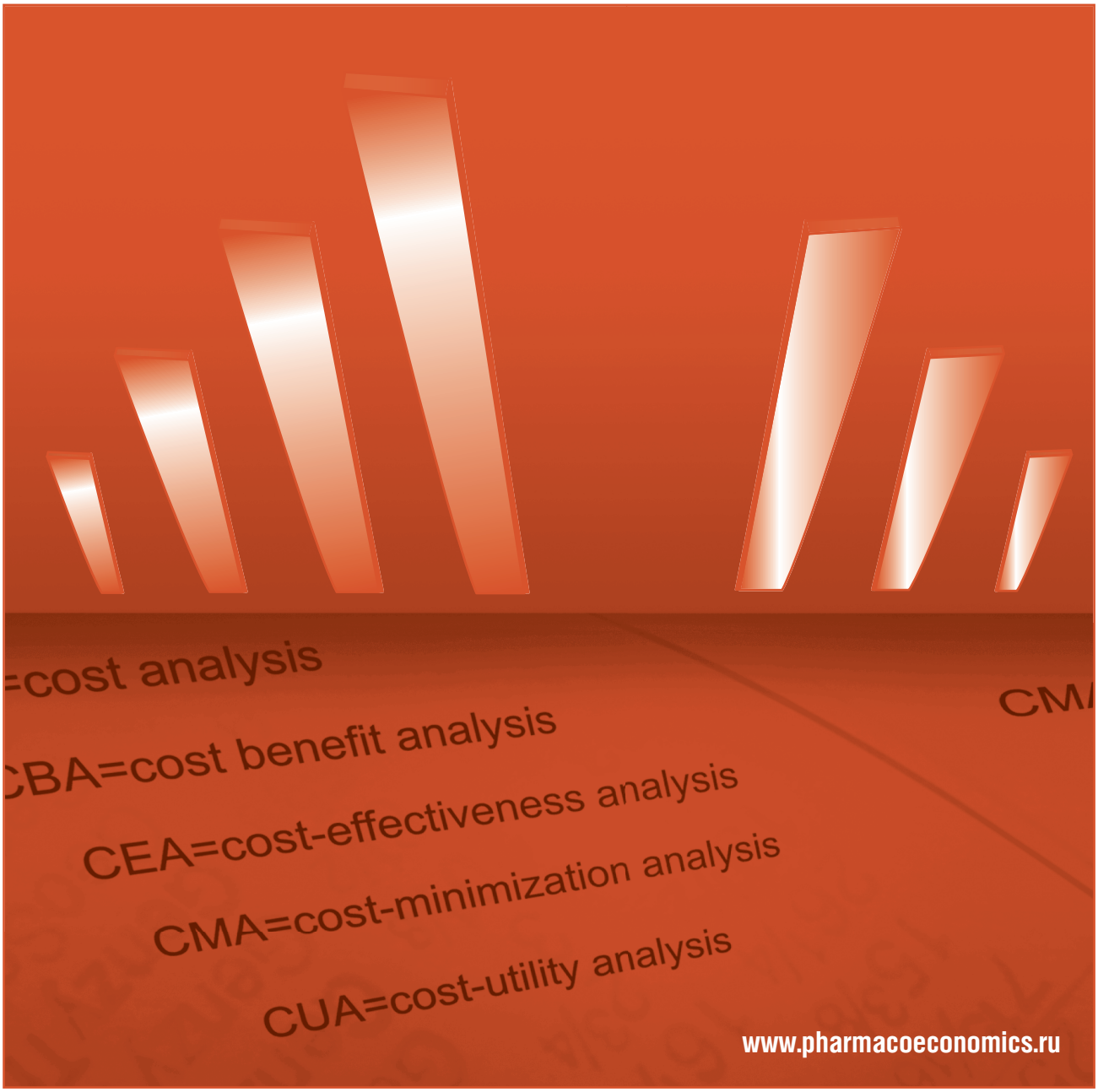
ISSN 2070-4909 (print)
ISSN 2070-4933 (online)

Фармакоэкономика

Современная фармакоэкономика и фармакоэпидемиология



FARMAKOEKONOMIKA

Modern Pharmacoeconomics and Pharmacoepidemiology

2021 Vol. 14 No. 3

№3

Том 14

2021

Данная интернет-версия статьи была скачана с сайта http://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях.
Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru

BY-NC-SA Check for updates

https://doi.org/10.17749/2070-4909/farmakoekonomika.2021.108
ISSN 2070-4909 (print)
ISSN 2070-4933 (online)

Модель прогнозирования пандемии COVID-19 на основе машинного обучения в отдельных регионах Российской Федерации

Гаврилов Д.В.[1], Абрамов Р.В.[1], Кирилкина А.В.[2], Ившин А.А.[3], Новицкий Р.Э.[1]

[1] *Общество с ограниченной ответственностью «К-Скай» (наб. Варкауса, д. 17, помещ. 62, Республика Карелия, Петрозаводск 185031, Россия)*

[2] *Государственное бюджетное учреждение здравоохранения «Республиканская инфекционная больница» (ул. Кирова, д. 42, Петрозаводск 185035, Россия)*

[3] *Федеральное государственное бюджетное образовательное учреждение высшего образования «Петрозаводский государственный университет» (пр-т Ленина, д. 33, Республика Карелия, Петрозаводск 185910, Россия)*

Для контактов: *Ившин Александр Анатольевич, e-mail: scipeople@mail.ru*

РЕЗЮМЕ

Актуальность. Прогнозирование распространения новой коронавирусной инфекции (COVID-19) имеет важное значение для принятия своевременных системных профилактических и противоэпидемических мер как на региональном, так и на федеральном уровне с целью снижения заболеваемости и смертности.

Цель: разработать модель краткосрочного прогнозирования зараженных и умерших от COVID-19 в Российской Федерации.

Материал и методы. Данные для обучения модели собраны с портала Стопкоронавирус.рф и ресурса Университета Джонса Хопкинса. Она включает 13 признаков для оценки динамики заражения и летальности, а также скорости их прироста в разных странах и отдельных регионах Российской Федерации. Модель обучена методом градиентного бустинга CatBoost и ежедневно переобучается на обновленных данных.

Результаты. Создана модель краткосрочного предсказания числа зараженных и умерших от COVID-19 на период до 14 дней. Оценка точности модели с учетом ошибки предсказания в процентах (англ. mean absolute percentage error, MAPE) составляет от 2,3% до 24% для 85 регионов России. Показано преимущество метода машинного обучения CatBoost перед линейной регрессией на примере величины среднеквадратичной ошибки (англ. root mean square error, RMSE). Модель показывает меньшую ошибку для регионов с большой численностью населения, чем для менее населенных областей.

Заключение. Модель может быть использована не только для прогнозирования пандемии новой коронавирусной инфекции, но и для контроля и оценки распространения заболеваний из группы новых инфекций на этапах их возникновения, пика заболеваемости и периода стабилизации.

КЛЮЧЕВЫЕ СЛОВА

Искусственный интеллект, машинное обучение, градиентный бустинг, эпидемиологический прогноз, пандемия COVID-19.

Статья поступила: 19.08.2021 г.; **в доработанном виде:** 30.08.2021 г.; **принята к печати:** 03.09.2021 г.; **опубликована онлайн:** 04.09.2021 г.

Конфликт интересов

Авторы заявляют об отсутствии необходимости раскрытия конфликта интересов в отношении данной публикации.

Финансирование

Исследование выполнено при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках Соглашения № 075-15-2021-665 на уникальной научной установке «Многокомпонентный программно-аппаратный комплекс для автоматизированного сбора, хранения, разметки научно-исследовательских и клинических биомедицинских данных, их унификации и анализа на базе ЦОД с использованием технологий искусственного интеллекта» (регистрационный номер 2075518).

Вклад авторов

Все авторы сделали эквивалентный вклад в подготовку публикации.

Для цитирования

Гаврилов Д.В., Абрамов Р.В., Кирилкина А.В., Ившин А.А., Новицкий Р.Э. Модель прогнозирования пандемии COVID-19 на основе машинного обучения в отдельных регионах Российской Федерации. *ФАРМАКОЭКОНОМИКА. Современная фармакоэкономика и фармакоэпидемиология.* 2021; 14 (3): 342–356. https://doi.org/10.17749/2070-4909/farmakoekonomika.2021.108.

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

COVID-19 pandemic prediction model based on machine learning in selected regions of the Russian Federation

Gavrilov D.V.[1], Abramov R.V.[1], Kirilkina A.V.[2], Ivshin A.A.[3], Novitskiy R.E.[1]

[1] *K-SkAI LLC (17 premises 62 Varkaus Qy, Republic of Karelia, Petrozavodsk 185031, Russia)*

[2] *Republican Infectious Diseases Hospital (42 Kirov Str., Republic of Karelia, Petrozavodsk 185035, Russia)*

[3] *Petrozavodsk State University (33 Lenin Ave., Republic of Karelia, Petrozavodsk 185910, Russia)*

Corresponding author: *Aleksandr A. Ivshin, e-mail: scipeople@mail.ru*

SUMMARY

Background. Prediction of the new coronavirus infection (COVID-19) spread is important to take timely measures and initiate systemic preventive and anti-epidemic actions both at the regional and state levels to reduce morbidity and mortality.

