

В.Г. Акимкин, К.Ф. Хафизов, Д.В. Дубоделов,
Е.М. Воронин, А.С. Черкашина, С.В. Углева,
Д.В. Стратулат, А.А. Самсонов



Центральный научно-исследовательский институт эпидемиологии,
Москва, Российская Федерация

Молекулярно-генетический мониторинг и технологии цифровой трансформации в современной эпидемиологии

Пандемия новой коронавирусной инфекции явилась серьезным вызовом для системы здравоохранения всех стран. Так, по состоянию на май 2023 г. в мире зарегистрировано 766 млн подтвержденных случаев заболевания и более 6,9 млн случаев смерти. В Российской Федерации зарегистрировано более 22 млн случаев заболевания и 398 тыс. летальных случаев от новой коронавирусной инфекции. Быстро меняющаяся эпидемиологическая ситуация обусловила потребность в систематизации материала для проведения аналитической работы. Стала очевидной необходимость создания специальных инструментов для агрегации массива разнородной информации. Высокая скорость накопления мутаций в геноме возбудителя сделала необходимым изучение циркулирующих геновариантов как с позиции ускользания их от постморбидного и поствакцинального иммунитета, так и с точки зрения особенностей эпидемического процесса, вызванного отдельными вариантами вируса, и их значимости для практического здравоохранения и организации противоэпидемических мероприятий. Значительная разнородность субъектов Российской Федерации по плотности населения, географическим и экономическим условиям обусловила необходимость организации работы по осуществлению постоянного молекулярно-генетического мониторинга изменчивости возбудителя COVID-19. Активное развитие вычислительных технологий и задачи, возникшие перед системой эпидемиологического надзора в период пандемии COVID-19, сформировали предпосылки к стремительному развитию процесса цифровой трансформации в эпидемиологии. На базе ЦНИИЭ были созданы три платформы — российская система агрегации данных VGARus, SOLAR и аналитическая платформа для анализа эпидемиологической обстановки по новой коронавирусной инфекции на базе программы Superset, которые явились основными инструментами эпидемиологического мониторинга новой коронавирусной инфекции.

Ключевые слова: эпидемиология, SARS-CoV-2, COVID-19, секвенирование, молекулярно-генетический мониторинг, VGARus, SOLAR, аналитическая платформа

363

Для цитирования: Акимкин В.Г., Хафизов К.Ф., Дубоделов Д.В., Воронин Е.М., Черкашина А.С., Углева С.В., Стратулат Д.В., Самсонов А.А. Молекулярно-генетический мониторинг и технологии цифровой трансформации в современной эпидемиологии. Вестник РАМН. 2023;78(4):363–369. doi: <https://doi.org/10.15690/vramn13672>

V.G. Akimkin, K.F. Khafizov, D.V. Dubodelov, E.M. Voronin, A.S. Cherkashina,
S.V. Ugleva, D.V. Stratulat, A.A. Samsonov

Central Research Institute of Epidemiology, Moscow, Russian Federation

Molecular Genetic Monitoring and Digital Transformation Technologies in Modern Epidemiology

The pandemic of a new coronavirus infection has become a serious challenge for the health system of all countries, so as of May 2023, 766 million confirmed cases of the disease and more than 6.9 million deaths have been registered in the world. More than 22 million cases and 398 thousand deaths from a new coronavirus infection have been registered in the Russian Federation. The rapidly changing epidemiological situation necessitated the systematization of the material for analytical work. The need to create special tools for aggregating an array of heterogeneous information has become obvious. The high rate of accumulation of mutations in the genome of the pathogen made it necessary to study circulating genovariants both from the standpoint of their escape from postmorbid and postvaccinal immunity, and from the point of view of the characteristics of the epidemic process caused by individual variants of the virus and their significance for practical health care and the organization of antiepidemic measures. The significant heterogeneity in terms of population density, geographical and economic conditions of the constituent entities of the Russian Federation necessitated the organization of work on the implementation of continuous molecular genetic monitoring of the variability of the COVID-19 pathogen. The active development of computing technologies and the challenges faced by the epidemiological surveillance system during the COVID-19 pandemic created the prerequisites for the rapid development of the process of digital transformation in epidemiology. Three platforms were created on the basis of Central Research Institute of Epidemiology — the Russian data aggregation system VGARus, SOLAR and an analytical platform for analyzing the epidemiological situation regarding the new coronavirus infection based on the Superset program, which were the main tools for epidemiological monitoring of the new coronavirus infection.

