данных объективного осмотра [15–18].

Вышеизложенное свидетельствует о том, что растет интерес к разработке и применению в сфере здравоохранения систем поддержки принятия решений, основанных на методах искусственного интеллекта. Отдельные модели находят применение в клинической практике и позволяют обеспечить врача дополнительной информацией при постановке диагноза и выборе схемы лечения. Определенным недостатком таких систем в настоящее время является их зависимость от качества, объема и репрезентативности обучающей выборки, на которые в значительной степени оказывает влияние человеческий фактор, поскольку занесение конкретных экземпляров обучающих данных осуществляется врачом, проводящим осмотр, специалистом, проводящим дополнительные обследования (ультразвуковое и лабораторные исследования, магнитно-резонансная томография и др.) в различных форматах, которые определяются каждым медицинским учреждением самостоятельно.

В настоящее время на государственном уровне в Российской Федерации планируется внедрение централизованной системы электронного обмена медицинской документации, что облегчит работу врачам-специалистам и позволит использовать

данные новых пациентов для разработки математических алгоритмов, направленных на оптимизацию диагностического и лечебного процесса [19].

Необходимость разработки новых моделей для решения конкретных диагностических задач у пациентов с COVID-19 предопределила цель настоящей работы.

Цель исследования – с использованием методов глубокого обучения построить предсказательную модель тяжелого течения COVID-19 у лиц молодого возраста.

Задачи исследования

1. Выявить характерные ранние лабораторные и инструментальные изменения у пациентов молодого возраста с тяжелым течением новой коронавирусной инфекции (COVID-19).

2. Установить предикторы тяжелого течения SARS-CoV-2-инфекции и их пороговые значения.

3. Построить предсказательную модель, которая могла бы с высокой вероятностью прогнозировать развитие тяжелого течения SARS-CoV-2-инфекции у лиц молодого возраста на этапе поступления в стационар.

Материалы и методы исследования

Материалом для настоящего исследования послужили данные историй болезни 906 пациентов (460 (50,8%) пациентов с легкой степенью тяжести заболевания, 354 (39%) пациента со средней степенью тяжести заболевания, 92 (10,2%) пациента с тяжелой степенью тяжести заболевания) в возрасте от 18 до 44 лет, поступивших на стационарное лечение в клиники Военно-медицинской академии им. С.М. Кирова за период 2020–2021 гг., а также данные 40 историй болезни пациентов молодого возраста за период 2022 г. [20]. Диагноз новой коронавирусной инфекции был лабораторно подтвержден методом полимеразной цепной реакции: случаи, вошедшие в обучающую выборку, были представлены штаммами альфа и дельта SARS-CoV-2, вадиационная выборка – преимущественно вариантом омикрон. Определение степени тяжести заболевания основывалось на действующих «Временных методических рекомендациях по профилактике, диагностике и лечению новой коронавирусной инфекции (COVID-19)» [21].

Критериями включения в выборку были:

- мужчины и женщины в возрасте от 18 до 44 лет (включительно);
- срок поступления на стационарное лечение не превышает 3 дней с момента начала болезни.

Критериями исключения были:

- возраст более 44 лет;

– отсутствие достаточного объема (более 60%) данных рассматриваемых лабораторных параметров.

В исследовании при построении признакового пространства для предсказательной модели на базе искусственного интеллекта рассматривались только объективные клинические показатели физического осмотра (частота сердечных сокращений (ЧСС), частота дыхательных движений (ЧДД), периферическая кислородная сатурация (SpO_2), систолическое и диастолическое артериальное давление (АД)). Характеристики, имеющие в своей природе субъективный фактор, исключены из финального набора данных.

Помимо рассмотрения клинических показателей физического осмотра, признаковое пространство расширено ключевыми лабораторными параметрами: проанализированы показатели общеклинического анализа крови (эритроциты (RBC), гемоглобин (HGB), среднее содержание гемоглобина в эритроците (MCHC), гематокрит (Ht), лейкоциты (WBC), абсолютное число эозинофилов (EOS), базофилов (BAS), лимфоцитов (LYMPH), нейтрофилов (NEUT), моноцитов (MON), тромбоциты (PLT), скорость оседания эритроцитов (СОЭ)); показатели биохимического анализа крови (глюкоза, общий белок, альбумин, общий билирубин, аланинаминотрансфераза (АЛТ), аспартатаминотрансфераза (АСТ), креатинфосфокиназа (КФК), креатинфосфокиназа-МВ (КФК-МВ), креатинин, мочевины, лактатдегидрогеназа (ЛДГ), холестерин, сывороточное железо (Fe), ферритин, натрий в сыворотке крови (Na^+), калий в сыворотке крови (K^+), С-реактивный белок (СРБ), прокальцитонин; коагулограмма (фибриноген, протромбиновый индекс (ПТИ), протромбиновое время (ПТВ), международное нормализованное отношение (МНО), активированное частичное тромбопластиновое время (АЧТВ), D-димер)). Оценка качества разработанной предсказательной модели проводилась с помощью ROC-анализа, а в качестве основной метрики выбрана ROC-AUC как интегральное значение чувствительности и специфичности модели.

Для статистического анализа собранных данных выбран непараметрический метод, U-критерий Манна – Уитни с уровнем статистической значимости $p \leq 0,05$. Для определения пороговых значений каждого из потенциальных предикторов визуализирован разброс значений характеристик в рассматриваемых группах с использованием коробчатых диаграмм. Дополнительная валидация гипотез о статистически значимых различиях в рассматриваемых группах проводилась с использованием бутстрапирования и р-перестановочного теста.

Статистический анализ проводился с использованием библиотеки SciPy¹ пакета для научных исследований Anaconda² языка Python, визуализация данных с применением пакета matplotlib³. Обучение нейронной сети проводилось с использованием фреймворка PyTorch⁴.

Исследование одобрено Комитетом по вопросам этики при Военно-медицинской академии им. С.М. Кирова (г. Санкт-Петербург) 26 апреля 2022 г. (протокол № 262).

Результаты исследования

Для определения предикторов тяжелого течения инфекции, вызванной SARS-CoV-2, пациенты с лег-

ким и среднетяжелым течением COVID-19 объединены в группу «Нетяжелые», которая сравнивалась с группой больных, у которых инфекционный процесс протекал тяжело («Тяжелые»).

Сравнение средних значений данных физического осмотра и показателей лабораторных анализов у пациентов из сравниваемых групп представлено в таблице 1.