Objective: to develop a model for short-term forecasting of COVID-19 cases and deaths in the Russian Federation.

Material and methods. The data for the model training were collected from the Stopcoronavirus.rf and Johns Hopkins University portals. It included 13 features to assess the infection dynamics and mortality, as well as the rate of morbidity and mortality in different countries and certain regions of the Russian Federation. The model was trained by the CatBoost gradient boosting method and retrained daily with updated data.

Results. The forecast model of COVID-19 cases and deaths for the period of up to 14 days was created. The mean absolute percentage error (MAPE) estimate of the model's accuracy ranged from 2.3% to 24% for 85 regions of the Russian Federation. The advantage of the CatBoost machine learning method over linear regression was shown using the example of the root mean square error (RMSE) value. The model showed less error for regions with a large population than for less populated ones.

Conclusion. The model can be used not only to predict the pandemic of the novel coronavirus infection but also to control and assess the spread of diseases from the group of new infections at their emergence, peak incidence, and stabilization period.

KEYWORDS

Artificial intelligence, machine learning, gradient boosting, epidemiological forecast, COVID-19 pandemic.

Received: 19.08.2021; **in the revised form:** 30.08.2021; **accepted:** 03.09.2021; **published online:** 04.09.2021

Conflict of interests

The authors declare they have nothing to disclose regarding the conflict of interests with respect to this manuscript.

Funding

The research was carried out with the financial support of the Ministry of Science and Higher Education of the Russian Federation under the Agreement No. 075-15-2021-665 at the unique scientific installation "Multicomponent software and hardware complex for automated collection, storage, marking of research and clinical biomedical data, their unification and analysis on the basis of the Data Processing Center using artificial intelligence technologies" (registration number 2075518).

Author's contrubution

The authors contributed equally to this article.

For citation

Gavrilov D.V., Abramov R.V., Kirilkina A.V., Ivshin A.A., Novitskiy R.E. COVID-19 pandemic prediction model based on machine learning in selected regions of the Russian Federation. *FARMAKOEKONOMIKA. Sovremennaya farmakoekonomika i farmakoepidemiologiya / FARMAKOEKONOMIKA. Modern Pharmacoeconomics and Pharmacoepidemiology.* 2021; 14 (3): 342–356 (in Russ.). https://doi.org/10.17749/2070-4909/farmakoekonomika.2021.108.

ВВЕДЕНИЕ / INTRODUCTION

В условиях пандемии новой коронавирусной инфекции (COVID-19) ключевой задачей является недопущение критической перегрузки системы здравоохранения за счет экспоненциального и непредсказуемого роста госпитализации больных и, как следствие, масштабного увеличения запросов на использование ограниченных медицинских ресурсов [1–6]. Важность данной задачи акцентируется на высшем государственном уровне. Подчеркивается необходимость прогноза пандемии новой коронавирусной инфекции, в особенности в краткосрочном периоде, с непрерывным обновлением данных прогноза в режиме реального времени исходя из объективно складывающейся обстановки в регионах [7].

Для определения рисков скорости распространения COVID-19, оптимального распределения ресурсов, принятия управленческих решений на федеральном, региональном и муниципальном уровнях необходим точный прогноз распространения инфекции. Показателями для прогнозирования с позиции эффективных противоэпидемических мер являются:

– прогноз числа зараженных;
– прогноз числа умерших;
– прогноз скорости развития эпидемии;
– прогноз нагрузки на медицинские организации.

При прогнозировании пандемии важно учитывать региональные особенности – численность населения и его плотность, которые напрямую влияют на скорость распространения инфекции и смертность от заболевания, а также готовность региональной системы здравоохранения обеспечить своевременную и полноценную помощь населению.

Технологии искусственного интеллекта, в том числе машинное обучение, являются одним из перспективных направлений для

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

Основные моменты

Что уже известно об этой теме?

- С 2019 г. по настоящее время COVID-19 распространяется и мутирует, имеет сложно предсказываемые волны подъема заболеваемости
- Используемые модели для прогнозирования распространения инфекций по типу SEIR не включают особенностей стран и ограничительных мер, также им сложно работать с многообразием штаммов из-за различных коэффициентов смертности и заболеваемости, что влияет на предсказательную точность. Модели, построенные при помощи машинного обучения, не имеют этих ограничений и потенциально обладают лучшей прогностической точностью
- Исследования предсказания течения инфекционного процесса немногочисленны, а комплексный подход к прогнозированию (заболеваемость, смертность, региональная особенности) применялся редко

Что нового дает статья?

- Поэтапно и детально описано создание модели машинного обучения для предсказания течения COVID-19 с итоговым использованием метода CatBoost для оценки заболеваемости и летальности по регионам РФ
- Модель комплексно оценивает различные прогностические признаки, используя не только динамику заболеваемости и смертности, но и демографические характеристики местности
- Показана возможность оценки технологии машинного обучения для краткосрочного прогнозирования развития событий, связанных с COVID-19

Как это может повлиять на клиническую практику в обозримом будущем?

- Благодаря построению модели при помощи машинного обучения нет необходимости в подборе коэффициентов для дифференциальных уравнений, а необходима лишь история количества заболевших. Это позволяет получить быстрый и точный краткосрочный прогноз течения инфекционного процесса
- Представленный алгоритм способен оценивать тренды течения COVID-19 как в целом, так и отдельно для каждого региона РФ
- Модель может быть использована не только для прогнозирования пандемии COVID-19, но и для контроля и оценки распространения заболеваний из группы новых инфекций на этапах их возникновения, пика заболеваемости и периода стабилизации

Highlights

What is already known about the subject?

- From 2019 to the present day, COVID-19 is spreading and mutating, and it is difficult to predict waves of an increase in morbidity
- The SEIR type models used to predict the spread of infections do not include the specifics of countries and restrictive measures, and it is difficult for them to work with a variety of strains either due to different mortality and morbidity rates, which affects the predictive accuracy. Models built using machine learning do not have these limitations and potentially have better predictive accuracy
- Studies on the prediction of the infectious process are few, and an integrated approach to forecasting (morbidity, mortality, regional characteristics) had rarely been used

What are the new findings?

- The creation of a machine learning model for predicting the course of COVID-19 is described in stages and in detail, with the final use of CatBoost method for assessing morbidity and mortality in the regions of the RF
- The model comprehensively evaluates various prognostic signs using not only the dynamics of morbidity and mortality, but also demographic characteristics of the area
- The possibility of evaluating machine learning technology was shown for short-term forecasting of the development of events related to COVID-19

How might it impact the clinical practice in the foreseeable future?

- Thanks to building a model using machine learning, there is no need to select coefficients for differential equations. Only a history of the number of cases is needed. This makes it possible to predict the course of the infectious process in the short term quickly and accurately
- The presented algorithm can evaluate the trends of COVID-19 course both as a whole and separately for each region of the RF
- The model can be used not only to predict the COVID-19 incidence, but also to control and assess the spread of diseases from the group of new infections at the stages of their emergence, peak incidence, and stabilization period

предикции заболеваний и построения прогнозов, в том числе через разработку и создание прогностических моделей [8–11]. В настоящее время в мире накоплен большой опыт создания различных предиктивных моделей для COVID-19. Исследователи, решающие эту задачу, использовали для прогнозирования распространения эпидемии многократно проверенные модели или же создавали новые модели и алгоритмы на основе различных математических подходов.

В проведенных масштабных исследованиях [12–16] была успешно использована модель SEIR (англ. Susceptible, Exposed, Infected, Recovered) – инструмент для прогнозирования эпидемий и применения противоэпидемических мер. Моделирование SEIR осуществляется с использованием дифференциальных уравнений, и для каждой из четырех групп населения (предрасположенные, вирусоносители, больные, выздоровевшие) создаются формулы расчета перехода из одной группы в другую. Подобный подход можно использовать для предсказания развития заражений на популяционном уровне, например для страны в целом, но с его помощью довольно сложно оценить эффективность ограничительных мер для регионов с небольшой численностью населения. Вместе с тем для такой недостаточно изученной болезни, как COVID-19, некоторые значимые переменные неизвестны или малоизвестны, поэтому их оценка на начальном этапе может привести к росту ошибок модели. В исследовании G. Pandey et al. были получены значения среднеквадратичной логарифмической ошибки (англ. root mean squared logarithmic error, RMSLE) в сравнении с фактическими данными: для SEIR – 1,52, для полиномиальной регрессии – 1,75 [17]. В работе российских исследователей А.В. Алешукина и др. использовалась интерактивная SEIR-модель и продемонстрирован успешный опыт ее применения для предсказания эпидемиологической ситуации по COVID-19 в Ростовской обл. на семнадцати тысячах заболевших [18].

В работе D. Fanelli и F. Piazzab использован аналог модели SEIR – модель SIRD (англ. Susceptible, Infectious, Recovered, Deceased) на дифференциальных уравнениях для предсказания количества предрасположенных, зараженных, выздоровевших и умерших в исследованной популяции. Данная модель позволяет определять общую тенденцию распространения болезни, но для отдельно взятого периода показывает неустойчивые результаты [13].