Keywords: epidemiology, SARS-CoV-2, COVID-19, sequencing, molecular genetic monitoring, VGARus, SOLAR, analytical platform

For citation: Akimkin V.G., Khafizov K.F., Dubodelov D.V., Voronin E.M., Cherkashina A.S., Ugleva S.V., Stratulat D.V., Samsonov A.A. Molecular Genetic Monitoring and Digital Transformation Technologies in Modern Epidemiology. *Annals of the Russian Academy of Medical Sciences*. 2023;78(4):363–369. doi: <https://doi.org/10.15690/vramn13672>

В конце декабря 2019 г. ВОЗ проинформировала общественность о вспышке «неизвестной пневмонии» в Китае в г. Ухань, а уже 11 марта 2020 г. активное распространение новой коронавирусной инфекции COVID-19 получило статус пандемии. На сегодняшний день в мире зарегистрировано более 766 млн случаев заболевания и более 6,9 млн смертельных исходов. В Российской Федерации зарегистрировано более 22 млн случаев заболевания и более 398 тыс. летальных исходов [1]. Основными научными инструментами эпидемиологического надзора за распространением инфекции и оценки ее масштабов являются молекулярно-генетический мониторинг и технологии цифровой трансформации [2, 3].

Молекулярно-генетический мониторинг вирусных возбудителей — важное научное направление в управлении эпидемическим процессом в современных условиях [4]. Молекулярно-генетические исследования позволяют выявлять изменения в геномах вирусов, идентифицировать и классифицировать различные варианты (штаммы), а также оценивать их патогенность и трансмиссивность. Эти научные данные могут быть использованы для разработки новых диагностических наборов и современных вакцин, а также для определения оптимальных стратегий противодействия эпидемиям [5].

В случае COVID-19 молекулярно-генетический мониторинг вируса стал особенно важным направлением эпидемиологических исследований, так как это дало возможность устанавливать связи между циркулирующими геновариантами вируса и особенностями проявления эпидемического процесса. Эта информация позволила лучше понимать факторы риска распространения инфекции и разрабатывать более эффективные меры по ее контролю и предотвращению.

При этом постоянное снижение затрат на секвенирование ДНК и применение современных программных продуктов открывают новые возможности для создания и анализа баз данных о геномах вирусов [6]. Эти базы данных могут быть использованы для выявления динамики и масштабов распространения инфекции, оценки влияния демографических и социальных факторов, а также для моделирования и прогнозирования развития эпидемического процесса. Это позволяет разрабатывать более эффективные стратегии борьбы с инфекционными болезнями. Следует отметить, что ранее подобные объемы аналитических данных, методики и подходы в борьбе с эпидемиями инфекционных болезней никогда не использовались.

Активное развитие вычислительных технологий и задачи, возникшие перед системой эпидемиологического надзора в период пандемии COVID-19, создали предпосылки к стремительному развитию процесса цифровой трансформации в эпидемиологии.

Цифровая трансформация включает в себя такие основные понятия и технологические векторы, как [7, 8]:

- BIG DATA (BD) [9, 10];
- искусственный интеллект (ИИ; artificial intelligence — AI) [11];
- эластичные облачные решения (достаточно дешевые и неограниченные возможности вычисления данных любых объемов и их хранение) [12].

Важной составляющей цифровой трансформации является ИИ. Наиболее употребимыми на сегодня методами ИИ выступают машинное обучение (МО) и нейронные сети (НС) [13].

В настоящее время во ФБУН «Центральный научно-исследовательский институт Эпидемиологии» Роспотреб-

надзора (ЦНИИ Эпидемиологии) в научной и практической работе используются все указанные составляющие цифровой трансформации. Ключевым требованием цифровой трансформации в эпидемиологии является сочетание междисциплинарных навыков и знаний квалифицированных специалистов в области эпидемиологии, прикладной математики, биоинформатики, физики, химии, биологии и других смежных наук.

Технологической основой для использования современных методов работы с большими объемами данных являются разработанные в ЦНИИ Эпидемиологии три программные платформы.

1. Национальная платформа агрегирования данных о геномах вирусов SARS-CoV-2 VGARus. Платформа разработана с использованием современных ИТ-технологий (в том числе Redis, PostgreSQL, Elastic Search) и размещена в сертифицированном и защищенном центре обработки данных. При развертывании платформы применяются лучшие практики CI/CD.
2. Платформа мониторинга и анализа сведений о результатах исследований населения Российской Федерации на наличие возбудителя новой коронавирусной инфекции SOLAR. Платформа использует методы обработки структурированных данных больших объемов для получения воспринимаемого человеком результатов. В качестве базы данных используется PostgreSQL, а для аналитики данных применяется Elastic Search.
3. Аналитическая платформа для оперативного и ретроспективного эпидемиологического анализа обстановки по новой коронавирусной инфекции Superset, которая позволяет оперативно и с минимальными затратами времени (несколько часов) осуществлять оценку уровня и динамики заболеваемости COVID-19 с учетом разделения на гендерно-возрастные группы, клинические формы заболевания и особенности структуры циркулирующих на данной территории геновариантов.

