

Научная статья

## ИСПОЛЬЗОВАНИЕ ИНДУКТИВНОГО ТИПА МАШИННОГО ОБУЧЕНИЯ ДЛЯ ВЫЯВЛЕНИЯ ФАКТОРОВ РИСКА ИНФИЦИРОВАНИЯ МЕДИЦИНСКИХ РАБОТНИКОВ ВЫСОКОКОНТАГИОЗНЫМИ ВИРУСАМИ (НА МОДЕЛИ COVID-19)

И.А. Егоров<sup>1</sup>, С.С. Смирнова<sup>1,2</sup>, А.В. Семенов<sup>1,2</sup>

<sup>1</sup>Федеральный научно-исследовательский институт вирусных инфекций «Виром», Российская Федерация, 620030, г. Екатеринбург, ул. Летняя, 23

<sup>2</sup>Уральский государственный медицинский университет, Российская Федерация, 620028, г. Екатеринбург, ул. Репина, 3

*Эпидемическое и пандемическое распространение высококонтагиозных вирусов (SARS-CoV, вирус гриппа А, вирус Эбола, MERS-CoV, SARS-CoV-2) – тенденция прошедших двух десятилетий XXI в.*

*Преимущественное воздействие биологического фактора производственной среды на медицинских работников определяет высокий профессиональный риск заражения, тяжелого течения болезни и летального исхода. Интеллектуальная обработка эпидемиологических данных, основанная на алгоритмах машинного обучения, находит успешное применение в эпидемиологической практике для выявления факторов, способствующих инфицированию (предикторов), у различных контингентов риска.*

*В ходе исследования проведен интеллектуальный анализ базы данных, сформированной по результатам анкетного опроса 1312 медицинских работников. Всего обучено 6912 моделей машинного обучения. Установлено, что инфицированию SARS-CoV-2 способствовало оказание медицинской помощи пациенту с COVID-19, использование полного комплекта СИЗ после непосредственного контакта с больным COVID-19, прямой контакт с предметами внешней (больничной) среды, вакцинация от COVID-19 после непосредственного контакта с больным COVID-19, выполнение функций младшего медицинского персонала (уборщиков), присутствие при проведении процедур, генерирующих аэрозоль.*

*Установлено четыре группы предикторов, определяющих инфицирование SARS-CoV-2 медицинских работников, – контакт с больным COVID-19 и предметами окружающей его среды, качество и комплексность СИЗ, профессиональная принадлежность медицинских работников и показатели ИМТ. Наличие одного предиктора установлено у 56,2 % медицинских работников, двух – у 19,2 %, трех – у 16,4 %, четырех – у 5,5 %, пяти – у 2,7 %.*

*Таким образом, интеллектуальная обработка эпидемиологических данных является современным этапом эпидемиологического анализа. Применение методов машинного обучения позволяет провести многофакторную оценку рисков инфицирования медицинских работников SARS-CoV-2, выявить и достоверно оценить наиболее значимые предикторы. Гибкость архитектуры интеллектуального анализа данных позволяет проводить как корректировку изучаемой модели, так и динамически дополнять сформированную базу новыми данными, улавливать изменения эпидемиологической ситуации и проводить актуальные профилактические и противоэпидемические мероприятия.*

**Ключевые слова:** интеллектуальный анализ данных, искусственный интеллект, машинное обучение, риск-ориентированный подход, профессиональные предикторы инфицирования, высококонтагиозные вирусы, SARS-CoV-2, медицинские работники.

© Егоров И.А., Смирнова С.С., Семенов А.В., 2024

**Егоров Иван Андреевич** – научный сотрудник Урало-Сибирского научно-методического центра по профилактике инфекций, связанных с оказанием медицинской помощи (e-mail: egorov\_ia@niivirom.ru; тел.: 8 (922) 127-66-88; ORCID: <https://orcid.org/0000-0002-7153-2827>).

**Смирнова Светлана Сергеевна** – кандидат медицинских наук, ведущий научный сотрудник, руководитель Урало-Сибирского научно-методического центра по профилактике инфекций, связанных с оказанием медицинской помощи; доцент кафедры эпидемиологии, социальной гигиены и организации госсанэпидслужбы (e-mail: smirnova\_ss@niivirom.ru; тел.: 8 (908) 917-59-86; ORCID: <https://orcid.org/0000-0002-9749-4611>).

**Семенов Александр Владимирович** – доктор биологических наук, директор; профессор кафедры медицинской микробиологии и клинической лабораторной диагностики (e-mail: semenov\_av@niivirom.ru; тел.: 8 (343) 261-99-47; ORCID: <https://orcid.org/0000-0003-3223-8219>).

Прошедшие два десятилетия XXI в. характеризуются эпидемическим и пандемическим распространением высококонтагиозных вирусных патогенов. Начало века (2002) ассоциировалось с распространением тяжелого острого респираторного синдрома (ТОРС). В период с 2009 по 2010 г. Всемирной организацией здравоохранения (ВОЗ) была объявлена пандемия вируса гриппа типа А (H1N1)<sup>1</sup>. Одна из самых крупных вспышек болезни, вызванной вирусом Эбола (БВВЭ), наблюдалась в 2014–2016 гг. на территории Западной Африки. В 2015 г. распространение получил коронавирус ближневосточного респираторного синдрома (БВРС) [1, 2]. Безусловно, одним из самых масштабных событий прошедших десятилетий XXI в. стала пандемия COVID-19 [3].

Эпидемическое и пандемическое распространение высококонтагиозных вирусов сопровождается усилением нагрузки на систему здравоохранения [4]. Преимущественное воздействие биологического фактора производственной среды на медицинских работников определяет их высокий профессиональный риск заражения высококонтагиозными вирусами, тяжелого течения болезни и летального исхода [5]. Доля медицинских работников в структуре заболевших в период эпидемий и пандемий, по данным различных исследований, достигает 30,0 % (ТОРС – 21,1 % [6]; вирус гриппа типа А (H1N1) – 27,1–30,0 % [7]; БВВЭ – 8,0 % [8]; БВРС – 18,7 % [9]; COVID-19 – 9,0–26,0 % [10]).

Технологии искусственного интеллекта успешно применяются при прогнозировании вспышечной заболеваемости различными инфекциями вирусной этиологии: болезни, вызванной вирусом Эбола [11], вирусными гепатитами и пневмонией, гриппом типа А [12, 13]. В пандемический и постпандемический периоды (COVID-19) интеллектуальный анализ накопленных данных помог решить задачи эпидемиологического характера: определение территорий, групп и факторов риска инфицирования (предикторов) COVID-19, прогнозирование заболеваемости и оценка эффективности профилактики COVID-19, прогнозирование мутаций, оценка тяжести поражения легких, дифференциальная диагностика по данным инструментальных методов обследования пациентов и моделирование молекулярного взаимодействия вируса SARS-CoV-2 [14–16]. Кроме того, практическое применение алгоритмов машинного обучения помогло решить вопросы, связанные с установлением факторов риска инфицирования медицинских работников SARS-CoV-2 [17, 18], определением приоритетных профессиональных групп медицинского сообщества для

проведения молекулярно-генетических исследований на РНК SARS-CoV-2, их изоляции [18, 19] и прогнозированием вероятности инфицирования SARS-CoV-2 по результатам интеллектуального анализа данных, полученных с носимых устройств медицинских работников [20].