Статистически значимые различия между «нетяжелыми» и «тяжелыми» пациентами наблюдались по следующим признакам: ЧДД, ЧСС, SpO₂, систолическое АД, диастолическое АД, RBC, HGB, Ht, СОЭ, EOS, LYMPH, NEUT, глюкоза, общий белок, общий билирубин, альбумин, АЛТ, АСТ, КФК,

Таблица 1

Характеристика пациентов с инфекцией, вызванной SARS-CoV-2, в зависимости от тяжести течения инфекционного процесса на момент поступления в стационар (Me (Q1–Q3), значение p для U-теста Манна – Уитни)

Признак	«Нетяжелые» (n = 814)	«Тяжелые» (n = 92)	p-value
ЧДД	16,33 (16,0 – 17,0)	18,38 (17,0 – 20,0)	< 0,001
SpO ₂	97,26 (96,0 – 98,0)	95,06 (93,0 – 98,0)	< 0,001
ЧСС	74,14 (68,0 – 80,0)	83,83 (74,5 – 92,0)	< 0,001
Систолическое АД	119,06 (115,0 – 120,0)	122,69 (115,0 – 130,0)	0,01
Диастолическое АД	75,56 (70,0 – 80,0)	78,4 (70,0 – 84,0)	0,01
RBC, 10 ¹² /л	5,07 (4,76 – 5,37)	4,8 (4,46 – 5,12)	< 0,001
HGB, г/л	147,56 (140,0 – 157,0)	138,59 (130,0 – 148,0)	< 0,001
MCHC, г/дл	339,47 (325,9 – 353,0)	334,32 (321,35 – 345,4)	0,1
Ht, л/л	0,44 (0,41 – 0,46)	0,41 (0,39 – 0,45)	< 0,001
WBC, 10 ⁹ /л	5,11 (4,1 – 6,2)	5,35 (3,9 – 6,7)	0,48
СОЭ, мм/ч	6,72 (3,0 – 10,5)	21,56 (10,0 – 31,5)	< 0,001
EOS, 10 ⁹ /л	0,1 (0,05 – 0,16)	0,06 (0,04 – 0,08)	0,01
BAS, 10 ⁹ /л	0,01 (0,0 – 0,03)	0,18 (0,0 – 0,35)	0,31
LYMPH, 10 ⁹ /л	1,65 (1,27 – 2,06)	1,39 (0,97 – 1,83)	< 0,001
NEUT, 10 ⁹ /л	2,73 (1,99 – 3,58)	3,47 (2,27 – 4,66)	0,01
MON, 10 ⁹ /л	0,43 (0,3 – 0,56)	0,47 (0,36 – 0,61)	0,14
PLT, 10 ⁹ /л	203,59 (167,0 – 242,0)	203,14 (154,0 – 252,5)	0,15
Глюкоза, ммоль/л	4,85 (4,33 – 5,28)	5,66 (4,72 – 6,69)	< 0,001
Общий белок, г/л	73,43 (69,9 – 77,0)	68,94 (64,7 – 72,85)	< 0,001
Общий билирубин, мкмоль/л	13,43 (8,85 – 17,6)	11,32 (8,9 – 13,4)	0,01
Альбумин, г/л	43,87 (41,2 – 46,3)	39,99 (37,4 – 42,4)	< 0,001
АЛТ, Ед/л	22,74 (14,3 – 31,05)	33,22 (20,4 – 46,0)	< 0,001
АСТ, Ед/л	27,05 (20,0 – 33,2)	35,33 (25,5 – 43,0)	< 0,001
КФК, Ед/л	113,18 (73,95 – 154,6)	221,23 (83,45 – 311,55)	< 0,001
КФК-МВ, Ед/л	13,05 (10,6 – 15,65)	13,38 (10,6 – 16,75)	0,99
Креатинин, ммоль/л	90,37 (82,5 – 98,05)	89,98 (80,0 – 102,0)	0,7

¹ <https://scipy.org/>

² <https://www.anaconda.com/>

³ <https://matplotlib.org/>

⁴ <https://pytorch.org/>

Признак	«Нетяжелые» (n = 814)	«Тяжелые» (n = 92)	p-value
Мочевина, ммоль/л	4,68 (3,92 – 5,4)	5,46 (4,15 – 6,82)	< 0,001
ЛДГ, Ед/л	170,28 (142,25 – 200,7)	246,06 (171,27 – 282,75)	< 0,001
Холестерин, ммоль/л	3,82 (3,32 – 4,31)	4,04 (3,3 – 4,64)	0,12
Ферритин, нг/мл	163,41 (81,35 – 251,65)	385,06 (159,8 – 591,36)	< 0,001
Fe, мкмоль/л	16,65 (9,6 – 24,1)	10,66 (5,25 – 16,0)	< 0,001
Трансферрин, г/л	2,42 (2,06 – 2,78)	1,95 (1,75 – 2,19)	< 0,001
Na ⁺ , ммоль/л	142,04 (140,45 – 143,5)	142,19 (140,7 – 144,1)	0,62
K ⁺ , ммоль/л	4,28 (3,99 – 4,6)	4,15 (3,82 – 4,51)	0,04
СРБ, мг/л	6,17 (1,18 – 10,3)	40,22 (9,29 – 66,74)	< 0,001
Фибриноген, г/л	3,79 (3,1 – 4,52)	5,32 (4,07 – 6,5)	< 0,001
ПТИ, %	92,42 (85,0 – 101,0)	88,98 (79,15 – 101,0)	0,14
ПТВ, с	12,23 (11,5 – 12,9)	12,54 (11,6 – 13,35)	0,07
МНО	1,07 (1,02 – 1,13)	1,1 (1,03 – 1,17)	0,19
АЧТВ, с	34,05 (31,5 – 36,7)	33,11 (30,0 – 35,6)	0,25
D-димер, нг/мл	239,85 (111,0 – 314,0)	592,1 (244,0 – 851,5)	< 0,001
Прокальцитонин, нг/мл	0,03 (0,02 – 0,05)	0,04 (0,03 – 0,06)	0,17

мочевина, ЛДГ, ферритин, Fe, трансферрин, K⁺, СРБ, фибриноген, D-димер.

Для дополнительной валидации гипотез о статистически значимых различиях в указанных группах осуществлена их проверка с использованием бутстрапирования и р-перестановочного теста, что

позволило выделить в качестве предикторов следующие признаки: SpO₂, RBC, HGB, EOS, LYMPH, NEUT, СОЭ, ЛДГ, ферритин, СРБ, D-димер. Визуальное представление данных по каждому из признаков представлено в виде boxplot-диаграмм (рис. 1).