На основе данных заболеваемости населения Китая исследователи Z. Yang et al. [19] успешно переобучили модель SEIR и создали вторую модель на основе метода долгой краткосрочной памяти (англ. long short-term memory, LSTM), который использует рекуррентную нейронную сеть для прогнозирования числа зараженных. Отличительной особенностью LSTM по сравнению с SEIR является отсутствие необходимости строить дифференциальные уравнения, но модель показала нестабильность и сложна в интерпретации.

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

Для прогнозирования эпидемии в Нью-Йорке была разработана модель с использованием методов экспоненциального сглаживания, которая краткосрочно, до 10 дней, предсказывает распространение COVID-19, а также дает прогноз по зараженным в период завершения пандемии, когда существенно замедляется прирост новых случаев [20].

В работах российских авторов Д.В. Мелик-Гусейнова и др. [21] представлены модели линейной регрессии для предсказания количества летальных исходов от новой коронавирусной инфекции на примере одного региона. Для предикции числа зараженных [22] в различных регионах Российской Федерации (РФ) использован метод градиентного бустинга CatBoost [23] с RMSLE 0,37, что дает возможность применения данного метода обучения для составления устойчивых прогнозов.

Результаты приведенных исследований демонстрируют возможность использования машинного обучения для предсказания количества зараженных и смертельных случаев от COVD-19 по отдельным регионам и странам в краткосрочном периоде.

МАТЕРИАЛ И МЕТОДЫ / MATERIAL AND METHODS

Исследование охватывает период распространения COVID-19 с 03.02.2020 г. (начало представления данных от источников). Сведения о количестве зараженных и умерших от COVID-19 собраны по 85 регионам РФ с портала Стопкоронавирус.рф [24] (примеры приведены в **таблицах 1, 2**). Информация по 178 странам (далее – другие страны) получена из публичных отчетов Университета Джонса Хопкинса (США) [25]. В набор данных были включены демографические параметры: общая численность населения, численность городского населения и плотность населения [26]. Достоверных сведений об использовании защитных масок и уровне самоизоляции найдено не было.

Для каждого из 108 дней были рассчитаны две базовые характеристики: количество новых зараженных и умерших и скорость прироста зараженных и умерших. Количество новых зараженных ля i-го дня (где i – номер дня начиная с 03.02.2020 г. (i = 0)) вычисляли как разность между числом зараженных за день i и числом зараженных за день i – 1 (i минус 1). Скорость прироста новых зараженных для i-го дня рассчитывали как разность между числом новых зараженных за день i и числом новых зараженных за день i – 1. Также были вычислены следующие параметры: количество дней после первого заражения и количество дней после 10^n (n = 0, 1, 2, 3, 4, 5) заражений в регионе. Аналогичным образом были рассчитаны характеристики для умерших. Полное описание данных содержится в **таблице 3**.

Полученный набор для первичного обучения содержал 45 признаков и данные по 108 дням наблюдения для всех регионов РФ и других стран – всего 36 056 записей. Все поля были прологарифмированы с целью уменьшения стандартного отклонения. Количество новых зараженных и умерших в день i были взяты в качестве меток. Тестовый набор состоял из четырех последних дней, все остальные были включены в тренировочный.

Таблица 1. Статистические параметры заражений SARS-CoV-2 на примере четырех регионов Российской Федерации (данные за период с 03.02.2020 г. по 17.06.2020 г.)

Table 1. Statistical parameters of disease cases of SARS-CoV-2 in four regions of the Russian Federation (data for the period of 03.02.2020–17.06.2020)

Регион / Region	Количество измерений (дни), n / Number of measurements (days), n	Максимальное количество, n / Maximal number, n	Среднее количество, n / Mean number, n	Средний ежедневный прирост, n / Mean daily increase, n	Стандартное отклонение / Standard deviation
Москва / Moscow	108	6703	1760	2016	1684
Санкт-Петербург / Saint Petersburg	108	541	216	206	150
Республика Карелия / Republic of Karelia	108	60	14	11	15
Ямало-Ненецкий автономный округ / Yamalo-Nenets Autonomous District	108	248	66	46	67

Таблица 2. Статистические параметры умерших от COVID-19 на примере четырех регионов Российской Федерации (данные за период с 03.02.2020 г. по 17.06.2020 г.)

Table 2. Statistical parameters of the lethal cases from COVID-19 in four regions of the Russian Federation (data for the period of 03.02.2020–17.06.2020)

Регион / Region	Количество измерений (дни), n / Number of measurements (days), n	Максимальное количество, n / Maximal number, n	Среднее количество, n / Mean number, n	Средний ежедневный прирост, n / Mean daily increase, n	Стандартное отклонение / Standard deviation
Москва / Moscow	108	78	32,00	33,01	25,00
Санкт-Петербург / Saint Petersburg	108	57	12,00	7,05	16,00
Республика Карелия / Republic of Karelia	108	1	0,02	0,02	0,15
Ямало-Ненецкий автономный округ / Yamalo-Nenets Autonomous District	108	4	0,45	0,27	0,91

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

Таблица 3 (начало). Статистические характеристики полученного набора данных

Table 3 (beginning). Statistical characteristics of the obtained set of data

№ п/п / No.	Показатель / Parameter	Среднее значение / Mean value	Стандартное отклонение / Standard deviation	Минимальное значение / Minimal value	Максимальное значение / Maximal value	Этап / Stage
1	Количество зараженных за день *i* – 1 Number of cases per Day *i* – 1	4502,88	20 516,9	0	289 140	1
2	Количество новых зараженных за день *i* Number of new disease cases per Day *i*	121,61	785,63	0	25 615	1
3	Количество новых зараженных за день *i* – 1 Number of new disease cases per Day *i* – 1	119,02	776,6	0	25 615	4
4	Скорость прироста новых зараженных за день *i* – 1 The rate of increase in new disease cases per Day *i* – 1	1,69	891,43	0	45 315	4
5	Количество новых зараженных за день *i* – 2 Number of new disease cases per Day *i* – 2	116,39	768,29	0	25 615	4
6	Скорость прироста новых зараженных за день *i* – 2 The rate of increase in new disease cases per Day *i* – 2	1,44	890,02	0	45 315	4
7	Количество новых зараженных за день *i* – 3 Number of new disease cases per Day *i* – 3	114,03	763,02	0	25 615	4
8	Скорость прироста новых зараженных за день *i* – 3 The rate of increase in new disease cases per Day *i* – 3	1,45	889,36	0	45 315	4
9	Количество новых зараженных за день *i* – 4 Number of new disease cases per Day *i* – 4	111,65	756,13	0	25 615	4
10	Скорость прироста новых зараженных за день *i* – 4 The rate of increase in new disease cases per Day *i* – 4	1,61	889,02	0	45 315	4
11	Количество новых зараженных за день *i* – 5 Number of new disease cases per Day *i* – 5	109,11	747,33	0	25 615	4
12	Скорость прироста новых зараженных за день *i* – 5 The rate of increase in new disease cases per Day *i* – 5	1,47	888,62	0	45 315	4
13	Количество новых зараженных за день *i* – 6 Number of new disease cases per Day *i* – 6	106,71	740,64	0	25 615	4
14	Скорость прироста новых зараженных за день *i* – 6 The rate of increase in new disease cases per Day *i* – 6	1,44	888,48	0	45 315	4
15	Дней после 1 зараженного в регионе Days after 1 disease case in the region	44,18	25,21	0	107	4
16	Дней после 10 зараженных в регионе Days after 10 disease cases in the region	44,18	25,21	0	107	4
17	Дней после 100 зараженных в регионе Days after 100 disease cases in the region	42,97	25,46	0	103	4
18	Дней после 1000 зараженных в регионе Days after 1,000 disease cases in the region	42,31	25,63	0	103	4
19	Дней после 10 000 зараженных в регионе Days after 10,000 disease cases in the region	41,57	25,8	0	103	4
20	Дней после 100 000 зараженных в регионе Days after 100,000 disease cases in the region	41,13	25,94	0	103	4