Применение методов машинного обучения в определении предикторов инфицирования медицинских работников высококонтагиозными вирусами является современным этапом эпидемиологического анализа, помогает реализовать риск-ориентированный подход к профилактике заражения не только при состоявшемся эпидемическом и пандемическом распространении известных патогенов, но и при потенциальном развитии эпидемических угроз.

**Цель исследования** – по данным анкетирования медицинских работников построить модели машинного обучения и определить предикторы инфицирования медицинских работников высококонтагиозными вирусами (на модели COVID-19).

**Материалы и методы.** Исследование проведено специалистами Урало-Сибирского научно-методического центра по профилактике инфекций, связанных с оказанием медицинской помощи, Федерального бюджетного учреждения науки «Федеральный научно-исследовательский институт вирусных инфекций “Виром”» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека. Проведение исследования одобрено локальным этическим комитетом ЕНИИ-ВИ ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора, протокол от 24.06.2022 № 3 (название учреждения приведено на момент одобрения исследования, учреждение изменило название в соответствии с приказом Роспотребнадзора № 599 от 11.11.2022).

В пандемический период (2020–2021) на территории крупного промышленного города был проведен опрос 1312 медицинских работников. Проведение опроса включало заполнение на бумажных носителях оригинальной деперсонифицированной анкеты «Выявление профессиональных и непрофессиональных факторов, влияющих на риски инфицирования SARS-CoV-2 персонала медицинских организаций», разработанной авторами<sup>2</sup>. Структура анкеты включала открытые и закрытые вопросы, разделенные на шесть тематических блоков: половая принадлежность и антропометрические данные (рост, вес), половозрастная характеристика, профессиональная занятость, риски инфицирования COVID-19, приверженность соблюдению мер специфической и неспецифической профилактики инфицирования SARS-CoV-2, обстоятельства выявления COVID-19. Медицинских работников включали

<sup>1</sup> Influenza A (H1N1), pandemic 2009–2010 [Электронный ресурс] // WHO. – URL: [https://www.who.int/emergencies/situations/influenza-a-\(h1n1\)-outbreak](https://www.who.int/emergencies/situations/influenza-a-(h1n1)-outbreak) (дата обращения: 18.01.2024).

<sup>2</sup> Анкета для медицинских работников: документ на Яндекс Диске [Электронный ресурс]. – URL: <https://disk.yandex.ru/i/nNFNjGaVLs5KdG> (дата обращения: 12.03.2024).

в исследование на основании личного добровольно-согласия, полученного от каждого работника.

Данные с каждого бумажного носителя в ручном режиме вносили в электронную таблицу формата Microsoft Excel (\*.xlxs). Первичная база данных содержала 1312 строк, численно соответствовавших количеству респондентов, (в том числе 366 строк для медицинских работников, перенесших заболевание новой коронавирусной инфекцией, и 946 – для интактных по COVID-19) и 45 столбцов, сопоставимых с вопросами в анкете. Один из столбцов выступал в качестве зависимой (целевой) переменной: значение «1» соответствовало перенесенному заболеванию COVID-19, «0» – интактность по COVID-19.

На этапе предварительной обработки из базы данных были исключены анкеты, заполненные административным персоналом и анкеты, имеющие дефекты внесения данных. Дополнительно, с целью устранения дисбаланса классов зависимой (целевой) переменной на каждую анкету медицинского работника, заболевшего COVID-19, была подобрана сопоставимая по параметрам анкета сотрудника, интактного по COVID-19. На основании данных о росте и весе респондентов рассчитывали индекс массы тела (ИМТ). Интерпретацию ИМТ проводили в соответствии с рекомендациями Всемирной организации здравоохранения<sup>3</sup>.

Подготовленная база данных включала 688 строк (анкет) и 28 столбцов, содержащих неперсонализированные данные, в том числе 27 столбцов – изучаемые предикторы и один столбец – зависимая (целевая) переменная.

Отбор предикторов для обучения моделей машинного обучения осуществляли путем определения связи между каждым предиктором зависимой (целевой) переменной с расчетом критерия  $\chi^2$  Пирсона. Было отобрано 22 предиктора, обладающих наибольшей степенью зависимости ( $p < 0,05$ ).

Сформированная окончательная база данных была разделена на обучающую (2/3,  $n = 460$ ) и тестовую (1/3,  $n = 228$ ) подвыборки. Интеллектуальный анализ проводили с использованием пяти алгоритмов машинного обучения, применяемых для задач классификации: сверхслучайные деревья, деревья решений, случайный лес, логистическая регрессия, экстремальный градиентный бустинг. Для всех алгоритмов использовали оригинальные авторские настройки. Воспроизведение алгоритмов машинного обучения осуществляли в среде разработки Jupyter notebook (v.6.0.0) на языке программирования Python (v.3.7.16). Предварительную обработку и анализ данных проводили с применением библиотеки pandas, реализацию математических и числовых операций – numpy, разделение данных на обучающую и тестовую подвыборки, нормализацию данных, расчет ста-

тистических показателей, построение матрицы несоответствий, подбор параметров для обучения моделей, отбор предикторов осуществляли с помощью библиотеки для машинного обучения scikit-learn. Функционал алгоритмов реализовывали с помощью библиотек с открытым исходным кодом. Визуализацию полученных данных проводили с применением библиотек matplotlib и seaborn, функционала визуализации – библиотек SHapley Additive exPlanations<sup>4</sup>.

Интерпретацию статистических показателей работы моделей машинного обучения проводили с построением ROC-кривых, расчетом ROC-AUC (area under the curve, площади под кривой) и ее 95%-ного доверительного интервала (95 % ДИ).

На основании матрицы несоответствий рассчитывали долю истинно положительных, истинно отрицательных, ложноположительных и ложноотрицательных предсказаний. Учитывали только модели, обладающие статистической значимостью ( $p < 0,05$ ), а также достаточной чувствительностью и специфичностью (более 60,0 %).

Определение важности предикторов проводили с помощью показателя F-score, рассчитанного встроенным методом библиотеки экстремального градиентного бустинга – feature importance.

Силу эффекта каждого исследуемого предиктора на результат работы изучаемой модели оценивали с помощью его среднего предельного вклада (SHAP-значения) с учетом всех возможных комбинаций. Предикторами, определяющими инфицирование медицинского работника вирусом SARS-CoV-2, считали те, для которых были получены положительные SHAP-значения (больше 0). Дополнительно была проведена кластеризация изучаемых предикторов с порогом 90,0 %. Дизайн исследования представлен на рис. 1.

**Результаты и их обсуждение.** Всего было обучено 6912 моделей машинного обучения, из них (рис. 2):

- алгоритм сверхслучайных деревьев (чувствительность – 66,0, специфичность – 85,6, AUC (area under curve, площадь под кривой) – 69,9, 95 % ДИ [62,1–76,9]);
- алгоритм деревьев решений (чувствительность – 66,0, специфичность – 77,6, AUC – 73,5, 95 % ДИ [67,6–79,3]);
- алгоритм случайного леса (чувствительность – 65,0, специфичность – 80,8, AUC – 75,1, 95 % ДИ [68,1–81,5]);
- алгоритм логистической регрессии (чувствительность – 69,9, специфичность – 79,2, AUC – 79,4, 95 % ДИ [73,3–85,4]);
- алгоритм экстремального градиентного бустинга (чувствительность – 70,9, специфичность – 80,8, AUC – 80,4, 95 % ДИ [74,4–85,8]).

<sup>3</sup> Body mass index (BMI) [Электронный ресурс] // WHO. – URL: <https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/body-mass-index> (дата обращения: 18.01.2024).