В декабре 2020 г., примерно через год после первого выявления нового коронавируса, Великобритания сообщила ВОЗ об обнаружении нового варианта коронавируса SARS-CoV-2 «VOC-20212/01», принадлежность к которому определялась наличием нескольких десятков мутаций в геноме. Геновариант сначала получил название «Британский», а позже был переименован в «альфа», так как ВОЗ рекомендовала отказаться от использования наименований стран. Среди выявленных мутаций наибольшее значение имели мутации в гене S-белка N501Y, P681H и Δ69-70, которые влияли на распространение вируса, т.е. способность заражать клетки и избегать иммунного ответа организма. Этот вариант был вскоре выявлен и на территории Российской Федерации — в период подъема заболеваемости, который пришелся на конец 2020 — начало 2021 г.

Чуть позже был обнаружен также и вариант «бета», а летом 2021 г. значительный рост заболеваемости и высокое число госпитализаций были вызваны появлением варианта «дельта».

После определенного этапа эпидемиологического благополучия в декабре 2021 г. был выявлен вариант «омикрон», и его распространение способствовало резкому подъему заболеваемости в Российской Федерации, который, впрочем, завершился достаточно быстро. В целом весной 2022 г. наблюдалось некоторое эпидемиологическое благополучие, которое характеризовалось невысоким уровнем заболеваемости COVID-19, но появление субвариантов омикрона — BA.4 и особенно BA.5 — вновь

вызвало рост числа заражений в стране, который спал лишь к концу октября 2022 г.

Конец 2022 г. и начало 2023 г. характеризовались появлением высокотрансмиссивных вариантов. При этом если сначала доминировали высокотрансмиссивные субварианты омикрона BA.5, такие как BQ.1 и BQ.1.1 (так называемый «Цербер»), то далее мы наблюдали уже возврат «новых форм старых штаммов», а именно омикрона BA.2 («Стелс»), который к нам вернулся в виде рекомбинантных форм ХВВ*. На апрель 2023 г. их доля составляла более 85%, из них только ХВВ.1.5 («Кракен») — около половины.

В Российской Федерации в 2021 г. на базе ЦНИИ Эпидемиологии в соответствии с Постановлением Правительства Российской Федерации от 23 марта 2021 г. № 448 «Об утверждении Временного порядка предоставления данных расшифровки генома возбудителя новой коронавирусной инфекции (COVID-19)» была разработана и создана (свидетельство о государственной регистрации № 2021621178 от 02 июня 2021 г.) национальная платформа VGARus с целью агрегирования данных о геномах вирусов SARS-CoV-2, выявленных в Российской Федерации. Платформа позволяет осуществлять централизованный сбор и анализ данных о структуре и динамике выявленных вариантов вируса, что помогает лучше оценивать эпидемиологическую ситуацию в стране и принимать эффективные меры по ее контролю (веб-сайт платформы VGARus — genome.crie.ru) [2, 14–16].

Приказом Роспотребнадзора от 09 апреля 2021 г. № 2160 «О порядке взаимодействия федерального бюджетного учреждения науки «Центральный научно-исследовательский институт эпидемиологии» Роспотребнадзора, организаций, подведомственных федеральным органам исполнительной власти субъектов Российской Федерации, в рамках которой проводятся молекулярно-генетические исследования вирусов, и федеральных органов исполнительной власти, осуществляющих получение и анализ данных о расшифровке генома возбудителя новой коронавирусной инфекции (COVID-19)» фактически был образован научный консорциум, в который вошли научные организации Роспотребнадзора, Министерства здравоохранения Российской Федерации, ФМБА и других ведомств. На сегодняшний день к работе платформы подключены более 150 организаций, из них около 50 проводят секвенирование геномов коронавируса и депонирование последовательностей геномов в базу данных (БД) VGARus для дальнейшего анализа. В проекте участвуют Республика Армения и Республика Беларусь, что позволяет проводить мониторинг изменчивости SARS-CoV-2 на территориях не только Российской Федерации, но и соседних государств, с которыми имеется активное транспортное сообщение. Последнее особенно значимо для более полного и точного анализа эпидемиологической ситуации и принятия эффективных мер по ее контролю на большой территории.

БД VGARus представляет собой яркий пример BIG DATA и является крупнейшей в Российской Федерации базой данных в области генетических исследований возбудителей инфекционных заболеваний, служащей для целей эпидемиологического надзора в стране. Важно, что объем данных в VGARus в 5 раз превышает количество геномных последовательностей в российском сегменте международной БД GISAID (т.е. тех, у которых в качестве места забора указана Российская Федерация). Кроме того, в отличие от данных в последней, VGARus

содержит также и большой объем значимых метаданных, необходимых для эпидемиологического анализа.