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

Таблица 3 (продолжение). Статистические характеристики полученного набора данных

Table 3 (continuation). Statistical characteristics of the obtained set of data

№ п/п / No.	Показатель / Parameter	Среднее значение / Mean value	Стандартное отклонение / Standard deviation	Минимальное значение / Minimal value	Максимальное значение / Maximal value	Этап / Stage
21	Количество умерших за день *i* – 1 Number of lethal cases per Day *i* – 1	280,32	2117,28	0	40 883	4
22	Количество новых умерших за день *i* Number of new lethal cases per Day *i*	6,65	64,5	0	4419	4
23	Количество новых умерших за день *i* – 1 Number of new lethal cases per Day *i* – 1	6,59	64,15	0	4419	4
24	Скорость прироста новых умерших за день *i* – 1 The rate of increase in new lethal cases per Day *i* – 1	0,04	61,33	0	5162	4
25	Количество новых умерших за день *i* – 2 Number of new lethal cases per Day *i* – 2	6,5	63,97	0	4419	4
26	Скорость прироста новых умерших за день *i* – 2 The rate of increase in new lethal cases per Day *i* – 2	0,01	61,3	0	5162	4
27	Количество новых умерших за день *i* – 3 Number of new lethal cases per Day *i* – 3	6,43	63,89	0	4419	4
28	Скорость прироста новых умерших за день *i* – 3 The rate of increase in new lethal cases per Day *i* – 3	0,02	61,28	0	5162	4
29	Количество новых умерших за день *i* – 4 Number of new lethal cases per Day *i* – 4	6,36	63,75	0	4419	4
30	Скорость прироста новых умерших за день *i* – 4 The rate of increase in new lethal cases per Day *i* – 4	0,03	61,26	0	5162	4
31	Количество новых умерших за день *i* – 5 Number of new lethal cases per Day *i* – 5	6,27	63,5	0	4419	4
32	Скорость прироста новых умерших за день *i* – 5 The rate of increase in new lethal cases per Day *i* – 5	0,03	61,24	0	5162	4
33	Количество новых умерших за день *i* – 6 Number of new lethal cases per Day *i* – 6	6,18	63,3	0	4419	4
34	Скорость прироста новых умерших за день *i* – 6 The rate of increase in new lethal cases per Day *i* – 6	0,03	61,23	0	5162	4
35	Дней после 1 умершего в регионе Days after the 1 lethal case in the region	44,18	25,21	0	107	4
36	Дней после 10 умерших в регионе Days after the 10 lethal cases in the region	30,26	27,45	0	87	4
37	Дней после 100 умерших в регионе Days after the 100 lethal cases in the region	23,05	25,87	0	87	4
38	Дней после 1000 умерших в регионе Days after the 1,000 lethal cases in the region	18,83	24,14	0	87	4
39	Дней после 10 000 умерших в регионе Days after the 10,000 lethal cases in the region	16,28	22,94	0	87	4

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

Таблица 3 (окончание). Статистические характеристики полученного набора данных

Table 3 (end). Statistical characteristics of the obtained set of data

№ п/п / No.	Показатель / Parameter	Среднее значение / Mean value	Стандартное отклонение / Standard deviation	Минимальное значение / Minimal value	Максимальное значение / Maximal value	Этап / Stage
40	Дней после 100 000 умерших в регионе Days after the 100,000 lethal cases in the region	14,42	21,99	0	87	4
41	Население региона Population in the region	14 567 844,01	70 000 000	801	1 380 000 000	3
42	Городское население, % Urban population, %	63,12	16,51	15	100	3
43	Плотность населения на км^2 Density of population per km^2	233,79	1443,05	0	26337	3
44	Дней с первого зараженного Days after the first disease case	65,16	25,2	1	108	4

Примечание. *i – номер дня начиная с 03.02.2020 г. (i = 0) и заканчивая 17.06.2020 г. (i=108); i – 1 – один день до дня i; i – 2 – два дня до дня i и т.д.*

Note. *i – number of the day starting from 03.02.2020 (i = 0) and ending on 17.06.2020 (i=108); i – 1 – one day before day i; i – 2 – two days before day i etc.*

Для решения задачи был использован метод машинного обучения – градиентный бустинг CatBoost. Он показывает сопоставимые с другими методами результаты при работе с временными рядами [26, 27]. Также была протестирована модель линейной регрессии, но ее ошибка на тестовых данных оказалась значительно больше (**табл. 4**), поэтому в дальнейшем она не рассматривалась.

Набор данных (см. табл. 3) создавался при помощи оценки на модели в четыре этапа. Оценка проходила посредством сравнения ошибки предсказания количества новых зараженных: уменьшение ошибки на этой задаче также уменьшало ее на следующей. В качестве меры была выбрана среднеквадратичная ошибка модели (англ. root mean square error, RMSE).

Таблица 4. Среднеквадратичная ошибка модели на тестовых данных при различных наборах признаков для обучения

Table 4. Root mean square error of the model on test data in different sets of parameters for education

Этап / Stage	Модель / Model	Количество признаков / Number of parameters	Характеристика признаков / Characteristic of parameters	RMSE при предсказании нового количества зараженных / RMSE during prediction of new number of cases
1	CatBoost	1	Количество зараженных за предыдущий день по регионам РФ Number of disease cases during the previous day in the regions of the RF	90,07
2	CatBoost	1	Количество зараженных за предыдущий день по регионам РФ и по другим странам Number of disease cases during the previous day in the regions of the RF and other countries	69,09
3	CatBoost	4	Количество зараженных за предыдущий день по регионам РФ и другим странам и данные о регионе Number of disease cases during the previous day in the regions of the RF and other countries and the data on the region	67,54
4	CatBoost	24	Количество зараженных за предыдущий день по регионам РФ и другим странам, данные о регионе и данные об изменении количества зараженных Number of disease cases during the previous day in the regions of the RF and other countries, the data on the region, and the data on the changes in the number of cases	33,04
4	Линейная регрессия Linear regression	24	Количество зараженных за предыдущий день по регионам РФ и другим странам, данные о регионе и данные об изменении количества зараженных Number of disease cases during the previous day in the regions of the RF and other countries, the data on the region, and the data on the changes in the number of cases	304,38

Примечание. *RMSE (англ. root mean square error) – среднеквадратичная ошибка модели; CatBoost – градиентный бустинг; РФ – Российская Федерация.*

Note. *RMSE – root mean square error; CatBoost – gradient boosting; RF – Russian Federation.*

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

На первом этапе использовали один признак – количество зараженных за прошлый день, что привело к RMSE 90,07. На втором этапе применяли первый признак и данные по другим странам и получили RMSE 69,09. На третьем этапе использовали информацию о численности регионов РФ и других стран – это позволило достичь RMSE 67,54. На четвертом этапе добавили следующие характеристики: дни после первого зараженного, дни после 10^n (n = 0, 1, 2, 3, 4, 5) заражений в регионе, количество новых зараженных для дня $i - n$ и скорость прироста новых зараженных для дня $i - n$ (n = 1, 2, 3, 4, 5, 6, 7). Таким образом, RMSE уменьшилась до 33,04 (см. табл. 4).

Полученный набор данных использовали в итоговой модели для прогнозирования количества новых зараженных. Предсказанные метки экспоненцировались для избавления от логарифма. Результаты прогноза числа зараженных представлены на **рисунке 1**.