<sup>4</sup> Welcome to the SHAP documentation [Электронный ресурс] // SHAP. – URL: <https://shap.readthedocs.io/en/latest/> (дата обращения: 19.01.2024).

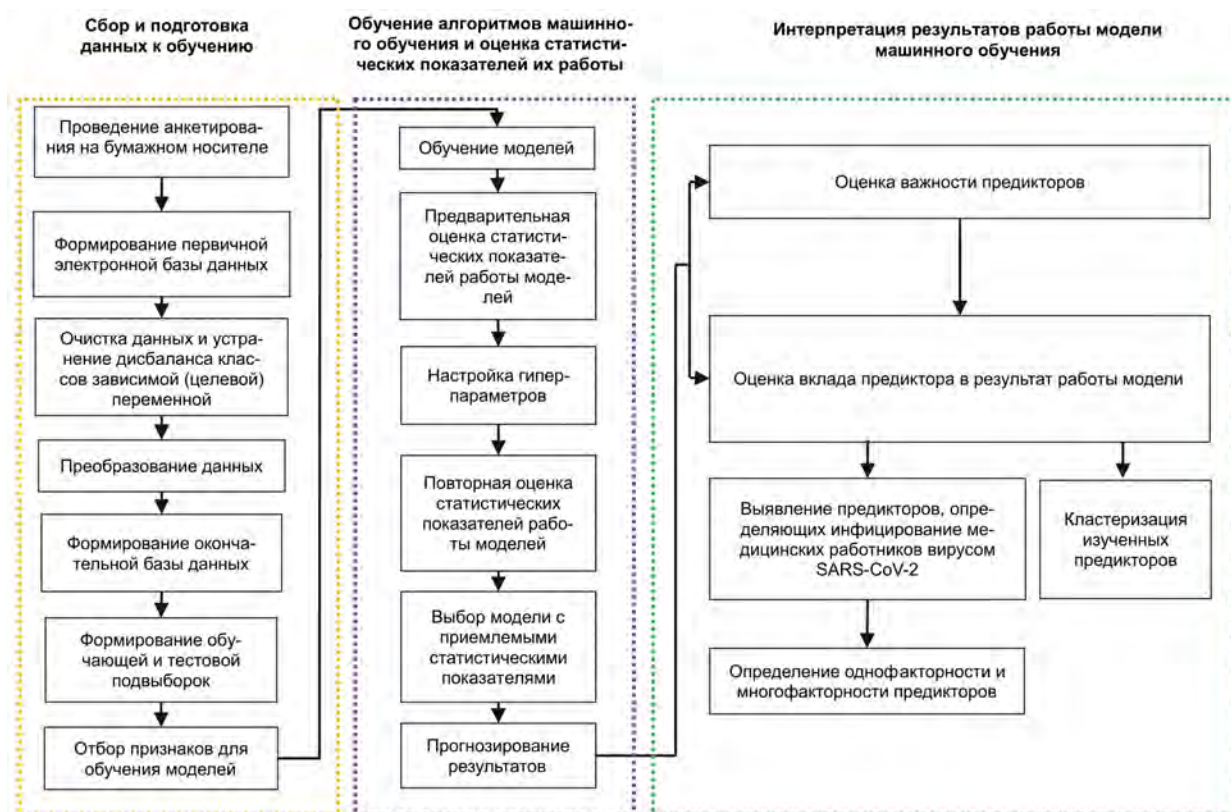


Рис. 1. Дизайн исследования

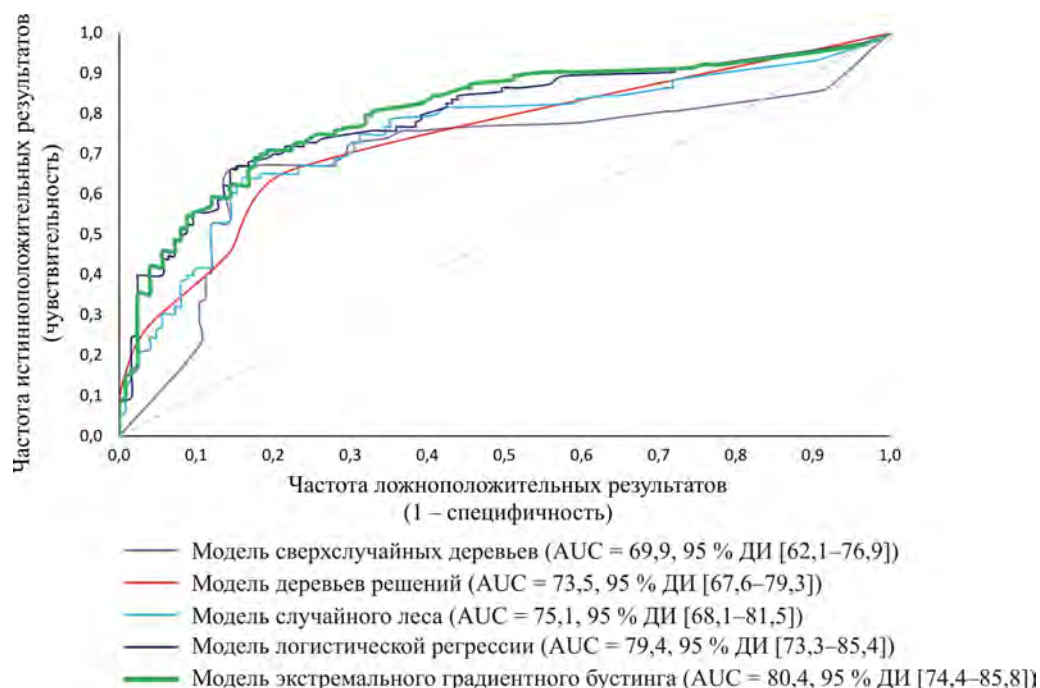


Рис. 2. ROC-кривые, описывающие статистические показатели работы алгоритмов машинного обучения

В ходе сравнительной оценки статистических показателей работы обученных моделей на исследуемом наборе персонифицированных данных установлена приемлемая чувствительность, специфичность и значение показателя AUC для алгоритма экстремального градиентного бустинга. Данный алгоритм использовался

для выявления предикторов, определяющих инфицирование медицинских работников SARS-CoV-2.

Анализ важности предикторов, проведенный с применением встроенного метода модели экстремального градиентного бустинга (показатель F-score), позволил выявить 19 предикторов (86,4 %) и сформиро-

вать несколько ранговых групп. Наибольшей важностью обладали: оказание амбулаторно-клинической медицинской помощи – 56,0, наличие заболевших COVID-19 в близком окружении медицинского работника – 46,0, оказание медицинской помощи пациенту с COVID-19 – 44,0, нормальный вес по показателю индекса массы тела (ИМТ) – 38,0, использование средств индивидуальной защиты (СИЗ) с неполной защитой органов зрения или дыхания – 32,0. Вторую ранговую позицию занимали такие предикторы, как продолжительность смены более 24 ч – 29,0, выполнение функций среднего медицинского персонала – 24,0, вакцинация от COVID-19 после непосредственного контакта с больным COVID-19 – 21,0. Третью – наличие аварийных ситуаций, связанных с экспозицией биоматериала пациента, – 19,0, избыточная масса тела (предожирение) – 17,0, наличие хронических соматических заболеваний – 16,0, прямой контакт с предметами внешней (больничной) среды – 13,0, использование полного комплекта СИЗ после непосредственного контакта с больным COVID-19 и ожирение первой степени – по 12,0. Четвертую – выполнение функций врача и ожирение второй степени – по 7,0, выполнение функций младшего медицинского персонала (уборщиков) – 6,0, присутствие при проведении процедур, генерирующих аэрозоль, – 5,0, проведение лабораторной и патолого-анатомической диагностики – 1,0.