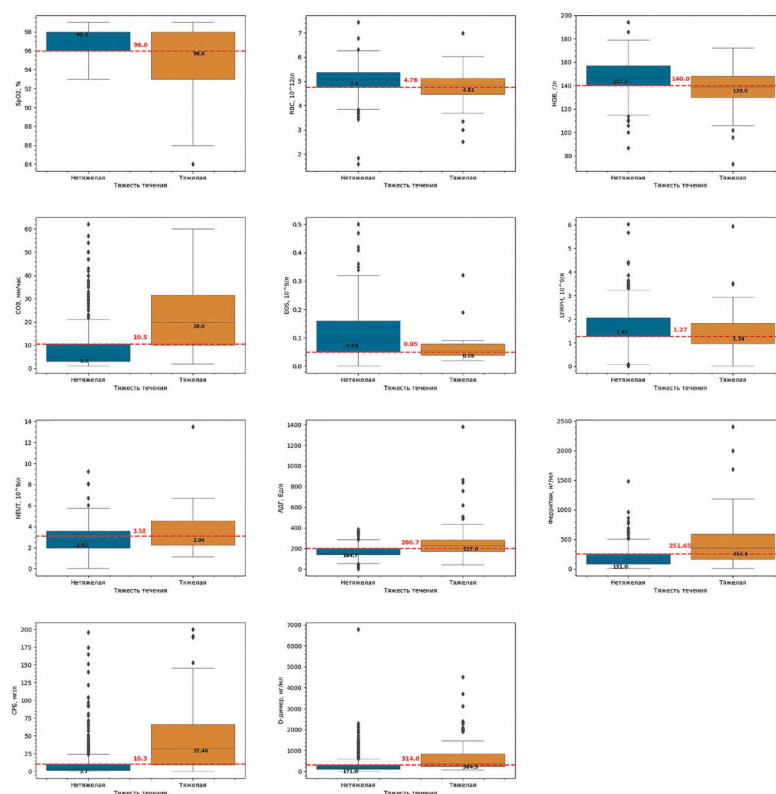


Рис. 1. Диаграммы размаха лабораторных показателей пациентов с SARS-CoV-2 инфекцией в группах «Нетяжелые» и «Тяжелые»

Признаки ЧДА, ЧСС, систолическое АД, диастолическое АД, Ht, глюкоза, общий белок, общий билирубин, альбумин, АЛТ, АСТ, КФК, мочевины, Fe, трансферрин, K^+ , фибриноген исключены из рассмотрения, поскольку не прошли валидацию на основе перестановочного теста, то есть статистически значимые различия этих признаков в 2 группах обусловлены особенностями собранной выборки из генеральной совокупности.

Таким образом, для ключевых прогностических показателей были определены следующие пороговые значения, представленные в таблице 2.

Таблица 2

Пороговые значения ключевых прогностических показателей на основе анализа диаграмм размаха

Прогностический показатель	Пороговое значение (ед. измерения)
Уровень сатурации	96%
Количество эритроцитов периферической крови	$4,76 \times 10^{12}/л$
Уровень гемоглобина	140 г/л
Скорость оседания эритроцитов	10,5 мм/ч
Абсолютные значения эозинофилов	$0,05 \times 10^9/л$
Абсолютные значения лимфоцитов	$1,27 \times 10^9/л$
Абсолютные значения нейтрофилов	$3,58 \times 10^9/л$
Уровень АДГ	200,7 ЕД/л
Уровень ферритина	251,65 нг/мл
Уровень С-реактивного белка	10,3 мг/литр
Уровень D-димера	314 нг/мл

Процесс разработки предсказательной модели тяжести течения SARS-CoV-2-инфекции на основе методов искусственного интеллекта состоял из 3 основных этапов, представленных на рисунке 2. На первом этапе осуществлен сбор данных историй болезни пациентов с лабораторно верифицированной SARS-CoV-2-инфекцией в СУБД SQLite, на основе анализа предметной области осуществлен отбор значимых признаков, сформированы обучающая и тестовая выборки, проведена проверка данных на непротиворечивость. На втором этапе выбран пул моделей машинного обучения и наборы их инициализирующих параметров, потенциально способных с высокой точностью решить поставленную задачу дихотомической классификации, проведено их обучение и оценка предсказательной способности. Характеристикой для оценки качества модели выбрана AUC-метрика, максимизация которой позволит максимально избежать ложноотрицательных и ложноположительных результатов работы классификатора. На третьем этапе модель с лучшими показателями

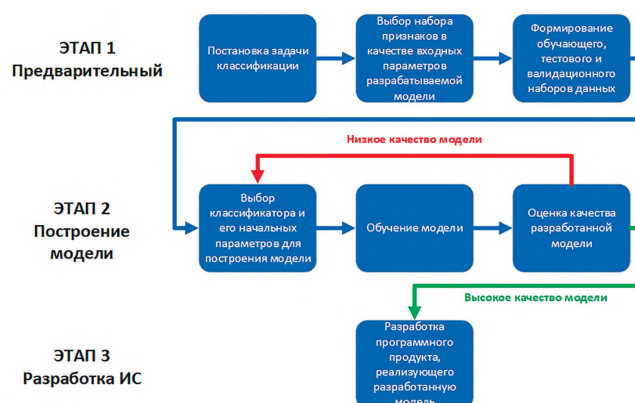


Рис. 2. Процесс разработки предсказательной модели тяжести течения SARS-CoV-2-инфекции на основе методов искусственного интеллекта

ROC-AUC-метрики для тестовой и валидационной выборок реализована в форме программного продукта.

Для формирования набора обучающих данных проведен дополнительный анализ собранных из историй болезни сведений о лабораторной характеристике заболевания у пациентов с инфекцией, вызванной SARS-CoV-2. В результате выявлено наличие пропусков в собранном наборе данных. В качестве критерия для исключения записи из набора данных обучающей выборки принят порог в 90% от всех исследуемых лабораторных параметров. В результате получен набор данных из 590 пациентов, из которых 90 (15,3%) пациентов с тяжелым течением инфекции, 500 (84,7%) пациентов с нетяжелым течением. Указанный набор данных можно считать репрезентативным и отражающим реальную структуру распределения по степени тяжести. Пропуски в данных были заполнены медианными значениями, а для признаков с высокой корреляцией использован метод k-ближайших соседей [22].

Полученный набор данных был разделен на обучающую и тестовую выборки в соотношении 70 к 30. В качестве метода кроссвалидации выбран k-fold-метод [23]. Для формирования валидационного набора данных, используемого только для оценки качества разрабатываемой модели собраны данные 40 пациентов с лабораторно верифицированной SARS-CoV-2-инфекцией за 2022 г., среди которых 50% имели тяжелое и 50% нетяжелое течение заболевания.

На втором этапе разработки информационной системы поддержки принятия врачебных решений о прогнозе тяжести течения заболевания у пациентов с SARS-CoV-2-инфекцией, с учетом имеющихся данных историй болезни, принято решение об использовании методов глубокого обучения для построения предсказательной модели, лежащей

в основе информационной системы. Выбор методов глубокого обучения обусловлен спецификой решаемой задачи: полученные 11 предикторов тяжелого течения заболевания имеют неявные статистические зависимости и различную значимость при принятии решения о прогнозе тяжести течения заболевания. Методы глубокого обучения позволяют моделировать абстракции высокого уровня с использованием архитектур из нескольких нелинейных преобразований за счет использования функций активации нейронов не только на выходном, но и на скрытых слоях нейронной сети. При этом непосредственно архитектура нейронной сети подбирается экспериментально. Поскольку в исследовании решалась задача дихотомической классификации, то для ее решения использована многослойная полносвязная нейронная сеть с 11 входами, то есть признаками, отобранными в качестве предикторов тяжелого течения. Каждый введенный предиктор соответствует определенному нейрону входного слоя и используется им в качестве входных данных для функции активации. Затем происходит последовательное вычисление функций активации скрытых слоев, результаты которого поступают на функции активации выходного слоя. Выходное значение функции активации выходного слоя и есть результат работы предсказательной модели. Обобщенная схема многослойной полносвязной нейронной сети представлена на рисунке 3.