Для прогноза количества умерших была построена вторая мо-

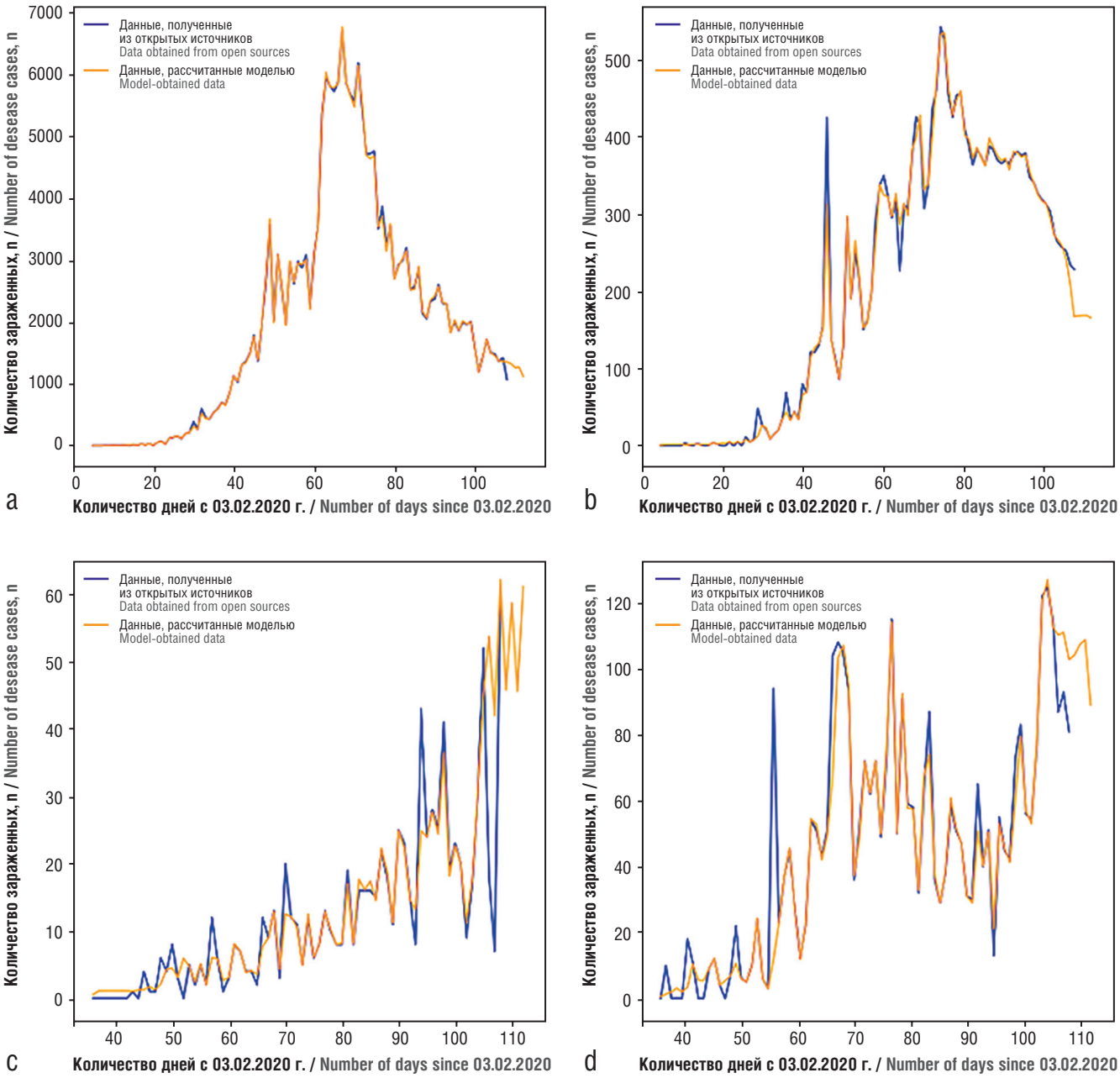


Рисунок 1. Прогноз числа зараженных SARS-CoV-2 на период 7 дней при использовании 24 признаков: число зараженных за прошлый день (логарифмированное), изменение ежедневного числа зараженных за 7 дней (логарифмированное), изменение скорости ежедневного числа зараженных за 7 дней (логарифмированное), период после появления в регионе первых (1, 10, 100, 1000, 10 000) зараженных, общая численность населения, плотность населения, процент городского населения:
a – для Москвы; b – для Санкт-Петербурга; c – для Республики Карелии; d – для Ямало-Ненецкого автономного округа.
100 дней демонстрируемого периода использовались в обучении, 7 дней – для составления прогноза

Figure 1. Prognosis on the number of disease cases with SARS-CoV-2 for 7 days based on 24 features: number of disease cases on the previous day (in logarithmic form), changes in the daily number of disease cases during 7 days (in logarithmic form), changes in the rate of daily increase in the disease cases during 7 days (in logarithmic form), a period after the appearance of the first disease cases in the region (1, 10, 100, 1,000, 10,000), total population, population density, percent of urban population:
a – for Moscow; b – for Saint Petersburg; c – for the Republic of Karelia; d – for Yamalo-Nenets Autonomous District.
100 days of the demonstrated period were used in training, 7 days – for prognosis

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

дель. Она включает все те же признаки из таблицы 3, но только по умершим. Итоги прогнозирования числа умерших показаны на **рисунке 2**.

Результаты прогнозирования полученной модели были оценены с помощью двух статистических метрик, применимых для анализа временных рядов: RMSE и средняя абсолютная ошибка предсказания модели в процентах (англ. mean absolute percentage error, MAPE).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ / RESULTS AND DISCUSSION

Метрики модели прогноза количества зараженных и умерших для различных по численности населения регионов РФ представлены в **таблице 5**. Модель показывает меньшую ошибку для регионов с большей плотностью населения, чем для менее населенных областей. Возможно, это происходит из-за выбросов, которые сложны в предсказаниях.

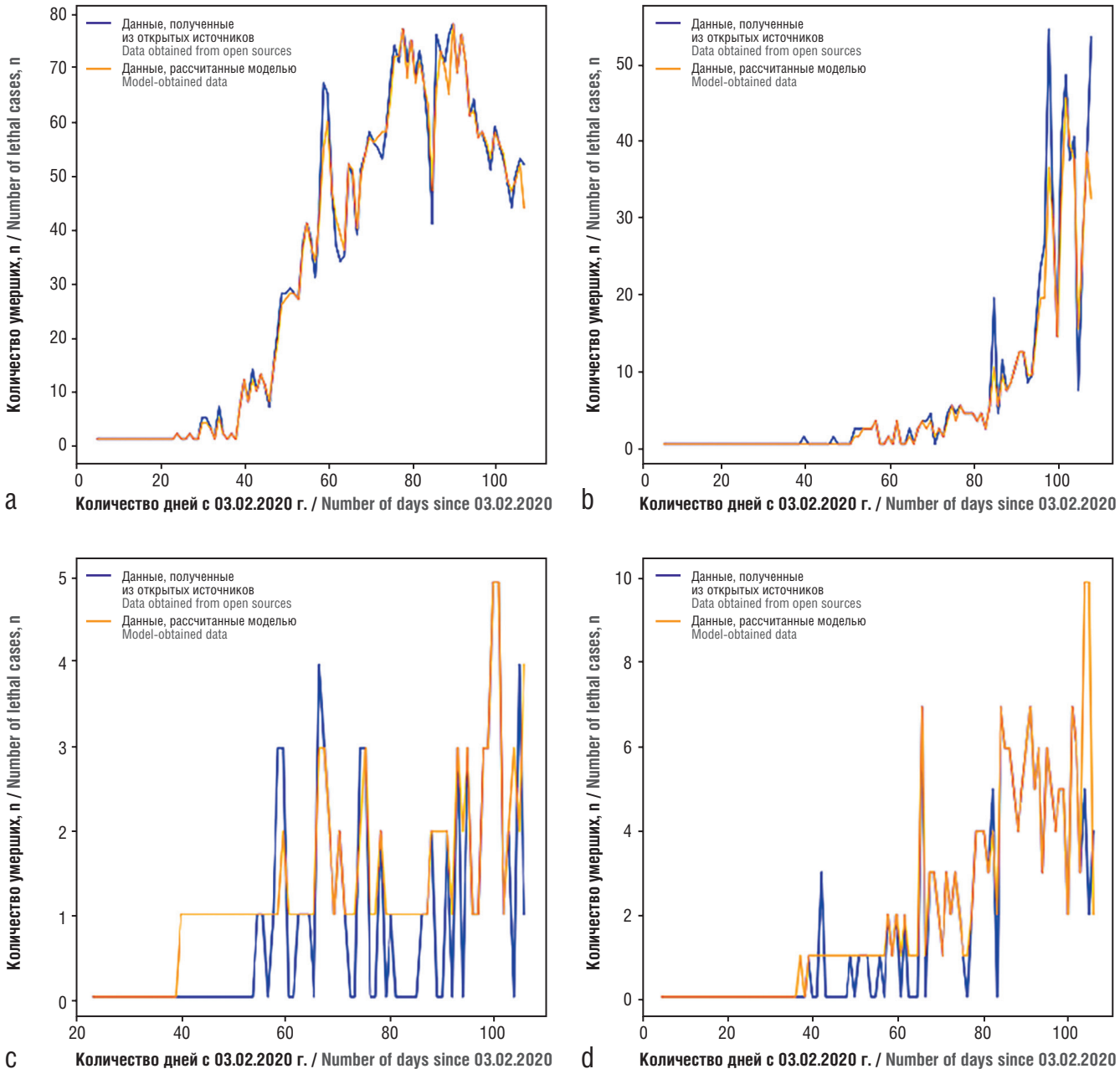


Рисунок 2. Прогноз числа умерших на период 7 дней при использовании 24 признаков: число умерших за прошлый день (логарифмированное), изменение ежедневного числа умерших за 7 дней (логарифмированное), изменение скорости ежедневного числа умерших за 7 дней (логарифмированное), период после появления в регионе первых (1, 10, 100, 1000, 10 000) умерших, общая численность населения, плотность населения, процент городского населения:
a – для Москвы; b – для Санкт-Петербурга; c – для Ростовской обл.; d – для Нижегородской обл.
100 дней демонстрируемого периода использовались в обучении, 7 дней – для составления прогноза

Figure 2. Prognosis on the number of lethal cases for 7 days based on a 24-feature model: number of lethal cases for the previous day (in logarithmic form), changes in the daily number of lethal cases for 7 days (in logarithmic form), changes in the increase in the daily rate of lethal cases for 7 days (in logarithmic form), a period after the appearance of the first lethal cases in the region (1, 10, 100, 1,000, 10,000), total population, population density, percent of urban population:
a – for Moscow; b – for Saint Petersburg, c – for Rostov Region; d – for Nizhny Novgorod Region.
100 days of the demonstrated period were used in training, 7 days – for prognosis

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

Таблица 5. Оценка результатов модели на основе метрик MAPE и RMSE на примере отдельных регионов Российской Федерации

Table 5. The evaluation of the modelled results based on MAPE and RMSE metrics in selected regions of the Russian Federation