В нашем исследовании оценка важности предикторов по показателю F-score имела ограничение при интерпретации результатов работы изучаемой модели. Определение моделью факта отсутствия дефицита или избытка веса в качестве предиктора обусловлено высокой частотой распространения данного признака в обучающей подвыборке (65,4 %,  $n = 301$ ).

В целях более четкой оценки важности предикторов, определяющих инфицирование SARS-CoV-2 медицинских работников, для дальнейших этапов исследования применялась стратегия оценки силы эффекта каждого предиктора с расчетом SHAP-значения. Данный подход позволил выявить предикторы, определяющие инфицирование, и провести их ранжирование в зависимости от величины SHAP-значения.

Сила эффекта изученных предикторов имела различную интенсивность – до 10,5 раза (0,9904–0,0943,  $p < 0,05$ ), в том числе (рис. 3): оказание медицинской помощи пациенту с COVID-19 ( $2,378 \pm 0,791$ ,  $p < 0,05$ ), использование полного комплекта СИЗ после непосредственного контакта с больным COVID-19 ( $0,565 \pm 0,17$ ,  $p < 0,05$ ), прямой контакт с предметами внешней (больничной) среды ( $0,547 \pm 0,146$ ,  $p < 0,05$ ), вакцинация от COVID-19 после непосредственного контакта с больным COVID-19 ( $0,304 \pm 0,072$ ,  $p < 0,05$ ), выполнение функций младшего медицинского персонала (уборщиков) ( $0,162 \pm 0,035$ ,  $p < 0,05$ ), присутствие при проведении процедур, генерирующих аэрозоль ( $0,109 \pm 0,022$ ,  $p < 0,05$ ). Отмечено, что ряд предикторов, определяющих инфициро-

вание SARS-CoV-2, не были связаны с профессией. Например, наличие заболевших COVID-19 в близком окружении медицинского работника ( $1,464 \pm 0,58$ ,  $p < 0,05$ ), ожирение второй степени ( $0,259 \pm 0,04$ ,  $p < 0,05$ ), наличие хронических соматических заболеваний ( $0,148 \pm 0,092$ ,  $p < 0,05$ ).

Проведенный кластерный анализ предикторов, определяющих инфицирование SARS-CoV-2 медицинских работников, позволил выявить четыре различных кластера (рис. 4):

- 1-й кластер был связан с непосредственным контактом с больным COVID-19 или предметами окружающей его внешней (больничной) среды: оказание медицинской помощи пациенту с COVID-19 ( $p < 0,05$ ), присутствие при проведении процедур, генерирующих аэрозоль ( $p < 0,05$ ), прямой контакт с предметами внешней (больничной) среды ( $p < 0,05$ );

- 2-й кластер характеризовал качество и комплектность применяемых СИЗ: использование СИЗ с неполной защитой органов зрения или дыхания ( $p > 0,05$ ), использование полного комплекта СИЗ после непосредственного контакта с больным COVID-19 ( $p < 0,05$ );

- 3-й кластер отражал профессиональную принадлежность медицинских работников: выполнение функций младшего медицинского персонала (уборщиков) ( $p < 0,05$ ), выполнение функций среднего медицинского персонала ( $p > 0,05$ ), выполнение функций врача ( $p > 0,05$ );

- 4-й кластер характеризовал личные признаки сотрудника, такие как ИМТ: нормальный вес по показателю индекса массы тела (ИМТ) ( $p > 0,05$ ), избыточная масса тела (предожирение) ( $p > 0,05$ ), ожирение первой степени ( $p > 0,05$ ).

Следующим этапом исследования явилось определение однофакторности и многофакторности взаимодействия изучаемых предикторов, связанных с профессиональной деятельностью.

В нашем исследовании наличие одного предиктора, определяющего инфицирование SARS-CoV-2, было установлено у 56,2 % медицинских работников, двух – у 19,2 %, трех – у 16,4 %, четырех – у 5,5 %, пяти – у 2,7 %.

Частота встречаемости предикторов при однофакторном воздействии различалась. Лидирующее место занимала вакцинация от COVID-19 лишь после непосредственного контакта с больным COVID-19 – 65,9 %, второе – выполнение функций младшего медицинского персонала (уборщиков) – 22,0 %, третье – оказание медицинской помощи пациенту с COVID-19 – 12,2 %.

При наличии двух предикторов первое ранговое место делили присутствие при проведении процедур, генерирующих аэрозоль, и использование полного комплекта СИЗ после непосредственного контакта с больным COVID-19 – по 32,1 %, второе место занимала вакцинация от COVID-19 после непосредственного контакта с больным COVID-19 – 17,9 %, третье – оказание медицинской помощи