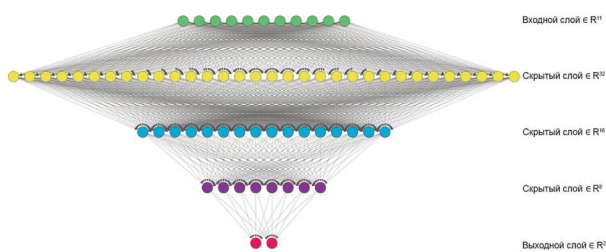


Рис. 3. Обобщенная схема многослойной полносвязной нейронной сети

Оценка полученных в ходе экспериментального подбора архитектуры нейронных сетей предсказательных моделей проводилась с использованием матрицы ошибок и ROC-кривых [24]. При этом если результаты оценки модели оказывались неудовлетворительными, то применялось изменение начальных параметров классификатора или менялся метод классификации в целом. Предварительная оценка качества модели проводилась на тестовой выборке: если качество модели на ней было удовлетворительным, то она проверялась на валидационной, которая также не участвовала в процессе обучения и была сформирована из данных историй болезни пациентов за 2022 г.

Результаты ROC-анализа для тестовой и валидационной выборок представлены на рисунке 4. Отметим, что для удобства восприятия в качестве координатных осей при построении ROC-кривых использованы чувствительность и специфичность, поэтому значения оси абсцисс начинаются с 1 и заканчиваются 0. График модели, использующей случайное равновероятное предсказание, представлен пунктирной линией.

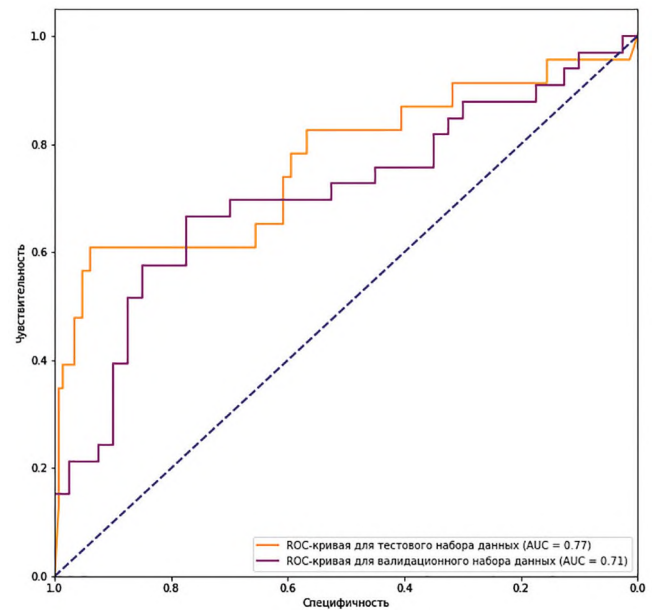


Рис. 4. Оценка качества разработанной модели с использованием ROC-кривых и метрики AUC

Для валидационного набора данных точность, чувствительность и специфичность составили 92%, 91% и 74%, для тестового — 88%, 92% и 87% соответственно, значение AUC-метрики — 0,82 и 0,94 соответственно. Поскольку на обоих наборах значение AUC-метрики превышает 0,8, то можно говорить о высоком качестве предсказательной способности разработанной модели.

Программа для ЭВМ «CASP(ER)» состоит из 3 основных частей: базы данных, в которой хранятся данные историй болезни пациентов и результаты предсказаний, серверной части приложения, разработанной с использованием фреймворка FASTapi, в которой реализовано занесение записей в базу данных и предварительные вычисления, необходимые для работы разработанного классификатора, и клиентской части в виде WEB-приложения.

Интерфейс основной вкладки программы для работы с предсказательной моделью представлен на рисунке 5.

Таким образом, проведенная оценка качества разработанной предсказательной модели на основе многослойной полносвязной искусственной нейронной сети позволяет говорить о высоком

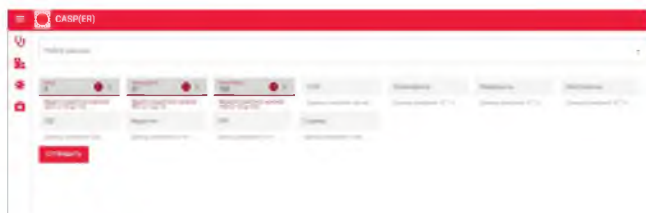


Рис. 5. Главное окно ввода данных предсказательной модели

качестве построенного на ее основе классификатора, который реализован в форме программы для ЭВМ «CASP(ER) — Covid-19 assisted severity prediction (early reliable)» [25].

Обсуждение

Выделенные в нашем исследовании лабораторные предикторы тяжелого течения SARS-CoV-2-инфекции соотносятся с данными других авторов — среди наиболее значимых характеристик при принятии решений отмечены эритроциты и лейкоцитарные показатели общеклинического анализа крови, неспецифические маркеры воспаления, индукторы цитолиза и показатели коагулограммы [26, 27]. При этом в указанных работах не приведены данные о пороговых значениях показателей, что значительно утяжеляет оценку конкретного клинического случая. Также следует отметить, что для более детальной оценки вклада каждого из выделенных параметров в выработку решения предсказательной моделью необходимо установить значимости каждого из предикторов тяжелого течения, выбранных при сравнении групп пациентов нетяжелого и тяжелого течения заболевания. В работах, где методы машинного обучения применяются для диагностики и раннего предсказания неблагоприятного течения SARS-CoV-2 инфекции, расчётные показатели чувствительности варьируют от 60 до 100% [28, 29]. При этом следует отметить, что высокие показатели чувствительности и специфичности (более 90%) могут быть обусловлены переобученностью модели. Одним из преимуществ нашей модели является построение на данных лабораторных анализов, полученных у пациентов в момент госпитализации (длительность заболевания не превышала 3 дней), что исключает возможность искажения точности прогноза в связи с изменением лабораторных показателей в разные периоды инфекционного процесса.

Нельзя исключать возможное влияние особенностей циркулирующего штамма на характер и выраженность клинических и лабораторных изменений у пациентов из тестовой и валидационной выборок, что, в свою очередь, могло повлиять на расчетную точность прогноза разработанной

модели. Как видно из рисунка 4, качество предсказания на валидационной выборке ниже (точность, чувствительность и специфичность составили 92%, 91% и 74% соответственно по сравнению с 88%, 92% и 87% для тестовой выборки), что, вероятно, обусловлено особенностями циркулирующих штаммов выбранного сезона. Отметим, что, исходя из специфики рассматриваемой задачи медицинской диагностики, при окончательном выборе нейросетевой модели для реализации приоритет отдавался минимизации ошибок 2 рода, то есть минимизации количества случаев отнесения пациента, склонному к тяжелому течению, к нетяжелым.