Регион / Region	Численность населения на начало 2020 г., тыс. чел. / Population in the beginning of 2020, thousand people	Число зараженных на 14.07.2020 г., тыс. чел. / Number of disease cases as of 14.07.2020, thousand people	MAPE прогноза количества зараженных, % / MAPE prognosis on the number of disease cases, %	MAPE прогноза количества умерших, % / MAPE prognosis on the number of lethal cases, %	RMSE прогноза количества зараженных / RMSE prognosis on the number of disease cases	RMSE прогноза количества умерших / RMSE prognosis on the number of lethal cases
Москва / Moscow	12 692	209, 7	7,54	3,72	169	12
Санкт-Петербург / Saint Petersburg	5 393	21,3	4,78	7,67	817	27
Республика Карелия / Republic of Karelia	615	0,87	20,32	0	46	0,0019
Ямало-Ненецкий автономный округ / Yamalo-Nenets Autonomous District	544	3,6	11,28	2,62	207	0,04
Ярославская обл. / Yaroslavl Region	1 253	4,2	13,34	3,60	58	0,26
Рязанская обл. / Ryazan Region	1 109	4,6	12,56	4,08	80	0,025
Новосибирская обл. / Novosibirsk Region	2 798	1,5	15,29	7,95	41	0,093
Чеченская Республика / Chechen Republic	1 477	4,6	11,01	3,19	59	0,015

Примечание. *MAPE (англ. mean absolute percentage error) – ошибка предсказания модели в процентах; RMSE (англ. root mean square error) – среднеквадратичная ошибка.*

Note. *MAPE – mean absolute percentage error; RMSE – root mean square error.*

Таблица 6. Количество регионов Российской Федерации с различными метриками MAPE

Table 6. The number of regions of the Russian Federation with different MAPE metrics

Значение MAPE / MAPE value	Количество регионов / Number of regions
Менее 5% / Less than 5%	7
От 5% до менее 10% / From 5% to less than 10%	21
От 10% до менее 15% / From 10% to less than 15%	24
15% и более / 15% or more	33

Примечание. *MAPE (англ. mean absolute percentage error) – ошибка предсказания модели в процентах.*

Note. *MAPE – mean absolute percentage error.*

Метрики MAPE для оценки точности работы модели по 52 регионам РФ составляют от 5% до 15% (**табл. 6**) и от 15% до 24% для остальных регионов. Таким образом, модель показывает высокое сходство прогнозных значений с фактическими данными для большинства регионов РФ [28].

Также прогностические модели были оценены на данных по зараженным и умершим, полученных за июнь–август 2020 г. (во время отмены карантинных мер). В **таблице 7** представлены результаты на основе метрик MAPE, RMSE на примере тех же регионов и видно, что их точность не стала ниже по сравнению с данными в таблице 5.

После встраивания в платформу прогнозной аналитики Webiomed.Analytics модель ежедневно переобучается на обновленных данных. На **рисунке 3** приведены графики распространения коронавирусной инфекции в Москве и Санкт-Петребурге в течение всего периода с момента начала пандемии.

Метрики прогноза количества зараженных за весь период пандемии COVID-19 составили 21,73 и 345,37 для MAPE и RMSE соответственно.

Кроме того, были отдельно изучены периоды так называемых волн – быстрого роста числа заболевших и последующего за ним спада заболеваемости. В тренировочный набор были включены данные всех дней волны, за исключением последних трех, в тестовый – данные последних трех дней. Даты начала роста и спада волн определялись по графику для Москвы. Дата начала отсчета – 03.02.2020 г. Результаты представлены на **рисунках 4** и **5**.

ЗАКЛЮЧЕНИЕ / CONCLUSION

Таким образом, нами разработана краткосрочная модель предсказания зараженных и умерших от COVID-19 на период до 14 дней. Оценка точности модели по MAPE составляет от 2,3% до 24% для 85 регионов РФ. Показано преимущество метода машин-

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

Таблица 7. Оценка результатов модели на основе метрик MAPE и RMSE в отдельных регионах Российской Федерации во время отмены карантинных мероприятий

Table 7. The evaluation of the modeled results based on MAPE and RMSE metrics in selected regions of the Russian Federation during the cancellation of quarantine measures

Регион / Region	Число зараженных на 04.09.2020 г., тыс. чел. / Number of disease cases as of 04.09.2020, thousand people	MAPE прогноза количества зараженных, % / MAPE prognosis on the number of disease cases, %	MAPE прогноза количества умерших, % / MAPE prognosis on the number of lethal cases, %	RMSE прогноза количества зараженных / RMSE prognosis on the number of disease cases	RMSE прогноза количества умерших / RMSE prognosis on the number of lethal cases
Москва / Moscow	220,1	6,85	3,83	161	2,59
Санкт-Петербург / Saint Petersburg	23,7	5,81	8,40	27	5,66
Республика Карелия / Republic of Karelia	1,2	16,99	3,24	7	0,03
Ямало-Ненецкий автономный округ / Yamalo-Nenets Autonomous District	4,1	12,17	3,45	7	0,17
Ярославская обл. / Yaroslavl Region	4,7	14,11	4,22	8	0,50
Рязанская обл. / Ryazan Region	5,1	19,10	4,84	8	0,23
Новосибирская обл. / Novosibirsk Region	1,7	13,48	3,81	5	0,09
Чеченская Республика / Chechen Republic	5,8	11,22	10,27	7	0,40

Примечание. *MAPE (англ. mean absolute percentage error) – ошибка предсказания модели в процентах; RMSE (англ. root mean square error) – среднеквадратичная ошибка.*

Note. *MAPE – mean absolute percentage error; RMSE – root mean square error.*

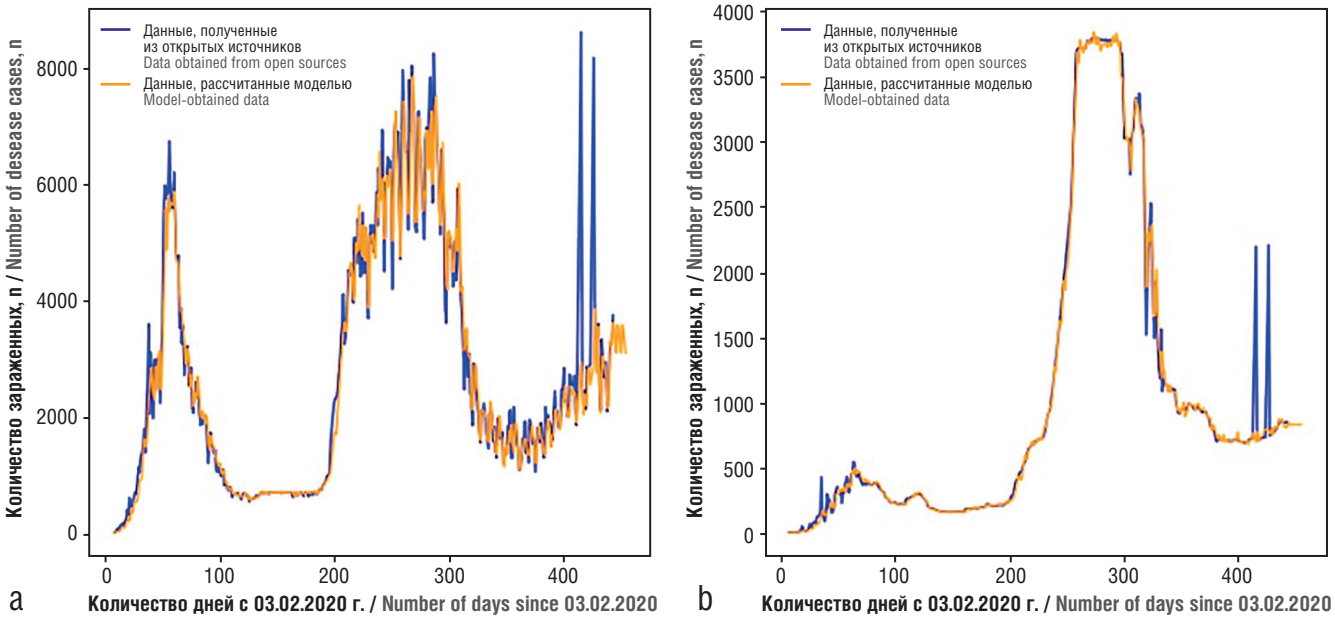


Рисунок 3. Распространение коронавирусной инфекции за весь исследуемый период (последние два пика – техническая ошибка, связанная с работой скрипта, собирающего данные):
a – в Москве; b – в Санкт-Петербурге

Figure 3. The spread of coronavirus infection during the studied period (the last two peaks – technical error associated with the operation of the script that collects the data):
a – in Moscow; b – in Saint Petersburg