В процессе работы все геномные последовательности SARS-CoV-2, полученные в рамках регулярного эпидемиологического мониторинга новой коронавирусной инфекции, загружаются в базу данных VGARus. Каждый образец в системе, помимо технической информации, информации о самой геномной последовательности, имеет данные о месте и времени взятия биоматериала, а также дополнительную метаинформацию (пол, регион, возраст, вакцинация, предположительное число контактов, сопутствующие заболевания, поездки за границу перед заболеванием), которая представляет эпидемиологическую значимость и используется для дальнейших исследований. Встроенные алгоритмы в автоматическом режиме с использованием международных и специально разработанных отечественных программ проводят анализ мутаций и идентификацию варианта SARS-CoV-2 в каждом из образцов. После загрузки нуклеотидной последовательности вируса система автоматически запускает процесс валидации (т.е. проверки на качество) последовательностей, анализ принадлежности к тому или иному варианту вируса.

Российская платформа агрегации данных о геномах вирусов VGARus стала важным инструментом в борьбе с пандемией COVID-19 и будет продолжать развиваться для предотвращения и контроля будущих инфекционных заболеваний. Она служит примером успешного сочетания научного подхода, технологической инновации и сотрудничества между организациями и специалистами различных ведомств.

Для анализа полных геномов применяется открытый и используемый мировым научным сообществом инструмент Pangolin (<https://cov-lineages.org>, <https://doi.org/10.1093/ve/veab064>). Pangolin периодически обновляется после выхода новых версий.

Для анализа фрагментных нуклеотидных последовательностей (участки или ген S-белка вируса целиком) применяется биоинформатическое средство анализа данных PARuS (государственная регистрация № 2022612845), разработанное и созданное в институте. Необходимость последнего обусловлена тем, что для фрагментного (таргетного) секвенирования, когда прочитывается лишь часть генома патогена, выявление варианта вируса существующими средствами (такими как Pangolin) затруднено. Разнообразие вариантов SARS-CoV-2 обусловлено уникальными комбинациями мутаций. Большая часть из них (и самая значимая с точки зрения детерминации штамма) находится в гене S-белка вируса. Тем не менее с появлением новых штаммов осуществлять стратификацию только по области S-гена становится все сложнее: совпадающие ключевые мутации на целевых участках и отличные мутации на непокрытых участках приводят к конфликтующим и неопределенным результатам [17].

Таким образом, специалистами-биоинформатиками ЦНИИ Эпидемиологии был доработан и создан новый алгоритм V-TRACE для выявления мутаций и отображения их принадлежности к новым эпидемиологически значимым вариантам. V-TRACE периодически дорабатывается в связи с появлением новых штаммов и мутаций.

В итоге научные результаты национальной платформы VGARus не только позволяют проводить ретроспективный анализ для объяснения причин развития пандемии в Российской Федерации, но и служат для оперативного прогнозирования эпидемиологической ситуации на ближайшие недели и месяцы, что дает возможность госу-

дарственным органам и организациям Роспотребнадзора и иных ведомств формировать стратегию мер противодействия распространению инфекции.

Российская национальная база данных геномных последовательностей возбудителя новой коронавирусной инфекции (VGARus) стала не только шагом в реализации программы импортозамещения в области IT-технологий, но и важным инструментом для обеспечения эпидемиологического благополучия и биобезопасности страны.

Благодаря введению в эксплуатацию VGARus удалось достичь значимых практических результатов, в том числе значительно сократить время между забором биоматериала и получением результата секвенирования. Множество лабораторий по всей стране осуществляют загрузку геномных данных, анализ выполняется централизованно и единообразно по всем доступным данным с возможностью оперативной выгрузки результатов, позволяющих прогнозировать эпидемиологическую обстановку. Объем накопленных геномных последовательностей SARS-CoV-2 превысил 270 тыс. по состоянию на август 2023 г., включая более 170 тыс. полных геномов вируса.

В настоящее время проводится активная работа по расширению портала VGARus как полноценного ресурса для загрузки и анализа геномных данных значимых возбудителей вирусных (а впоследствии — бактериальных, паразитарных) заболеваний, а также получения эпидемиологической информации. Информация о результатах расшифровки генома возбудителя инфекционных заболеваний будет сопровождаться сведениями санитарно-эпидемиологического характера, что выгодно отличает VGARus от традиционных международных баз данных геномной информации. Уже полностью реализован раздел для депонирования геномов вируса гепатита В, проводятся работы по добавлению раздела для вируса гепатита С, вирусов гриппа А и В, респираторно-синцициального вируса и др.

В рамках исполнения Постановления Правительства Российской Федерации от 27 марта 2021 г. № 452 «Об обеспечении уведомления физических лиц о результатах исследований на наличие возбудителей новой коронавирусной инфекции (COVID-19) с использованием федеральной государственной информационной системы «Единый портал государственных и муниципальных услуг (функций)» и обмена информацией о результатах таких исследований» на базе ЦНИИ Эпидемиологии в 2020–2022 гг. действует платформа мониторинга и анализа сведений о результатах исследований населения Российской Федерации на наличие возбудителя новой коронавирусной инфекции SOLAR.