Однако несмотря на описанные в литературе различия в частоте встречаемости некоторых клинических проявлений COVID-19 у пациентов, переносивших заболевания, вызванные вариантами альфа, дельта и омикрон, вопрос о влиянии конкретного штамма на частоту развития тяжелого течения заболевания остается открытым: ключевую роль в развитии летального исхода на сегодняшний день отводят возрасту и особенностям преморбидного фона пациента, а сценарии прогрессирования заболевания при развитии тяжелого течения вне зависимости от выделенного штамма признаны идентичными. Также следует отметить, что данные о различиях лабораторной картины при развитии заболевания, вызванного различными вариантами возбудителя в отдельных возрастных группах, в доступной литературе не представлены. Еще одним фактором, обуславливающим особенности течения инфекции, вызванной SARS-CoV-2, является вакцинация: установлено, что вакцинированные пациенты статистически значимо реже нуждаются в госпитализации, кислородной поддержке, а также имеют меньшую продолжительность заболевания по сравнению с невакцинированными. При этом при развитии прорывной инфекции тяжелого течения риск развития летального исхода у рассматриваемых групп значимо не отличается [30].

Заключение

У пациентов с легкой и средней степенью тяжести инфекции, вызванной SARS-CoV-2, периферическая кислородная сатурация, содержание эритроцитов, гемоглобина, общего белка, альбумина, уровень гематокрита, сывороточного железа, трансферрина, а также абсолютное число эозинофилов и лимфоцитов периферической крови на момент поступления в стационар были статистически значимо выше, чем у пациентов с тяжелой степенью тяжести заболевания ($p < 0,001$). Значения абсолютного числа нейтрофилов, СОЭ, глюкозы, АЛТ, АСТ, КФК, мочевины, ЛДГ, ферритина, СРБ, фибриногена, D-димера, ЧДД, ЧСС, уровень артериального давления в группе пациентов легкой

и средней степени тяжести были статистически значимо ниже, чем в группе тяжелых пациентов ($p < 0,001$).

Выделено 11 показателей, являющихся предикторами тяжелого течения, и их пороговые значения: уровень периферической кислородной сатурации (с пороговым значением 96%), количество эритроцитов периферической крови (пороговое значение $4,76 \times 10^{12}/л$), уровень гемоглобина (пороговое значение 140 г/л), СОЭ (пороговое значение 10,5 мм/ч), абсолютные значения эозинофилов (пороговое значение $0,05 \times 10^9/л$), абсолютное число лимфоцитов (пороговое значение $1,27 \times 10^9/л$), абсолютное значение нейтрофилов (пороговое значение $3,58 \times 10^9/л$), уровень ЛДГ (пороговое значение 200,7 Ед/л), значение ферритина (пороговое значение 251,65 нг/мл), уровень С-реактивного белка (пороговое значение 10,3 мг/литр), содержание D-димера (пороговое значение 314 нг/мл) периферической крови.

Разработанная с использованием методов искусственного интеллекта предсказательная модель с высоким качеством (на валидационном наборе данных точность, чувствительность, специфичность, значение AUC-метрики составили 92%, 91%, 74%, 0,82, на тестовом наборе — 88%, 92%, 87%, 0,94 соответственно) способна предсказать развитие тяжелого течения SARS-CoV-2-инфекции у пациентов молодого возраста.

Литература

1. Crevier, D. AI: the tumultuous history of the search for artificial intelligence / D. Crevier — DOI: 10.3233/AIC-1995-8108. — Текст: электронный // Basic Books, Inc. — 1993. — URL: https://www.researchgate.net/publication/233820788_AI_The_Tumultuous_History_of_the_Search_for_Artificial_Intelligence (дата обращения: 05.09.2022).
2. Mitchell, T. M. Machine learning / T. M. Mitchell — DOI: — Текст: электронный // Burr Ridge, IL: McGraw Hill. — 1997. — №45(37). — P. 870-877. — URL: <https://www.cin.ufpe.br/~cavmj/Machine%20-%20Learning%20-%20Tom%20Mitchell.pdf> (дата обращения: 05.09.2022).
3. Shortliffe, E. Computer-based medical consultations: MYCIN / E. Shortliffe — DOI: 10.1016/B978-0-444-00179-5.X5001-X. — Текст: электронный // Elsevier. — 2012. — №2. — URL: https://books.google.ru/books/about/Computer_Based_Medical_Consultations_MYC.html?id=i9QXugPQw6C&redir_esc=y (дата обращения: 05.09.2022).
4. Hinton, G. Deep learning—a technology with the potential to transform health care / G. Hinton — DOI: 10.1001/jama.2018.11100. — Текст: электронный // Jama. — 2018. — №320(11). — P. 1101-1102. — URL: <https://jamanetwork.com/journals/jama/article-abstract/2701666> (дата обращения: 05.09.2022).
5. Силкина, У.И. Дискриминантный анализ параметров вариационной пульсометрии / У.И. Силкина, В.А. Баландин // Russian Technological Journal. — 2020. — № 8(3). — С. 81–91. — <https://doi.org/10.32362/2500-316X-2020-8-3-81-91>
6. Rawson, T.M. A systematic review of clinical decision support systems for antimicrobial management: are we failing to investigate these interventions appropriately / T. M. Rawson, L. S. Moore, B. Hernandez [et. al.]. — DOI: 10.1016/j.cmi.2017.02.028. — Текст: электронный // Clinical Microbiology and Infection. — 2017. — №23(8). — P. 524-532. — URL: [https://www.clinicalmicrobiologyandinfection.com/article/S1198-743X\(17\)30125-8/fulltext](https://www.clinicalmicrobiologyandinfection.com/article/S1198-743X(17)30125-8/fulltext) (дата обращения: 05.09.2022).
7. Martínez-Aguero, S. Machine learning techniques to identify antimicrobial resistance in the intensive care unit / S. Martínez-Aguero, I. Mora-Jiménez, J. Llerida-García [et. al.]. — DOI: 10.3390/e21060603. — Текст: электронный // Entropy. — 2019. — №21(6). — URL: <https://www.mdpi.com/1099-4300/21/6/603> (дата обращения: 05.09.2022).
8. Hartvigsen, T. Early Prediction of MRSA Infections using Electronic Health Records / T. Hartvigsen, C. Sen, S. Brownell [et. al.]. — DOI: 10.5220/0006599601560167. — Текст: электронный // HEALTHINF. — 2018. — P. 156-167. — URL: <https://www.scitepress.org/Link.aspx?doi=10.5220/0006599601560167> (дата обращения: 05.09.2022).
9. Revuelta-Zamorano, P. Prediction of healthcare associated infections in an intensive care unit using machine learning and big data tools / P. Revuelta-Zamorano, A. Sánchez, J.L. Rojo-Álvarez [et. al.]. — DOI: 10.1007/978-3-319-32703-7_163. — Текст: электронный // XIV Mediterranean conference on medical and biological engineering and computing, Springer International Publishing. — 2016. — №57. — P. 840–845. — URL: https://link.springer.com/chapter/10.1007/978-3-319-32703-7_163 (дата обращения: 05.09.2022).
10. Hernandez, B. Supervised learning for infection risk inference using pathology data / B. Hernandez, P. Herrero, T. M. Rawson [et. al.]. — DOI: 10.1186/s12911-017-0550-1. — Текст: электронный // BMC medical informatics and decision making. — 2017. — №17(168). — P. 1-12. — URL: <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-017-0550-1> (дата обращения: 05.09.2022).
11. Agbehadj, I. E. Review of big data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing / I. E. Agbehadj, B. O. Awuzie, A. B. Ngowi [et. al.]. — DOI: 10.3390/ijerph17155330. — Текст: электронный // Int J Environ Res Publ Health. — 2020. — №17(15). — P. 5330. — URL: <https://www.mdpi.com/1660-4601/17/15/5330> (дата обращения: 05.09.2022).
12. Naudé, W. Artificial intelligence vs Covid-19: limitations, constraints and pitfalls / W. Naudé — DOI: 10.1007/s00146-020-00978-0. — Текст: электронный // AI Soc. — 2020. — №35. — P. 761–765. — URL: <https://link.springer.com/article/10.1007/s00146-020-00978-0#citeas> (дата обращения: 05.09.2022).
13. Vaid, A. Machine learning to predict mortality and critical events in covid-19 positive New York city patients: a cohort study / A. Vaid, S. Somani, A. J. Russak [et. al.]. — DOI: 10.2196/24018. — Текст: электронный // J Med Internet Res. — 2020. — №22(11). — URL: <https://www.jmir.org/2020/11/e24018> (дата обращения: 05.09.2022).
14. Rechtman, E. Vital signs assessed in initial clinical encounters predict covid-19 mortality in an NYC hospital system / E. Rechtman, P. Curtin, E. Navarro [et. al.]. — DOI: 10.1038/s41598-020-78392-1. — Текст: электронный // Sci Rep. — 2020. — №10(21545). — P. 1-6. — URL: <https://www.nature.com/articles/s41598-020-78392-1> (дата обращения: 05.09.2022).
15. Cabitza, F. Development, evaluation, and validation of machine learning models for covid-19 detection based on routine blood tests / F. Cabitza, A. Campagner, D. Ferrari [et. al.]. — DOI: 10.1515/cclm-2020-1294. — Текст: электронный // Clin Chem Lab Med. — 2020. — №59(2). — P. 421-431. — URL:

<https://www.degruyter.com/document/doi/10.1515/cclm-2020-1294/html> (дата обращения: 05.09.2022).

16. Goodman-Meza, D. A machine learning algorithm to increase COVID-19 inpatient diagnostic capacity / D. Goodman-Meza, A. Rudas, J. N. Chiang [et. al.]. — DOI: 10.1371/journal.pone.0239474. — Текст: электронный // PLoS One. — 2020. — №15(9). — URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0239474> (дата обращения: 05.09.2022).

17. AlJame, M. Ensemble learning model for diagnosing COVID-19 from routine blood tests / M. AlJame, I. Ahmad, A. Imtiaz [et. al.]. — DOI: 10.1016/j.imu.2020.100449. — Текст: электронный // Inform Med Unlocked. — 2020. — №21(100449). — URL: <https://www.sciencedirect.com/science/article/pii/S2352914820305992?via%3Dihub> (дата обращения: 05.09.2022).

18. Поляков, А.С. Прогностическое значение некоторых гематологических синдромов при инфекции, вызванной SARS-CoV-2K. / А.С. Поляков [и др.] — DOI 10.34883/PI.2020.6.2.001. — EDN IAVLYV — Текст: электронный // Гематология. Трансфузиология. Восточная Европа. — 2020. — Т. 6, № 2. — С. 161–171. — URL: https://www.elibrary.ru/download/elibrary_43065813_69549691.pdf (дата обращения: 05.09.2022).

19. Национальный проект «Здравоохранение». Федеральный проект «Создание единого цифрового контура в здравоохранении на основе единой государственной информационной системы в сфере здравоохранения (ЕГИСЗ)», утв. Министерством здравоохранения России 16.10.2019

20. Dyussenbayev, A. Age Periods Of Human Life / A. Dyussenbayev. — DOI: 10.14738/assrj.46.2924_4(6)_258-263. — Текст: электронный // Advances in Social Sciences Research Journal. — 2017. — №4(6). — URL: <https://journals.scholarpublishing.org/index.php/ASSRJ/article/view/2924>. (дата обращения: 05.09.2022).

21. Временные методические рекомендации по профилактике, диагностике и лечению новой коронавирусной инфекции (COVID-19) версия 16 (URL: https://static-0.minzdrav.gov.ru/system/attachments/attaches/000/060/193/original/%D0%92%D0%9C%D0%A0_COVID-19_V16.pdf) (дата обращения: 05.09.2022).

22. Kuhn, M. Feature Engineering and Selection: A Practical Approach for Predictive Models / M. Kuhn, K. Johnson. — Chapman & Hall/CRC, 2019. — 310 с.

23. Воронцов, К.В. Машинное обучение (курс лекций) / К.В. Воронцов. — URL: http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_%28курс_лекций%2C_К.В.Воронцов%29 (дата обращения 20.08.2022)

24. Брюс П. Практическая статистика для специалистов Data Science. 2-е изд., перераб.и доп. / П. Брюс, Э. Брюс, П. Гедек. — O'Reilly, 2021. — 352 с.

25. Свидетельство о государственной регистрации программы для ЭВМ No2022665747. Российская Федерация. «CASP(ER) — Covid-19 assisted severity prediction (early reliable)» / К. Касьяненко; заявитель и правообладатель Касьяненко Кристина. — No2022665747; заявка 06.08.2022; зарегистр. 19.08.2022; опублик. 19.08.2022, Бюл. No 8. — 1 с.

26. Bayat, V. A severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) prediction model from standard laboratory tests / V. Bayat, S. Phelps, R. Ryono [et. al.]. — DOI: 10.1093/cid/ciaa1175. — Текст: электронный // Clin Infect Dis. — 2020. — №73(9). — P. 2901-2907. — URL: <https://academic.oup.com/cid/article/73/9/e2901/5891814?login=false> (дата обращения: 05.09.2022).

27. Kukar, M. Covid-19 diagnosis by routine blood tests using machine learning / M. Kukar, G. Gunčar, T. Vovko [et. al.].

— DOI: 10.1038/s41598-021-90265-9. — Текст: электронный // Sci Rep. — 2021. — №11(10738). — URL: <https://www.nature.com/articles/s41598-021-90265-9#citeas> (дата обращения: 05.09.2022).

28. Wu, J. Rapid and accurate identification of covid-19 infection through machine learning based on clinical available blood test results / J. Wu, P. Zhang, L. Zhang [et. al.]. — DOI: 10.1101/2020.04.02.20051136. — Текст: электронный // medRxiv. — 2020. — URL: <https://www.medrxiv.org/content/10.1101/2020.04.02.20051136v1> (дата обращения: 05.09.2022).