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

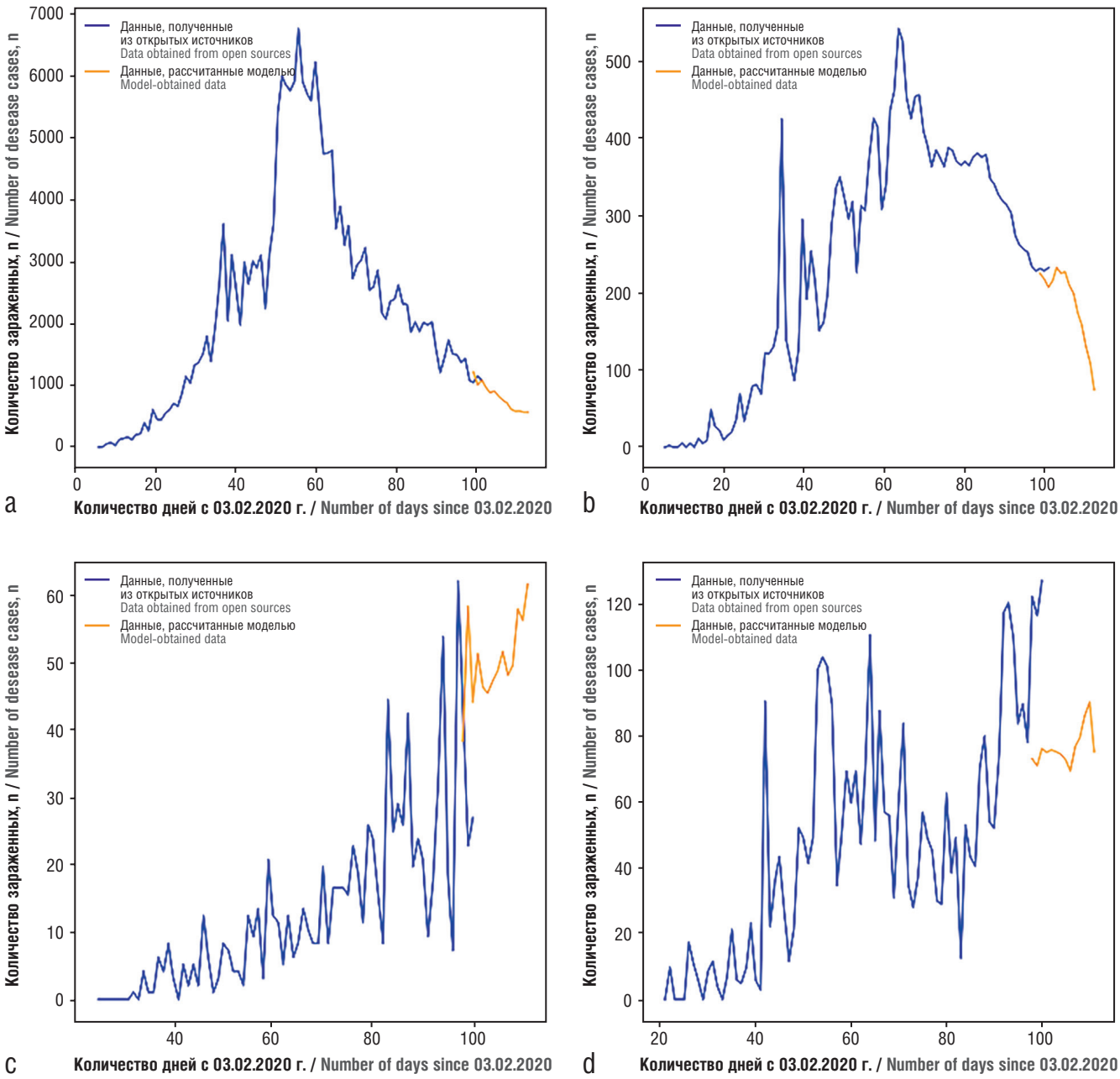


Рисунок 4. Полученная и предсказанная заболеваемость COVID-19 для «первой» волны пандемии (первые 100 сут с даты начала отсчета). Метрики ошибки для этого периода составили: MAPE 34,24 и RMSE 72,70:
a – для Москвы; b – для Санкт-Петербурга; c – для Республики Карелии; d – для Ямало-Ненецкого автономного округа.
MAPE (англ. mean absolute percentage error) – ошибка предсказания модели в процентах; RMSE (англ. root mean square error) – среднеквадратичная ошибка

Figure 4. The obtained and predicted morbidity rate with COVID-19 for the "first" wave of pandemic (first 10 days from the date of the starting point). Metrics of the error for this period were MAPE 34.24 and RMSE 72.70:
a – for Moscow; b – for Saint Petersburg; c- for the Republic of Karelia; d – Yamalo-Nenets Autonomous District.
MAPE – mean absolute percentage error; RMSE – root mean square error

ного обучения CatBoost перед линейной регрессией на примере величины ошибки RMSE.

Созданная модель реализована в виде веб-приложения в рамках платформы прогнозной аналитики Webiomed для системы здравоохранения. Данные на платформе ежедневно автоматически обновляются, пользователь может выбрать интересующий регион и построить прогноз по количеству зараженных или умерших.

Разработанная модель прогнозирования числа зараженных и умерших может быть использована не только для прогнозирования пандемии новой коронавирусной инфекции, но и для контроля и оценки распространения заболеваний из группы новых инфекций на этапах их возникновения, пика заболеваемости и периода стабилизации. Регулярный мониторинг при помощи данной модели позволит оперативно оценивать эпидемиологическую ситуацию и предпринимать обоснованные управленческие решения для предотвращения и купирования распространения инфекции.

Перспективы дальнейшего совершенствования модели видятся в интеграции существующих в РФ медицинских информационных систем на единой платформе для сбора обезличенных биомедицинских данных, что позволило бы учитывать особенности здоровья пациентов и тем самым повышать прогностическую точность модели.

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

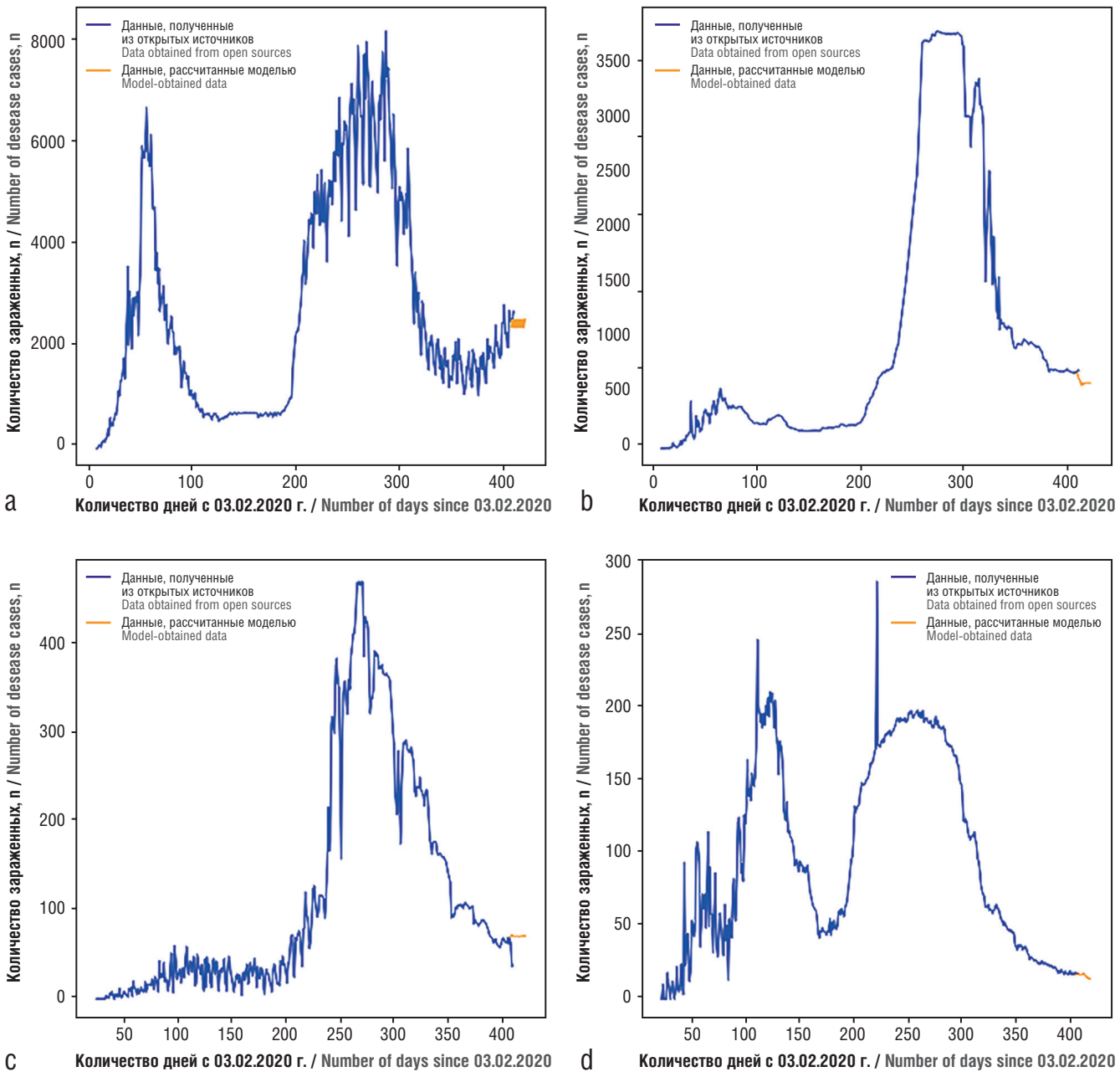


Рисунок 5. Полученная и предсказанная заболеваемость COVID-19 для периода 350 дней с начала пандемии, включающего две волны и плато между ними. Метрики ошибки для этого периода составили: MAPE 24,99 и RMSE 372,53:
a – для Москвы; b – для Санкт-Петербурга; c – для Республики Карелии; d – для Ямало-Ненецкого автономного округа.
MAPE (англ. mean absolute percentage error) – ошибка предсказания модели в процентах; RMSE (англ. root mean square error) – среднеквадратичная ошибка