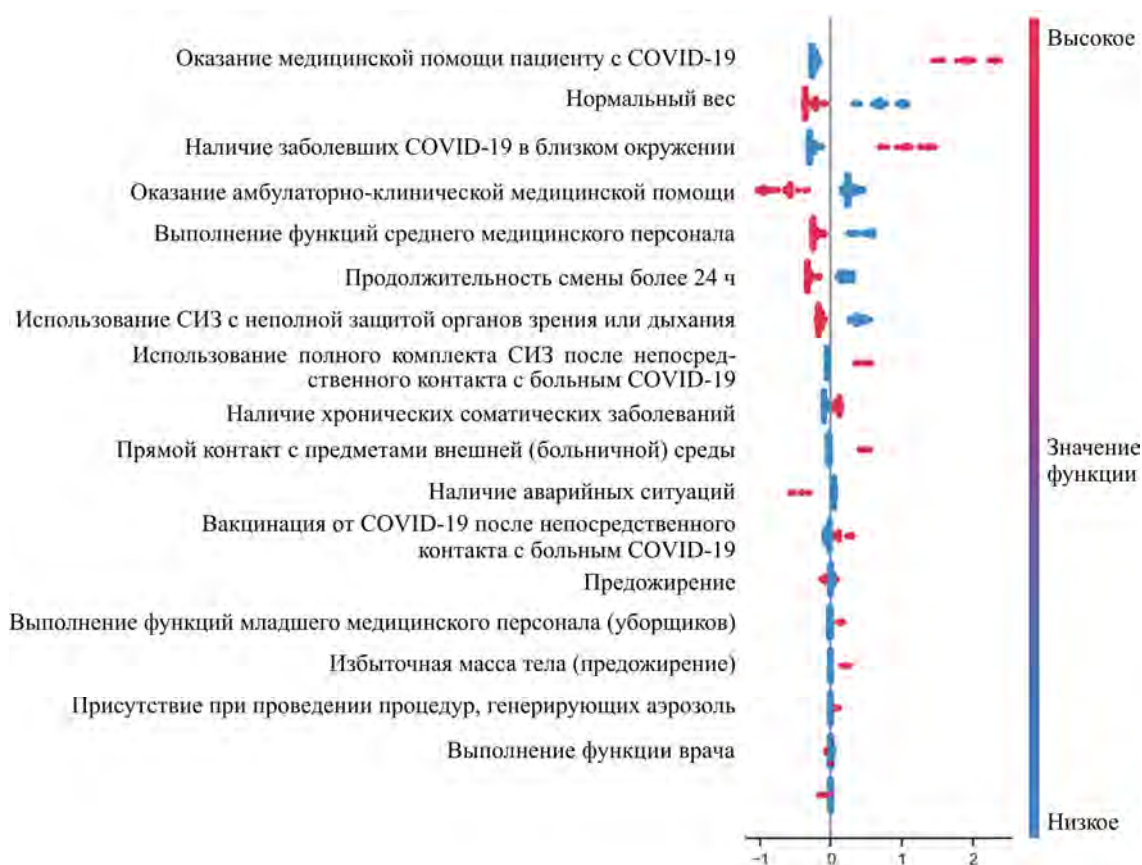


Рис. 3. Предикторы, определяющие инфицирование медицинских работников SARS-CoV-2

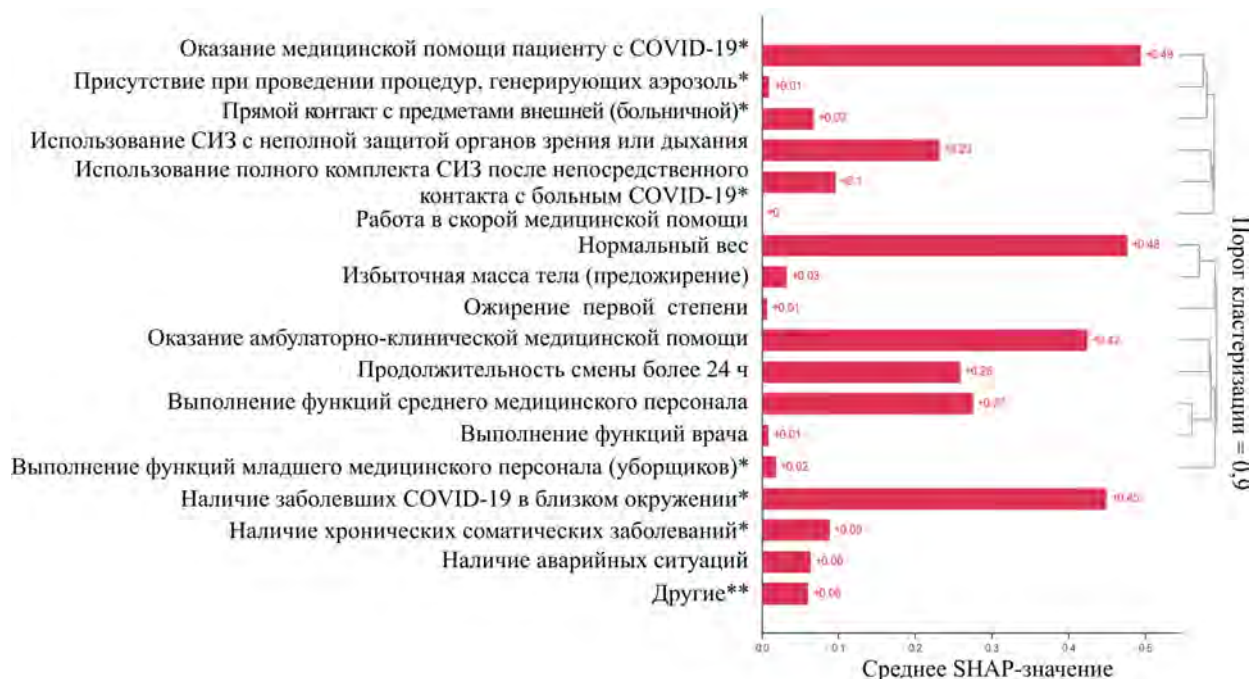


Рис. 4. Кластеризация предикторов, определяющих инфицирование медицинских работников SARS-CoV-2:  
 \* – предикторы, имеющие положительное влияние на инфицирование медицинских работников SARS-CoV-2;  
 \*\* – сумма SHAP-значений: вакцинация от COVID-19 после непосредственного контакта с больным COVID-19; ожирение второй и третьей степени; недостаточный вес; проведение лабораторной и патолого-анатомической диагностики

пациенту с COVID-19 – 10,7 %, четвертое – выполнение функций младшего медицинского персонала (уборщиков) – 7,1 %.

Структура многофакторного взаимодействия при наличии трех предикторов выглядела следующим образом: присутствие при проведении процедур, генерирующих аэрозоль, – 27,8 %, оказание медицинской помощи пациенту с COVID-19 и использование полного комплекта СИЗ после непосредственного контакта с больным COVID-19 – по 19,4 %, вакцинация от COVID-19 после непосредственного контакта с больным COVID-19 и прямой контакт с предметами внешней (больничной) среды – по 13,9 %, выполнение функций младшего медицинского персонала (уборщиков) – 5,6 %.

При наличии четырех предикторов многофакторное взаимодействие определялось присутствием при проведении процедур, генерирующих аэрозоль, и прямым контактом с предметами внешней (больничной) среды – по 25,0 %, оказанием медицинской помощи пациенту с COVID-19 и использованием полного комплекта СИЗ после непосредственного контакта с больным COVID-19 – 18,8 %, вакцинацией от COVID-19 после непосредственного контакта с больным COVID-19 и выполнением функций младшего медицинского персонала (уборщиков) – по 6,3 %.

Равная частота встречаемости была установлена при одновременном наличии пяти предикторов: вакцинация от COVID-19 после непосредственного контакта с больным COVID-19, оказание медицинской помощи пациенту с COVID-19, присутствие при проведении процедур, генерирующих аэрозоль, использование полного комплекта СИЗ после непосредственного контакта с больным COVID-19, пря-

мой контакт с предметами внешней (больничной) среды – по 20,0 %.

**Выводы.** Таким образом, интеллектуальная обработка эпидемиологических данных является современным этапом эпидемиологического анализа. Применение методов машинного обучения позволяет провести многофакторную оценку рисков инфицирования медицинских работников SARS-CoV-2, выявить и достоверно оценить наиболее значимые предикторы, сформировать группы риска с возможностью реализации персонализированного подхода к профилактике профессионального заражения вирусными патогенами.

Гибкость архитектуры интеллектуального анализа эпидемиологических данных позволяет проводить как корректировку изучаемой модели, так и динамически дополнять сформированную базу новыми данными, улавливать изменения эпидемиологической ситуации и проводить актуальные профилактические и противоэпидемические мероприятия. Качество и точность результата работы модели машинного обучения достигается полным и качественным сбором первоначальных данных, проведением предварительной обработки и использованием для обучения модели выверенных баз (наборов) данных.

**Финансирование.** Исследование выполнено в рамках научно-исследовательской работы «Изучение эпидемиологического процесса и профилактика вирусных инфекций, связанных с оказанием медицинской помощи (на примере ветряной оспы, норо- и ротавирусной инфекции и др.)», Рег. № НИОКТР 121040500099-5.