29. Brinati, D. Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study / D. Brinati, A. Campagner, D. Ferrari [et. al.]. — DOI: 10.1007/s10916-020-01597-4. — Текст: электронный // J Med Syst. — 2020. — №44(135). — URL: <https://link.springer.com/article/10.1007/s10916-020-01597-4> (дата обращения: 05.09.2022).

30. Butt AA. Rate and risk factors for severe/critical disease among fully vaccinated persons with breakthrough severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in a high-risk national population / A.A. Butt, P. Yan, O.S. Shaikh [et al.]. — DOI: 10.1093/cid/ciab1023. — Текст: электронный // Clin Infect Dis. — 2021. — №75(1). — URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8689859/> (дата обращения: 05.09.2022)

References

1. Crevier, D. AI: the tumultuous history of the search for artificial intelligence / D. Crevier — DOI: 10.3233/AIC-1995-8108. — Текст: электронный // Basic Books, Inc. — 1993. — URL: https://www.researchgate.net/publication/233820788_AI_The_Tumultuous_History_of_the_Search_for_Artificial_Intelligence (data obrashheniya: 05.09.2022).

2. Mitchell, T. M. Machine learning / T. M. Mitchell — DOI: — Текст: электронный // Burr Ridge, IL: McGraw Hill. — 1997. — №45(37). — P. 870-877. — URL: <https://www.cin.ufpe.br/~cavmj/Machine%20-%20Learning%20-%20Tom%20Mitchell.pdf> (data obrashheniya: 05.09.2022).

3. Shortliffe, E. Computer-based medical consultations: MYCIN / E. Shortliffe — DOI: 10.1016/B978-0-444-00179-5.X5001-X. — Текст: электронный // Elsevier. — 2012. — №2. — URL: https://books.google.ru/books/about/Computer_Based_Medical_Consultations_MYCIN.html?id=i9QXugPQw6oC&redir_esc=y (data obrashheniya: 05.09.2022).

4. Hinton, G. Deep learning—a technology with the potential to transform health care / G. Hinton — DOI: 10.1001/jama.2018.11100. — Текст: электронный // Jama. — 2018. — №320(11). — P. 1101-1102. — URL: <https://jamanetwork.com/journals/jama/article-abstract/2701666> (data obrashheniya: 05.09.2022).

5. Silkina, U.I. Discriminant analysis of variational pulsemetry parameters / U.I. Silkina, V.A. Balandin // Russian Technological Journal. — 2020. — №8(3). — P. 81-91. (In Russ.) <https://doi.org/10.32362/2500-316X-2020-8-3-81-91>

6. Rawson, T. M. A systematic review of clinical decision support systems for antimicrobial management: are we failing to investigate these interventions appropriately / T. M. Rawson, L. S. Moore, B. Hernandez [et. al.]. — DOI: 10.1016/j.cmi.2017.02.028. — Текст: электронный // Clinical Microbiology and Infection. — 2017. — №23(8). — P. 524-532. — URL: [https://www.clinicalmicrobiologyandinfection.com/article/S1198-743X\(17\)30125-8/fulltext](https://www.clinicalmicrobiologyandinfection.com/article/S1198-743X(17)30125-8/fulltext) (data obrashheniya: 05.09.2022).