Система использует методы обработки структурированных и неструктурированных данных огромных объемов для получения воспринимаемых человеком результатов, эффективных в условиях непрерывно увеличивающегося объема поступающих данных. Для обработки данных используются рукописные методы валидации, а также высокомасштабируемая распределенная поисковая система полнотекстового поиска и анализа данных с веб-интерфейсом, REST API и неформализованными JSON-документами, которая работает в режиме реального времени.

К платформе сегодня подключены более 1800 организаций из 85 субъектов Российской Федерации. С 01 ноября 2020 г. на Единый портал государственных услуг было передано более **180 млн** результатов исследований. Данные ежедневно пополняются в среднем на **30 тыс.** результатов исследований, в пиковые периоды 2020–2021 гг. — на **150 тыс.** ежедневно.

Результаты идентифицируются на портале госуслуг, и граждане получают информацию о результатах в цифровом виде в личном кабинете. Платформа SOLAR позволяет оперативно анализировать и предоставлять данные о распространении коронавирусной инфекции среди населения Российской Федерации, что способствует принятию своевременных решений по предотвращению и контролю инфекции.

SOLAR играет важную роль в сборе, анализе и предоставлении данных о коронавирусе, что облегчает аналитику. Система SOLAR способствует прозрачности и оперативности предоставления системной информации о коронавирусной инфекции, что улучшает сотрудничество между учреждениями, учеными и специалистами в области здравоохранения.

В дальнейшем платформа SOLAR может быть расширена для мониторинга и анализа данных о других инфекционных болезнях, что усилит ее роль в обеспечении здоровья населения. Разработка и внедрение новых алгоритмов и инструментов анализа данных на платформе SOLAR позволяют обеспечить еще большую точность и оперативность принятия решений для борьбы с инфекционными болезнями.

Платформа мониторинга и анализа сведений о результатах исследований на наличие возбудителя новой коронавирусной инфекции SOLAR является важным современным инструментом в борьбе с пандемией COVID-19 в России. Благодаря использованию передовых технологий и методов анализа данных SOLAR позволяет оперативно собирать, анализировать и предоставлять информацию о распространении новой коронавирусной инфекции на территории всех субъектов Российской Федерации, что, в свою очередь, облегчает принятие своевременных и обоснованных решений специалистами Роспотребнадзора, местными властями и медицинскими организациями. Кроме того, система SOLAR обеспечивает гражданам Российской Федерации доступ к информации о результатах своих исследований в удобном цифровом формате через портал госуслуг, что свидетельствует о полном выполнении требований Постановления Правительства Российской Федерации № 452 от 27 марта 2021 г.

Для решения задач оперативного и ретроспективного анализа необходимо применение комплекса программных средств для автоматизации сбора, хранения, анализа и интерпретации сведений эпидемиологического характера. В роли ключевого элемента (платформы) такой системы может выступать BI-система (Business Intelligence). Такая платформа позволяет объединить данные из других программных платформ, а также информацию из отчетных форм различных ведомств и открытых источников и предоставить доступ к ней в виде, пригодном для проведения эпидемиологического анализа, специалистам, осуществляющим эпидемиологический надзор за новой коронавирусной инфекцией.

В ЦНИИ Эпидемиологии в период пандемии была разработана и успешно используется аналитическая платформа для анализа эпидемиологической обстановки по новой коронавирусной инфекции. Платформа построена на базе BI-системы Apache Superset — программного обеспечения с открытым исходным кодом для исследования и визуализации данных. Все загружаемые данные и программный код находятся на серверах BIG DATA института, что обеспечивает их сохранность и контролируемый доступ.

Входными данными для платформы являются данные отчетных форм Роспотребнадзора, данные по заболевае-

мости Росстата и данные платформы VGARus по генетическому разнообразию образцов коронавируса SARS-CoV-2. Данные на аналитическую платформу автоматически загружаются с необходимой периодичностью (в основном один раз в сутки) и группируются в соответствии с решаемой задачей. Статистические показатели, необходимые для анализа, рассчитываются в фоновом режиме. Также показатели могут быть сгруппированы по субъектам или федеральным округам.

В настоящее время на платформе для анализа доступны более **1292 млн** загруженных данных о количестве заболевших, данных эпидемиологического анамнеза, сведений о клинической форме и степени тяжести заболевания, исходах. Платформа динамично пополняется новыми сведениями — более чем **14 тыс.** показателей ежедневно.