**Конфликт интересов.** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

### Список литературы

1. Infectious disease in an era of global change / R.E. Baker, A.S. Mahmud, I.F. Miller, M. Rajeev, F. Rasambainarivo, B.L. Rice, S. Takahashi, A.J. Tatem [et al.] // *Nat. Rev. Microbiol.* – 2022. – Vol. 20, № 4. – P. 193–205. DOI: 10.1038/s41579-021-00639-z
2. Emerging paradigms of viral diseases and paramount role of natural resources as antiviral agents / R. Sagaya Jansi, A. Khusro, P. Agastian, A. Alfathan, N.A. Al-Dhabi, M.V. Arasu, R. Rajagopal, D. Barcelo, A. Al-Tamimi // *Sci. Total Environ.* – 2021. – Vol. 759. – P. 143539. DOI: 10.1016/j.scitotenv.2020.143539
3. Ashmore P., Sherwood E. An overview of COVID-19 global epidemiology and discussion of potential drivers of variable global pandemic impacts // *J. Antimicrob. Chemother.* – 2023. – Vol. 78, Suppl. 2. – P. ii2–ii11. DOI: 10.1093/jac/dkad311
4. Risk factors and protective measures for healthcare worker infection during highly infectious viral respiratory epidemics: A systematic review and meta-analysis / C. Tian, O. Lovrics, A. Vaisman, K.J. Chin, G. Tomlinson, Y. Lee, M. Englesakis, M. Parotto, M. Singh // *Infect. Control Hosp. Epidemiol.* – 2022. – Vol. 43, № 5. – P. 639–650. DOI: 10.1017/ice.2021.18
5. COVID-19 and healthcare workers: A systematic review and meta-analysis / M. Gholami, I. Fawad, S. Shadan, R. Rowaiee, H. Ghanem, A. Hassan Khamis, S.B. Ho // *Int. J. Infect. Dis.* – 2021. – Vol. 104. – P. 335–346. DOI: 10.1016/j.ijid.2021.01.013
6. Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS) / W.H. Seto, D. Tsang, R.W.H. Yung, T.Y. Ching, T.K. Ng, M. Ho, L.M. Ho, J.S.M. Peiris, Advisors of Expert SARS group of Hospital Authority // *Lancet.* – 2003. – Vol. 361, № 9368. – P. 1519–1520. DOI: 10.1016/s0140-6736(03)13168-6
7. Factors associated with the transmission of pandemic (H1N1) 2009 among hospital healthcare workers in Beijing, China / Y. Zhang, H. Seale, P. Yang, C.R. MacIntyre, B. Blackwell, S. Tang, Q. Wang // *Influenza Other Respir. Viruses.* – 2013. – Vol. 7, № 3. – P. 466–471. DOI: 10.1111/irv.12025
8. Epidemiology of Ebola virus disease transmission among health care workers in Sierra Leone, May to December 2014: a retrospective descriptive study / O. Olu, B. Kargbo, S. Kamara, A.H. Wurie, J. Amone, L. Ganda, B. Ntsama, A. Poy [et al.] // *BMC Infect. Dis.* – 2015. – Vol. 15. – P. 416. DOI: 10.1186/s12879-015-1166-7
9. Risk Factors for Middle East Respiratory Syndrome Coronavirus Infection among Healthcare Personnel / B.M. Al-raddadi, H.S. Al-Salmi, K. Jacobs-Slifka, R.B. Slayton, C.F. Estivariz, A.I. Geller, H.H. Al-Turkistani, S.S. Al-Rehily [et al.] // *Emerg. Infect. Dis.* – 2016. – Vol. 22, № 11. – P. 1915–1920. DOI: 10.3201/eid2211.160920
10. Сравнительная оценка рисков инфицирования SARS-CoV-2 у работников медицинских организаций крупного промышленного города в период пандемии / С.С. Смирнова, И.А. Егоров, Н.Н. Жуйков, Л.Г. Вяткина, А.Н. Харитонов, А.В. Семенов, О.В. Морозова // *Анализ риска здоровью.* – 2022. – № 2. – С. 139–150. DOI: 10.21668/health.risk/2022.2.13

11. The Large Scale Machine Learning in an Artificial Society: Prediction of the Ebola Outbreak in Beijing / P. Zhang, B. Chen, L. Ma, Z. Li, Z. Song, W. Duan, X. Qiu // *Comput. Intell. Neurosci.* – 2015. – Vol. 2015. – P. 531650. DOI: 10.1155/2015/531650
12. Alqaissi E.Y., Alotaibi F.S., Ramzan M.S. Modern Machine-Learning Predictive Models for Diagnosing Infectious Diseases // *Comput. Math. Methods Med.* – 2022. – Vol. 2022. – P. 6902321. DOI: 10.1155/2022/6902321
13. Singh R., Singh R. Applications of sentiment analysis and machine learning techniques in disease outbreak prediction – A review // *Materials Today: Proceedings.* – 2021. – Vol. 81, № 2. – P. 1006–1011. DOI: 10.1016/j.matpr.2021.04.356
14. Artificial Intelligence in the Fight Against COVID-19: Scoping Review / A. Abd-Alrazaq, M. Alajlani, D. Alhuwail, J. Schneider, S. Al-Kuwari, Z. Shah, M. Hamdi, M. Househ // *J. Med. Internet Res.* – 2020. – Vol. 22, № 12. – P. e20756. DOI: 10.2196/20756
15. Comprehensive Overview of the COVID-19 Literature: Machine Learning-Based Bibliometric Analysis / A. Abd-Alrazaq, J. Schneider, B. Mifsud, T. Alam, M. Househ, M. Hamdi, Z.A. Shah // *J. Med. Internet Res.* – 2021. – Vol. 23, № 3. – P. e23703. DOI: 10.2196/23703
16. The application of artificial intelligence and data integration in COVID-19 studies: a scoping review / Y. Guo, Y. Zhang, T. Lyu, M. Prosperi, F. Wang, H. Xu, J. Bian // *J. Am. Med. Inform. Assoc.* – 2021. – Vol. 28, № 9. – P. 2050–2067. DOI: 10.1093/jamia/ocab098
17. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread // *Lancet Digit. Health.* – 2020. – Vol. 2, № 4. – P. e166–e167. DOI: 10.1016/S2589-7500(20)30054-6
18. Strong associations and moderate predictive value of early symptoms for SARS-CoV-2 test positivity among healthcare workers, the Netherlands, March 2020 / A. Tostmann, J. Bradley, T. Bousema, W.-K. Yiek, M. Holwerda, C. Bleeker-Rovers, J. Ten Oever, C. Meijer [et al.] // *Euro Surveill.* – 2020. – Vol. 25, № 16. – P. 2000508. DOI: 10.2807/1560-7917.ES.2020.25.16.2000508
19. Routine Laboratory Blood Tests Predict SARS-CoV-2 Infection Using Machine Learning / H.S. Yang, Y. Hou, L.V. Vasovic, P.A.D. Steel, A. Chadburn, S.E. Racine-Brzostek, P. Velu, M.M. Cushing [et al.] // *Clin. Chem.* – 2020. – Vol. 66, № 11. – P. 1396–1404. DOI: 10.1093/clinchem/hvaa200
20. Evaluation of a machine learning approach utilizing wearable data for prediction of SARS-CoV-2 infection in health-care workers / R.P. Hirten, L. Tomalin, M. Danieleto, E. Golden, M. Zweig, S. Kaur, D. Helmus, A. Biello [et al.] // *JAMIA Open.* – 2022. – Vol. 5, № 2. – P. ooac041. DOI: 10.1093/jamiaopen/ooac041

*Егоров И.А., Смирнова С.С., Семенов А.В. Использование индуктивного типа машинного обучения для выявления факторов риска инфицирования медицинских работников высококонтагиозными вирусами (на модели COVID-19) // Анализ риска здоровью. – 2024. – № 2. – С. 122–131. DOI: 10.21668/health.risk/2024.2.11*

UDC 613.6.02, 616-036.22, 004.8  
DOI: 10.21668/health.risk/2024.2.11.eng



Research article

## USING INDUCTIVE MACHINE LEARNING TO IDENTIFY RISK FACTORS FOR HEALTHCARE WORKERS TO GET INFECTED WITH HIGHLY CONTAGIOUS VIRUSES (BASED ON COVID-19 MODEL)

**I.A. Egorov<sup>1</sup>, S.S. Smirnova<sup>1,2</sup>, A.V. Semenov<sup>1,2</sup>**

<sup>1</sup>Federal Scientific Research Institute of Viral Infections «Virome», 23 Letnyaya St., Ekaterinburg, 620030, Russian Federation

<sup>2</sup>Ural State Medical University, 3 Repina St., Ekaterinburg, 620028, Russian Federation

*Epidemic and pandemic spread of highly contagious viruses (SARS-CoV, influenza A virus, Ebola virus, MERS-CoV, and SARS-CoV-2) has been a trend observed in the first two decades of the 21<sup>st</sup> century.*

© Egorov I.A., Smirnova S.S., Semenov A.V., 2024

**Ivan A. Egorov** – Researcher at Ural-Siberian Research and Methodology Center for Prevention of Healthcare-Associated Infections (e-mail: egorov\_ia@niivirom.ru; tel.: +7 (922) 127-66-88; ORCID: <https://orcid.org/0000-0002-7153-2827>).