7. Martínez-Agüero, S. Machine learning techniques to identify antimicrobial resistance in the intensive care unit / S.

- Martínez-Aguero, I. Mora-Jiménez, J. Llerida-García [et. al.]. — DOI: 10.3390/e21060603. — Текст: электронный // *Entropy*. — 2019. — №21(6). — URL: <https://www.mdpi.com/1099-4300/21/6/603> (дата обращения: 05.09.2022).
8. Hartvigsen, T. Early Prediction of MRSA Infections using Electronic Health Records / T. Hartvigsen, C. Sen, S. Brownell [et. al.]. — DOI: 10.5220/0006599601560167. — Текст: электронный // *HEALTHINF*. — 2018. — P. 156-167. — URL: <https://www.scitepress.org/Link.aspx?doi=10.5220/0006599601560167> (data obrashheniya: 05.09.2022).
9. Revuelta-Zamorano, P. Prediction of healthcare associated infections in an intensive care unit using machine learning and big data tools / P. Revuelta-Zamorano, A. Sánchez, J. L. Rojo-Álvarez [et. al.]. — DOI: 10.1007/978-3-319-32703-7_163. — Текст: электронный // XIV Mediterranean conference on medical and biological engineering and computing, Springer International Publishing. — 2016. — №57. — P. 840–845. — URL: https://link.springer.com/chapter/10.1007/978-3-319-32703-7_163 (data obrashheniya: 05.09.2022).
10. Hernandez, B. Supervised learning for infection risk inference using pathology data / B. Hernandez, P. Herrero, T. M. Rawson [et. al.]. — DOI: 10.1186/s12911-017-0550-1. — Текст: электронный // *BMC medical informatics and decision making*. — 2017. — №17(168). — P. 1-12. — URL: <https://bmc-medinformdecismak.biomedcentral.com/articles/10.1186/s12911-017-0550-1> (data obrashheniya: 05.09.2022).
11. Agbehadj, I. E. Review of big data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing / I. E. Agbehadj, B. O. Awuzie, A. B. Ngowi [et. al.]. — DOI: 10.3390/ijerph17155330. — Текст: электронный // *Int J Environ Res Publ Health*. — 2020. — №17(15). — P. 5330. — URL: <https://www.mdpi.com/1660-4601/17/15/5330> (data obrashheniya: 05.09.2022).
12. Naudé, W. Artificial intelligence vs Covid-19: limitations, constraints and pitfalls / W. Naudé — DOI: 10.1007/s00146-020-00978-0. — Текст: электронный // *AI Soc*. — 2020. — №35. — P. 761–765. — URL: <https://link.springer.com/article/10.1007/s00146-020-00978-0#citeas> (data obrashheniya: 05.09.2022).
13. Vaid, A. Machine learning to predict mortality and critical events in covid-19 positive New York city patients: a cohort study / A. Vaid, S. Somani, A. J. Russak [et. al.]. — DOI: 10.2196/24018. — Текст: электронный // *J Med Internet Res*. — 2020. — №22(11). — URL: <https://www.jmir.org/2020/11/e24018> (data obrashheniya: 05.09.2022).
14. Rechtman, E. Vital signs assessed in initial clinical encounters predict covid-19 mortality in an NYC hospital system / E. Rechtman, P. Curtin, E. Navarro [et. al.]. — DOI: 10.1038/s41598-020-78392-1. — Текст: электронный // *Sci Rep*. — 2020. — №10(21545). — P. 1-6. — URL: <https://www.nature.com/articles/s41598-020-78392-1> (data obrashheniya: 05.09.2022).
15. Cabitza, F. Development, evaluation, and validation of machine learning models for covid-19 detection based on routine blood tests / F. Cabitza, A. Campagner, D. Ferrari [et. al.]. — DOI: 10.1515/cclm-2020-1294. — Текст: электронный // *Clin Chem Lab Med*. — 2020. — №59(2). — P. 421-431. — URL: <https://www.degruyter.com/document/doi/10.1515/cclm-2020-1294/html> (data obrashheniya: 05.09.2022).
16. Goodman-Meza, D. A machine learning algorithm to increase COVID-19 inpatient diagnostic capacity / D. Goodman-Meza, A. Rudas, J. N. Chiang [et. al.]. — DOI: 10.1371/journal.pone.0239474. — Текст: электронный // *PloS One*. — 2020. — №15(9). — URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0239474> (data obrashheniya: 05.09.2022).
17. AlJame, M. Ensemble learning model for diagnosing COVID-19 from routine blood tests / M. AlJame, I. Ahmad, A. Imtiaz [et. al.]. — DOI: 10.1016/j.imu.2020.100449. — Текст: электронный // *Inform Med Unlocked*. — 2020. — №21(100449). — URL: <https://www.sciencedirect.com/science/article/pii/S2352914820305992?via%3Dihub> (data obrashheniya: 05.09.2022).
18. Polyakov, A.S. Prognostic Significance of Some Hematological Syndromes in SARS-CoV-2 Infection / A.S. Polyakov, K.V. Kozlov // *Hematology. Transfusiology*. Eastern Europe. — 2020. — №6(2). — P. 161-171. (In Russ.)
19. Nacional'nyj proekt «Zdravoohranenie». Federal'nyj proekt «Sozdanie edinogo cifrovogo kontura v zdravoohranении na osnove edinoj gosudarstvennoj informacionnoj sistemy v sfere zdravoohraneniya (EGISZ)», utv. Ministerstvom zdravoohraneniya Rossii 16.10.2019
20. Dyussenbayev, A. (2017). Age Periods Of Human Life / A. Dyussenbayev. — DOI: 10.14738/assrj.46.2924 4(6) 258-263. — Текст: электронный // *Advances in Social Sciences Research Journal*. — 2017. — №4(6). — URL: <file:///D:/Downloads/2924-Article%20Text-7575-1-10-20170401.pdf> data obrashheniya: 05.09.2022).
21. Vremennye metodicheskie rekomendacii po profilaktike, diagnostike i lecheniju novoj koronavirusnoj infekcii (COVID-19) versija 16 (URL: https://static-0.minzdrav.gov.ru/system/attachments/attaches/000/060/193/original/%D0%92%D0%9C%D0%A0_COVID-19_V16.pdf) (data obrashheniya: 05.09.2022).
22. Kuhn, M. Feature Engineering and Selection: A Practical Approach for Predictive Models / M. Kuhn, K. Johnson. — Chapman & Hall/CRC, 2019. — 310 p.
23. Voroncov K.V. Mashinnoe obuchenie (kurs lekciij) / K.V. Voroncov URL: http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_%28курс_лекций%2C_K.B.Воронцов%29 (data obrashheniya: 05.09.2022).
24. Bruce, P. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python / P. Bruce, A. Bruce, P. Gedeck. — O'Reilly, 2020. — 368 p.
25. Svidetel'stvo o gosudarstvennoj registracii programmy dlja JeVM No2022665747. Rossijskaja Federacija. «CASP(ER) — Covid-19 assisted severity prediction (early reliable)» / K. Kas'janenko; zajavitel' i pravooobladatel' Kas'janenko Kristina. — No2022665747; zajavka 06.08.2022; zaregistr. 19.08.2022; opubl. 19.08.2022, Bjul. No 8. — 1 s.
26. Bayat, V. A severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) prediction model from standard laboratory tests / V. Bayat, S. Phelps, R. Ryono [et. al.]. — DOI: 10.1093/cid/ciaa1175. — Текст: электронный // *Clin Infect Dis*. — 2020. — №73(9). — P. 2901-2907. — URL: <https://academic.oup.com/cid/article/73/9/e2901/5891814?login=false> (data obrashheniya: 05.09.2022).
27. Kukar, M. Covid-19 diagnosis by routine blood tests using machine learning / M. Kukar, G. Gunčar, T. Vovko [et. al.]. — DOI: 10.1038/s41598-021-90265-9. — Текст: электронный // *Sci Rep*. — 2021. — №11(10738). — URL: <https://www.nature.com/articles/s41598-021-90265-9#citeas> (data obrashheniya: 05.09.2022).
28. Wu, J. Rapid and accurate identification of covid-19 infection through machine learning based on clinical available blood test results / J. Wu, P. Zhang, L. Zhang [et. al.]. — DOI: 10.1101/2020.04.02.20051136. — Текст: электронный // *medRxiv*. — 2020. — URL: <https://www.medrxiv.org/content/10.1101/2020.04.02.20051136v1> (data obrashheniya: 05.09.2022).
29. Brinati, D. Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study / D. Brinati, A. Campagner, D. Ferrari [et. al.]. — DOI: 10.1007/

s10916-020-01597-4. — Текст: электронный // J Med Syst. — 2020. — №44(135). — URL: <https://link.springer.com/article/10.1007/s10916-020-01597-4> (data obrashhenija: 05.09.2022).

30. Butt AA. Rate and risk factors for severe/critical disease among fully vaccinated persons with breakthrough severe acute

respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in a high-risk national population / A.A. Butt, P. Yan, O.S. Shaikh [et al]. — DOI: 10.1093/cid/ciab1023. — Текст: электронный // Clin Infect Dis. — 2021. — №75(1). — URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8689859/> (data obrashhenija: 05.09.2022).

Авторский коллектив:

Касьяненко Кристина Валерьевна — преподаватель кафедры инфекционных болезней (с курсом медицинской паразитологии и тропических заболеваний) Военно-медицинской академии им. С.М. Кирова; тел.: +7-911-262-06-33, e-mail: dr.snegur@gmail.com

Козлов Константин Вадимович — профессор кафедры инфекционных болезней (с курсом медицинской паразитологии и тропических заболеваний) Военно-медицинской академии им. С.М. Кирова, д.м.н., доцент; тел.: +7-921-657-27-49, e-mail: kosttiak@mail.ru

Жданов Константин Валерьевич — начальник кафедры инфекционных болезней (с курсом медицинской паразитологии и тропических заболеваний) Военно-медицинской академии им. С.М. Кирова, д.м.н., профессор, член-корреспондент РАН; тел.: +7-921-939-82-95

Лапиков Игорь Игоревич — доцент кафедры информационного противоборства Института кибербезопасности и цифровых технологий МИРЭА — Российского технологического университета, к.т.н.; тел.: 8(499)215-65-65, e-mail: lapikov.i.i@yandex.ru

Беликов Владимир Вячеславович — доцент кафедры информационного противоборства Института кибербезопасности и цифровых технологий МИРЭА — Российского технологического университета, к.воен.н., доцент; тел.: 8(499)215-65-65, e-mail: belikov@mirea.ru