Использование такого подхода обеспечивает решение следующих задач:

- система объединяет сведения из большого количества разных источников;
- предоставляет единую точку доступа к сведениям эпидемиологического характера и необходимым справочным данным (количество источников данных растет, они изменяются, все привлекаемые к работе специалисты должны иметь все актуальные сведения);
- решает проблему различия форм представления данных в используемых источниках;
- эпидемиологи освобождаются от необходимости предварительной обработки и структурирования данных;
- статистические показатели рассчитываются автоматически, без участия исследователя;
- акцент в представлении данных может быть сделан как на элементы визуализации, так и на табличные данные;
- обработанные и отфильтрованные через аналитическую платформу данные могут быть использованы для дальнейшего анализа с использованием методов МО и НС, загружены для дальнейшего анализа в программы статистической обработки;
- платформа позволяет совмещать различные способы отображения информации, в том числе с использованием геоданных;
- возможно расширение применяемого для анализа инструментария за счет открытости платформы.

Использование аналитической платформы позволяет значительно увеличить информативность и снизить затраты времени на проведение эпидемиологического анализа. Так, анализ эпидемиологической обстановки по заболеваемости новой коронавирусной инфекцией на территории всех субъектов Российской Федерации занимает всего несколько часов.

Использование VI-систем при проведении оперативного и ретроспективного эпидемиологического анализа позволяет внести в работу элемент интерактивности (система меняет параметры отображаемых данных в ответ на действия пользователя), помогает исследователю задавать вопросы и находить на них ответы максимально просто, смещает фокус внимания исследователя со сбора и предварительной обработки данных на их анализ.

Ведется разработка аналогичной аналитической платформы для анализа эпидемиологической обстановки по вирусным гепатитам. Полученный в ходе работы опыт может быть использован в организации системы эпидемиологического надзора за другими инфекционными заболеваниями.

Помимо перечисленных программных платформ, в ЦНИИ Эпидемиологии в практической деятельности активно используются технологии ИИ. На сегодняшний день это преимущественно набор методов МО, методы агломеративной иерархической кластеризации, современные методики регрессионного анализа с расширенным набором статистических методов, эпидемиолого-экономический анализ.

В качестве примеров можно привести такие методы, как:

- методы прогнозирования темпов заболеваемости COVID-19 в странах мира;
- кластеризация регионов Российской Федерации по заболеваемости COVID-19;
- взаимосвязь заболеваемости COVID-19 в кластерах регионов России с их социально-экономическими характеристиками;
- прогнозирование заболеваемости населения г. Москвы COVID-19 математическими методами;
- оценка факторов риска развития пневмоний при COVID-19 методами МО;
- прогнозирование исходов заболевания COVID-19 у детей методами МО;
- экономическая значимость выявления и лабораторного подтверждения наследуемого хромосомно-интегрируемого вируса герпеса человека 6A/B у новорожденных;
- эпидемиологический анализ гендерно-возрастной структуры госпитализируемых с COVID-19 и их летальности за 2020–2021 гг.;
- разработка методики оценки и оценка предотвращенного экономического ущерба от COVID-19 при использовании наборов реагентов на основе МАНК производства ЦНИИ Эпидемиологии Роспотребнадзора;
- разработка методики расчета и расчет примерного экономического ущерба от COVID-19 для Российской Федерации в целом за 2020–2022 гг.;
- анализ значимости интернет-запросов, связанных с обонянием, как прогностических маркеров изменения динамики заболеваемости COVID-19 [18];
- оценка возможности эпидемиологического прогнозирования динамики заболеваемости COVID-19 на основе анализа интернет-запросов «Антибиотики» [19].

Заключение

Таким образом, пандемия COVID-19 явилась мощным импульсом к развитию современной эпидемиологии, технологий цифровой трансформации данных, к изучению генетических свойств возбудителей вирусных инфекций, что позволило в кратчайшие сроки создать отечественные диагностические тест-системы, вакцинные препараты, разработать и применить на практике оперативные эффективные модели управления эпидемиологическим процессом.

Дополнительная информация

Источник финансирования. Рукопись подготовлена и опубликована за счет финансирования по месту работы авторов.

Конфликт интересов. Авторы данной статьи подтвердили отсутствие конфликта интересов, о котором необходимо сообщить.

Участие авторов. В.Г. Акимкин — разработка концепции исследования, руководство научным исследованием; К.Ф. Хафизов — разработка концепции исследования, подбор методик, написание статьи; Д.В. Дубоделов — разработка концепции исследования, подбор методик, написание статьи; Е.М. Воронин — разработка концепции исследования, написание статьи; А.С. Черкашина — разработка концепции исследования, подбор

методик, написание статьи; С.В. Углева — написание и редактирование статьи, работа с редакцией журнала; Д.В. Стратулат — разработка концепции исследования; А.А. Самсонов — разработка концепции исследования, написание статьи. Все авторы внесли значимый вклад в проведение поисково-аналитической работы и подготовку рукописи, прочли и одобрили финальную версию текста перед публикацией.