**Svetlana S. Smirnova** – Candidate of Medical Sciences, Leading Researcher, Head of the Ural-Siberian Research and Methodology Center for Prevention of Healthcare-Associated Infections; Associate Professor of the Department of Epidemiology, Social Hygiene and Organization of Sanitary and Epidemiological Service (e-mail: smirnova\_ss@niivirom.ru; tel.: +7 (908) 917-59-86; ORCID: <https://orcid.org/0000-0002-9749-4611>).

**Aleksandr V. Semenov** – Doctor of Biological Sciences, director; Professor of the Department of Medical Microbiology and Clinical Laboratory Diagnostics (e-mail: semenov\_av@niivirom.ru; tel.: +7 (343) 261-99-47; ORCID: <https://orcid.org/0000-0003-3223-8219>).



The predominant impact made by the biological occupational factor on healthcare workers determines their high occupational risk of infection, a severe disease course and a fatal outcome. Epidemiological data mining based on machine learning algorithms is successfully used in epidemiological practice to identify factors (predictors) contributing to infection in various risk populations.

In this study, the database generated from a survey of 1312 healthcare workers was analyzed intelligently. A total of 6912 machine learning models were implemented. SARS-CoV-2 infection was found to be facilitated by providing medical care to a COVID-19 patient, using a full set of PPE after direct contact with a COVID-19 patient, direct contact with items in the external (hospital) environment, vaccination against COVID-19 after direct contact with a COVID-19 patient, acting as nursing staff (cleaners) and being present during aerosol-generating procedures.

The study identified four groups of predictors determining SARS-CoV-2 infection in healthcare workers: contact with a COVID-19 patient and environmental items, PPE quality and complexity, occupational affiliation of healthcare workers and their BMI values. One predictor was found in 56.2 % of healthcare workers; two, in 19.2 %; three, in 16.4 %; four, in 5.5 %; and five predictors, in 2.7 %.

Thus, epidemiological data mining is a modern stage in epidemiological analysis. The use of machine learning methods allows for multifactorial assessment of SARS-CoV-2 infection risks in healthcare workers and enables identifying and reliably estimating the most significant predictors. Intelligent data analysis has flexible architecture, which allows adjusting the model under study and supplementing new data to the existing database, detecting changes in an epidemiological situation and accomplishing relevant preventive and anti-epidemic activities.

**Keywords:** data mining, artificial intelligence, machine learning, risk-based approach, occupational predictors of infection, highly contagious viruses, SARS-CoV-2, healthcare workers.

## References

1. Baker R.E., Mahmud A.S., Miller I.F., Rajeev M., Rasambainarivo F., Rice B.L., Takahashi S., Tatem A.J. [et al.]. Infectious disease in an era of global change. *Nat. Rev. Microbiol.*, 2022, vol. 20, no. 4, pp. 193–205. DOI: 10.1038/s41579-021-00639-z
2. Sagaya Jansi R., Khusro A., Agastian P., Alfarhan A., Al-Dhabi N.A., Arasu M.V., Rajagopal R., Barcelo D., Al-Tamimi A. Emerging paradigms of viral diseases and paramount role of natural resources as antiviral agents. *Sci. Total Environ.*, 2021, vol. 759, pp. 143539. DOI: 10.1016/j.scitotenv.2020.143539
3. Ashmore P., Sherwood E. An overview of COVID-19 global epidemiology and discussion of potential drivers of variable global pandemic impacts. *J. Antimicrob. Chemother.*, 2023, vol. 78, suppl. 2, pp. ii2–ii11. DOI: 10.1093/jac/dkad311
4. Tian C., Lovrics O., Vaisman A., Chin K.J., Tomlinson G., Lee Y., Englesakis M., Parotto M., Singh M. Risk factors and protective measures for healthcare worker infection during highly infectious viral respiratory epidemics: A systematic review and meta-analysis. *Infect. Control Hosp. Epidemiol.*, 2022, vol. 43, no. 5, pp. 639–650. DOI: 10.1017/ice.2021.18
5. Gholami M., Fawad I., Shadan S., Rowaiee R., Ghanem H., Hassan Khamis A., Ho S.B. COVID-19 and healthcare workers: A systematic review and meta-analysis. *Int. J. Infect. Dis.*, 2021, vol. 104, pp. 335–346. DOI: 10.1016/j.ijid.2021.01.013
6. Seto W.H., Tsang D., Yung R.W.H., Ching T.Y., Ng T.K., Ho M., Ho L.M., Peiris J.S.M., Advisors of Expert SARS group of Hospital Authority. Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS). *Lancet*, 2003, vol. 361, no. 9368, pp. 1519–1520. DOI: 10.1016/s0140-6736(03)13168-6
7. Zhang Y., Seale H., Yang P., MacIntyre C.R., Blackwell B., Tang S., Wang Q. Factors associated with the transmission of pandemic (H1N1) 2009 among hospital healthcare workers in Beijing, China. *Influenza Other Respir. Viruses*, 2013, vol. 7, no. 3, pp. 466–471. DOI: 10.1111/irv.12025
8. Olu O., Kargbo B., Kamara S., Wurie A.H., Amone J., Ganda L., Ntsama B., Poy A. [et al.]. Epidemiology of Ebola virus disease transmission among health care workers in Sierra Leone, May to December 2014: a retrospective descriptive study. *BMC Infect. Dis.*, 2015, vol. 15, pp. 416. DOI: 10.1186/s12879-015-1166-7
9. Alraddadi B.M., Al-Salmi H.S., Jacobs-Slifka K., Slayton R.B., Estivariz C.F., Geller A.I., Al-Turkistani H.H., Al-Rehily S.S. [et al.]. Risk Factors for Middle East Respiratory Syndrome Coronavirus Infection among Healthcare Personnel. *Emerg. Infect. Dis.*, 2016, vol. 22, no. 11, pp. 1915–1920. DOI: 10.3201/eid2211.160920
10. Smirnova S.S., Egorov I.A., Zhuikov N.N., Vyatkina L.G., Kharitonov A.N., Semenov A.V., Morova O.V. Risks of becoming infected with SARS-CoV-2 for medical personnel in a large industrial city during the pandemic: comparative assessment. *Health Risk Analysis*, 2022, no. 2, pp. 139–150. DOI: 10.21668/health.risk/2022.2.13.eng
11. Zhang P., Chen B., Ma L., Li Z., Song Z., Duan W., Qiu X. The Large Scale Machine Learning in an Artificial Society: Prediction of the Ebola Outbreak in Beijing. *Comput. Intell. Neurosci.*, 2015, vol. 2015, pp. 531650. DOI: 10.1155/2015/531650
12. Alqaissi E.Y., Alotaibi F.S., Ramzan M.S. Modern Machine-Learning Predictive Models for Diagnosing Infectious Diseases. *Comput. Math. Methods Med.*, 2022, vol. 2022, pp. 6902321. DOI: 10.1155/2022/6902321
13. Singh R., Singh R. Applications of sentiment analysis and machine learning techniques in disease outbreak prediction – A review. *Materials Today: Proceedings*, 2021, vol. 81, no. 2, pp. 1006–1011. DOI: 10.1016/j.matpr.2021.04.356
14. Abd-Alrazaq A., Alajlani M., Alhuwail D., Schneider J., Al-Kuwari S., Shah Z., Hamdi M., Househ M. Artificial Intelligence in the Fight Against COVID-19: Scoping Review. *J. Med. Internet Res.*, 2020, vol. 22, no. 12, pp. e20756. DOI: 10.2196/20756
15. Abd-Alrazaq A., Schneider J., Mifsud B., Alam T., Househ M., Hamdi M., Shah Z.A. Comprehensive Overview of the COVID-19 Literature: Machine Learning-Based Bibliometric Analysis. *J. Med. Internet Res.*, 2021, vol. 23, no. 3, pp. e23703. DOI: 10.2196/23703