ЛИТЕРАТУРА

1. *Weekly epidemiological update on COVID-19 — 18 May 2023*. Ed. 143. Available online: <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---18-may-2023>. Accessed on August 30, 2023.
2. Акимкин В.Г., Семенов Т.А., Углева С.В. и др. COVID-19 в России: эпидемиология и молекулярно-генетический мониторинг // *Вестник РАМН*. — 2022. — Т. 77. — № 4. — С. 254–260. — doi: <https://doi.org/10.15690/vramn2121> [Akimkin VG, Semenenko TA, Ugleva SV, et al. COVID-19 in Russia: epidemiology and molecular genetic monitoring. *Annals of the Russian Academy of Medical Sciences*. 2022;77(4):254–260. (In Russ).] doi: <https://doi.org/10.15690/vramn2121>
3. Ступак В.С., Зубко А.В., Манюшкина Е.М. и др. Здравоохранение России в период пандемии COVID-19: вызовы, системные проблемы и решение первоочередных задач // *Профилактическая медицина*. — 2022. — Т. 25. — № 11. — С. 21–27. — doi: <https://doi.org/10.17116/profmed20222511121> [Stupak VS, Zubko AV, Manoshkina EM, et al. Healthcare in Russia during the COVID-19 pandemic: challenges, systemic issues, and addressing priorities. *Profilakticheskaya Meditsina*. 2022;25(11):21–27. (In Russ).] doi: <https://doi.org/10.17116/profmed20222511121>
4. Hill V, Ruis C, Bajaj S, et al. Progress and challenges in virus genomic epidemiology. *Trends Parasitol*. 2021;37(12):1038–1049. doi: 10.1016/j.pt.2021.08.007
5. Attwood SW, Hill SC, Aanensen DM, et al. Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. *Nat Rev Genet*. 2022;23(9):547–562. doi: <https://doi.org/10.1038/s41576-022-00483-8>
6. Khare S, Gurry C, Freitas L, et al. GISAID’s Role in Pandemic Response. *China CDC Weekly*. 2021;3(49):1049–1051. doi: <https://doi.org/10.46234/ccdcw2021.255>
7. Catlin T, Lorenz JT, Sternfels B, Willmott P. A Roadmap for a Digital Transformation. In: *McKinsey & Company*. March 2017. Available online: <https://www.mckinsey.com/industries/financial-services/our-insights/roadmap-for-a-digital-transformation>. Accessed on August 30, 2023.
8. Digital, Technology, and Data. In: *Boston Consulting Group*. 2023. Available online: <https://www.bcg.com/en-us/digital-bcg/overview.aspx>. Accessed on August 30, 2023.
9. Laney D. *3D Data Management: Controlling Data Volume, Velocity and Variety*. META Group. February 6, 2001. 4 p. Available online: <https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--an...> Accessed on August 30, 2023.
10. ГОСТ Р ИСО/МЭК 20546-2021. Национальный стандарт Российской Федерации. *Информационные технологии. Большие данные. Обзор и словарь*. [All-Union State Standard Р ИСО/МЭК 20546-2021. National standard of the Russian Federation. *Information technology. Big data. Overview and vocabulary*. (In Russ).]
11. ГОСТ 33707-2016 (ISO/IEC 2382:2015). Межгосударственный стандарт. *Информационные технологии. Словарь*. [All-Union State Standard 33707-2016 (ISO/IEC 2382:2015). Interstate standard. *Information technologies. Vocabulary*. (In Russ).]
12. ГОСТ ISO/IEC 17788-2016. Межгосударственный стандарт. *Информационные технологии. Облачные вычисления. Общие положения и терминология*. [All-Union State Standard ISO/IEC 17788-2016. Interstate standard. *Information technology. Cloud computing. Overview and vocabulary*. (In Russ).]
13. ГОСТ Р 59895-2021. Национальный стандарт Российской Федерации. *Технологии искусственного интеллекта в образовании. Общие положения и терминология*. [All-Union State Standard Р 59895-2021. National standard of the Russian Federation. *Artificial intelligence technologies in education. General provisions and terminology*. (In Russ).]
14. Акимкин В.Г., Попова А.Ю., Плоскирева А.А. и др. COVID-19: эволюция пандемии в России. Сообщение I: проявления эпидемического процесса COVID-19 // *Журнал микробиологии, эпидемиологии и иммунобиологии*. — 2022. — Т. 99. — № 3. — С. 269–286. doi: <https://doi.org/10.36233/0372-9311-276> [Akimkin VG, Popova AYU, Ploskireva AA, et al. COVID-19: the evolution of the pandemic in Russia. Report I: manifestations of the COVID-19 epidemic process. *Journal of microbiology, epidemiology and immunobiology = Zhurnal mikrobiologii, epidemiologii i immunobiologii*. 2022;99(3):269–286. (In Russ).] doi: <https://doi.org/10.36233/0372-9311-276>
15. Акимкин В.Г., Попова А.Ю., Хафизов К.Ф. и др. COVID-19: эволюция пандемии в России. Сообщение II: динамика циркуляции геновариантов вируса SARS-CoV-2 // *Журнал микробиологии, эпидемиологии и иммунобиологии*. — 2022. — Т. 99. — № 4. — С. 381–396. — doi: <https://doi.org/10.36233/0372-9311-295> [Akimkin VG, Popova AYU, Khafizov KF, et al. COVID-19: evolution of the pandemic in Russia. Report II: dynamics of the circulation of SARS-CoV-2 genetic variants. *Journal of microbiology, epidemiology and immunobiology = Zhurnal mikrobiologii, epidemiologii i immunobiologii*. 2022;99(4):381–396. (In Russ).] doi: <https://doi.org/10.36233/0372-9311-295>
16. Meyerowitz EA, Richterman A, Gandhi RT, Sax PE. Transmission of SARS-CoV-2: A Review of Viral, Host, and Environmental Factors. *Ann Intern Med*. 2021;174(1):69–79. doi: <https://doi.org/10.7326/M20-5008>
17. O’Toole Á, Pybus OG, Abram ME, et al. Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. *BMC Genomics*. 2022;23(1):121. doi: <https://doi.org/10.1186/s12864-022-08358-2>
18. Хорошун Д.К., Момыналиев К.Т., Воронин Е.М., Акимкин В.Г. Анализ поисковых запросов в «Яндекс», связанных с COVID-19 в Российской Федерации // *Медицинский алфавит*. — 2022. — № 18. — С. 14–22. doi: <https://doi.org/10.33667/2078-5631-2022-18-14-22> [Khoroshun DK, Momynaliev KT, Voronin EM, Akimkin VG. Analysis of Yandex search queries related to COVID-19 in Russian Federation. *Medical alphabet*. 2022;(18):14–22. (In Russ).] doi: <https://doi.org/10.33667/2078-5631-2022-18-14-22>
19. Акимкин В.Г., Тутельян А.В., Шулакова Н.И., Воронин Е.М. Пандемия COVID-19: новый виток нарастания антибиотикоре-

зистентности // *Инфекционные болезни*. — 2021. — Т. 19. — № 3. — С. 133–138. — doi: <https://doi.org/10.20953/1729-9225-2021-3-133-138>
[Akimkin VG, Tutelyan AV, Shulakova NI, Voronin EM. Панде-

мия COVID-19 pandemic: a new round of antibiotic resistance. *Infectious diseases*. 2021;19(3):133–138. (In Russ.) doi: <https://doi.org/10.20953/1729-9225-2021-3-133-138>

КОНТАКТНАЯ ИНФОРМАЦИЯ

Углева Светлана Викторовна, д.м.н., профессор [*Svetlana V. Ugleva*, MD, PhD, Professor];
адрес: 111123, Москва, ул. Новогиреевская, д. 3а [address: 3a Novogireevskaya Str., Moscow, 111123, Russian Federation]; e-mail: uglevas@bk.ru; eLibrary SPIN: 8840-5814; ORCID: <https://orcid.org/0000-0002-1322-0155>

Акимкин Василий Геннадьевич, д.м.н., профессор, академик РАН [*Vasily G. Akimkin*, MD, PhD, Professor, Academician of the RAS]; e-mail: vgakimkin@yandex.ru; eLibrary SPIN: 4038-7455, ORCID: <https://orcid.org/0000-0003-4228-9044>

Хафизов Камил Фаридович, к.б.н. [*Kamil F. Khafizov*, PhD in Biology]; e-mail: khafizov@cmd.su, eLibrary SPIN: 9082-5749, ORCID: <http://orcid.org/0000-0001-5524-0296>

Дубоделов Дмитрий Васильевич, к.м.н., с.н.с. [*Dmitry V. Dubodelov*, MD, PhD, Senior Research Associate]; e-mail: dubodelov@cmd.su, eLibrary SPIN: 4860-7909, ORCID: <https://orcid.org/0000-0003-3093-5731>

Воронин Евгений Михайлович, к.м.н. [*Evgeny M. Voronin*, MD, PhD]; ORCID: <https://orcid.org/0000-0001-5925-7757>

Черкашина Анна Сергеевна, к.х.н. [*Anna S. Cherkashina*, PhD in Chemistry]; e-mail: cherkashina@pcr.ms, eLibrary SPIN: 7854-7358, ORCID: <https://orcid.org/0000-0001-7970-7495>

Стратулат Денис Валерьевич [*Denis V. Stratulat*]; e-mail: stratylat@cmd.su, ORCID: <https://orcid.org/0000-0003-0988-4466>

Самсонов Александр Андреевич [*Alexander A. Samsonov*]; e-mail: samsonov@cmd.su, ORCID: <https://orcid.org/0009-0007-8036-1883>