16. Guo Y., Zhang Y., Lyu T., Prosperi M., Wang F., Xu H., Bian J. The application of artificial intelligence and data integration in COVID-19 studies: a scoping review. *J. Am. Med. Inform. Assoc.*, 2021, vol. 28, no. 9, pp. 2050–2067. DOI: 10.1093/jamia/ocab098
17. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. *Lancet Digit. Health*, 2020, vol. 2, no. 4, pp. e166–e167. DOI: 10.1016/S2589-7500(20)30054-6
18. Tostmann A., Bradley J., Bousema T., Yiek W.-K., Holwerda M., Bleeker-Rovers C., Ten Oever J., Meijer C. [et al.]. Strong associations and moderate predictive value of early symptoms for SARS-CoV-2 test positivity among healthcare workers, the Netherlands, March 2020. *Euro Surveill.*, 2020, vol. 25, no. 16, pp. 2000508. DOI: 10.2807/1560-7917.ES.2020.25.16.2000508
19. Yang H.S., Hou Y., Vasovic L.V., Steel P.A.D., Chadburn A., Racine-Brzostek S.E., Velu P., Cushing M.M. [et al.]. Routine Laboratory Blood Tests Predict SARS-CoV-2 Infection Using Machine Learning. *Clin. Chem.*, 2020, vol. 66, no. 11, pp. 1396–1404. DOI: 10.1093/clinchem/hvaa200
20. Hirten R.P., Tomalin L., Danieletto M., Golden E., Zweig M., Kaur S., Helmus D., Biello A. [et al.]. Evaluation of a machine learning approach utilizing wearable data for prediction of SARS-CoV-2 infection in healthcare workers. *JAMIA Open*, 2022, vol. 5, no. 2, pp. ooac041. DOI: 10.1093/jamiaopen/ooac041

*Egorov I.A., Smirnova S.S., Semenov A.V. Using inductive machine learning to identify risk factors for healthcare workers to get infected with highly contagious viruses (based on COVID-19 model). Health Risk Analysis, 2024, no. 2, pp. 122–131. DOI: 10.21668/health.risk/2024.2.11.eng*

Получена: 02.02.2024

Одобрена: 22.04.2024

Принята к публикации: 20.06.2024

Научная статья

## МНОЖЕСТВЕННАЯ ЛЕКАРСТВЕННАЯ УСТОЙЧИВОСТЬ УРОПАТОГЕНОВ КАК ФАКТОР РИСКА ПРИ ОКАЗАНИИ МЕДИЦИНСКОЙ ПОМОЩИ (ОПЫТ ЦЕНТРАЛЬНОЙ БОЛГАРИИ)

С. Алекова, Р. Койчева

Тракийски университет, Болгария, 6000, г. Стара-Загора, ул. Армейская, 11

*Возникновение устойчивости уропатогенных бактерий к множественным антимикробным препаратам является значимой проблемой здравоохранения. Указанные микроорганизмы вызывают инфекции мочевыводящих путей – распространенных заболеваний, подлежащих амбулаторному лечению.*

*Проведено одновременное поперечное углубленное исследование 353 пациентов поликлиник в возрасте старше 18 лет, анализ мочи которых оказался положительным на наличие бактерий. В рамках исследования был определен индекс множественной лекарственной устойчивости (МЛУ) уропатогенов, вызывающих инфекции у таких пациентов. Исследование проводилось в период с января по июнь 2023 г. в частном клиническом лабораторном центре в г. Стара-Загора, Болгария.*

*Одна треть штаммов E. coli и более половины штаммов Klebsiella pneumoniae имели значение измеренного индекса МЛУ выше 0,2, а также подтвержденный высокий уровень устойчивости к аминопенициллинам и цефалоспорином. Настораживающе высокие уровни индекса МЛУ были определены для всех представителей Enterococcus faecalis, видов Enterobacter и Pseudomonas aeruginosa. Устойчивость данных патогенов к некоторым антимикробным препаратам, наиболее часто назначаемым при амбулаторном лечении, варьировалась от 58,3 до 100 %.*

*Результаты исследования подчеркивают значимость и серьезность существующей проблемы возрастающего преобладания уропатогенных микроорганизмов с МЛУ к антибиотикам, принимаемым перорально, что повышает риски неэффективности медицинской помощи пациентам.*

**Ключевые слова:** инфекции мочевыводящих путей, бактерии, индекс множественной лекарственной устойчивости, уропатогены, антимикробные препараты, фактор риска.

Подавляющее большинство рецептов на антибиотики выписываются пациентам в рамках первичной и специализированной медицинской помощи [1]. Безответственное и некомпетентное поведение пациентов при самолечении является огромной ошибкой, приводящей к массовому применению антибиотиков [2]. Инфекции мочевыводящих путей (ИМП) являются наиболее распространенными бактериальными заболеваниями, которым в одинаковой степени подвержены и мужчины и женщины. Инфекции становятся все труднее и труднее лечить в амбулаторных условиях, все больше времени требуется на их устранение по причине приобретенной

лекарственной устойчивости выделяемых бактериальных штаммов. К сожалению, лекарственная устойчивость является одной из самых серьезных угроз здоровью населения, поскольку она увеличивает затраты на здравоохранение и повышает частоту неблагоприятных исходов лечения [3]. Лекарственная устойчивость – это способность микроорганизма сопротивляться действию одного или нескольких антимикробных препаратов в результате мутации хромосомных генов или приобретения внешних генов устойчивости [4, 5]. Путем приобретения механизмов множественной лекарственной устойчивости (МЛУ) бактериальные штаммы становятся устойчи-

© Алекова С., Койчева Р., 2024

**Алекова Севдалина Тодорова** – доктор медицинских наук, главный ассистент кафедры внутренних болезней и общей медицины медицинского факультета (e-mail: sevdalina.alekova@abv.bg; тел.: +359 896-610-001; ORCID: <https://orcid.org/0000-0002-0443-5891>).

**Койчева Ренета** – доктор медицинских наук, главный ассистент кафедры внутренних болезней и общей медицины медицинского факультета (e-mail: koychevar@abv.bg; тел.: +359 888-990-135).