Figure 5. The obtained and predicted morbidity rate with COVID-19 for 350 days from the beginning of pandemics that included two waves and plateau between them. Metrics for this period were MAPE 24.99 and RMSE 372.53:
a – for Moscow; b – for Saint Petersburg; c – for the Republic of Karelia; d – Yamalo-Nenets Autonomous District.
MAPE – mean absolute percentage error; RMSE – root mean square error

ЛИТЕРАТУРА:

1. WHO Coronavirus (COVID-19) Dashboard. URL: https://covid19.who.int/ (дата обращения 17.06.2020).
2. Временные методические рекомендации «Профилактика, диагностика и лечение новой коронавирусной инфекции (COVID 19). Версия 11» (утв. Министерством здравоохранения РФ 7 мая 2021 г.). URL: https://base.garant.ru/400738625/ (дата обращения 17.06.2021).
3. Huang C., Wang Y., Li X., et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020; 395 (10223): 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.
4. Onder G., Rezza G., Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA*. 2020; 323 (18): 1775–6. https://doi.org/10.1001/jama.2020.4683.
5. Mehta P., McAuley D.F., Brown M., et al. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet*. 2020; 395 (10229): 1033–4. https://doi.org/10.1016/S0140-6736(20)30628-0.

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях.
Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

6. Hollander J.E., Carr B.G. Virtually perfect? Telemedicine for COVID-19. *N Engl J Med.* 2020; 382 (18): 1679–81. https://doi.org/10.1056/NEJMp2003539.
7. Стенограмма совещания о санитарно-эпидемиологической обстановке в России (13.04.2020 г.). URL: http://prezident.org/tekst/stenogramma-soveschanija-o-sanitarno-epidemiologicheskoi-obstanovke-v-rossii-13-04-2020.html (дата обращения 17.06.2021).
8. Гусев А.В., Добриднюк С.Л. Искусственный интеллект в медицине и здравоохранении. *Информационное общество.* 2017; 4-5: 78–93.
9. Гусев А.В., Плисс М.А. Основные рекомендации к созданию и развитию информационных систем в здравоохранении на базе искусственного интеллекта. *Врач и информационные технологии.* 2018; 3: 45–60.
10. Гусев А.В., Кузнецова Т.Ю., Корсаков И.Н. Искусственный интеллект в оценке рисков развития сердечно-сосудистых заболеваний. *Журнал телемедицины и электронного здравоохранения.* 2018; 3 (8): 85–90.
11. Гусев А.В., Гаврилов Д.В., Корсаков И.Н. и др. Перспективы использования методов машинного обучения для предсказания сердечно-сосудистых заболеваний. *Врач и информационные технологии.* 2019; 3: 41–7.
12. Тамм М.В. Коронавирусная инфекция в Москве: прогнозы и сценарии. *ФАРМАКОЭКОНОМИКА. Современная фармакоэкономика и фармакоэпидемиология.* 2020; 13 (1): 43–51. https://doi.org/10.17749/2070-4909.2020.13.1.43-51.
13. Fanelli D., Piazzab F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals.* 2020; 134: 109761. https://doi.org/10.1016/j.chaos.2020.109761.
14. Petropoulos F., Makridakis S. Forecasting the novel coronavirus COVID-19. *PLoS One.* 2020; 15 (3): e0231236. https://doi.org/10.1371/journal.pone.0231236.
15. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci The Total Environ.* 2020; 729: 138817. https://doi.org/10.1016/j.scitotenv.2020.138817.
16. Roda W.C., Varughese M.B., Han D., Li M.Y. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect Dis Model.* 2020; 5: 271–81. https://doi.org/10.1016/j.idm.2020.03.001.
17. Pandey G., Chaudhary P., Gupta R., Pal S. SEIR and Regression Model based COVID-19 outbreak predictions in India. arXiv:2004.00958. https://doi.org/10.1101/2020.04.01.20049825.
18. Алешукина А.В., Денисенко В.В., Алешукин Г.С., Голошва Е.В. Применение математической модели для прогнозирования эпидемиологической ситуации по COVID-19 в Ростовской области. COVID19-Preprints.Microbe.ru; 2020. https://doi.org/10.21055/preprints-3111736.
19. Yang Z., Zeng Z., Wang K., et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis.* 2020; 12 (3): 165–74. https://doi.org/10.21037/jtd.2020.02.64.
20. Petropoulos F., Makridakis S. Forecasting the novel coronavirus COVID-19. *PLoS One.* 2020; 15 (3): e0231236. https://doi.org/10.1371/journal.pone.0231236.
21. Мелик-Гусейнов Д.В., Карякин Н.Н., Благонравова А.С. и др. Регрессионные модели прогнозирования количества летальных исходов при новой коронавирусной инфекции. *Современные технологии в медицине.* 2020; 12 (2): 6–13. https://doi.org/http://doi.org/10.17691/stm2020.12.2.01.
22. Forecast the global spread of COVID-19. URL: https://ods.ai/competitions/sberbank-covid19-forecast (дата обращения 17.06.2021).
23. Prokhorenkova L., Gusev G., Vorobev A., et al. CatBoost: unbiased boosting with categorical features. 32nd Conference on Neural Information Processing Systems. NeurIPS; 2018.
24. Стопкоронавирус.РФ. URL: https://стопкоронавирус.рф (дата обращения 17.06.2021).
25. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. URL: https://github.com/CSSEGISandData/COVID-19 (дата обращения 17.06.2021).
26. Pavlyshenko B.M. Linear, machine learning and probabilistic approaches for time series analysis. The 1st IEEE International Conference on Data Stream Mining & Processing, 23–27 August 2016, Lviv, Ukraine.
27. Ji C., Zoua X., Hu Y., et al. XG-SF: an XGBoost classifier based on shapelet features for time series classification. *Procedia Comput Sci.* 2019; 147: 24–8. https://doi.org/10.1016/j.procs.2019.01.179.
28. Botchkarev A. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: properties and typology. arXiv:1809.03006. https://doi.org/10.28945/4184.

REFERENCES:

1. WHO Coronavirus (COVID-19) Dashboard. Available at: https://covid19.who.int/ (accessed 17.06.2021).
2. Temporary methodological recommendations “Prevention, diagnosis and treatment of a new coronavirus infection (COVID 19). Version 11” (approved by the Ministry of Health of the Russian Federation on May 7, 2021) (in Russ.) Available at: https://base.garant.ru/400738625/ (accessed 17.06.2021).
3. Huang C., Wang Y., Li X., et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020; 395 (10223): 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.
4. Onder G., Rezza G., Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA.* 2020; 323 (18): 1775–6. https://doi.org/10.1001/jama.2020.4683.
5. Mehta P., McAuley D.F., Brown M., et al. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet.* 2020; 395 (10229): 1033–4. https://doi.org/10.1016/S0140-6736(20)30628-0.
6. Hollander J.E., Carr B.G. Virtually perfect? Telemedicine for COVID-19. *N Engl J Med.* 2020; 382 (18): 1679–81. https://doi.org/10.1056/NEJMp2003539.
7. Stenogram of the meeting on the sanitary and epidemiological situation in Russia (April 13, 2020). Available at: http://prezident.org/tekst/stenogramma-soveschanija-o-sanitarno-epidemiologicheskoi-obstanovke-v-rossii-13-04-2020.html (in Russ.) (accessed 17.06.2021).
8. Gusev A.V., Dobridnyuk S.L. Artificial intelligence in medicine and healthcare. *Information Society.* 2017; 4-5: 78–93 (in Russ.).
9. Gusev A.V., Pliss M.A. The basic recommendations for the creation and development of information systemsin health care based on artificial intelligence. *Information Technologies for the Physician.* 2018; 3: 45–60 (in Russ.).
10. Gusev A.V., Kuznetsova T.Yu., Korsakov I.N. Artificial intelligence for cardiovascular risks assessment. *The Journal of Telemedicine and e-Health.* 2018; 3 (8): 85–90 (in Russ.).
11. Gusev A.V., Gavrilov D.V., Korsakov I.N., et al. Prospects for the use of machine learning methods for predicting cardiovascular disease. *Information Technologies for the Physician.* 2019; 3: 41–7 (in Russ.).
12. Tamm M.V. COVID-19 in Moscow: prognoses and scenarios. *FARMAKOEKONOMIKA. Sovremennaya farmakoekonomika i farmakoepidemiologiya / FARMAKOEKONOMIKA. Modern Pharmacoeconomics and Pharmacoepidemiology.* 2020; 13 (1): 43–51 (in Russ.). https://doi.org/10.17749/2070-4909.2020.13.1.43-51.
13. Fanelli D., Piazzab F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals.* 2020; 134: 109761. https://doi.org/10.1016/j.chaos.2020.109761.

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

14. Petropoulos F., Makridakis S. Forecasting the novel coronavirus COVID-19. *PLoS One.* 2020; 15 (3): e0231236. https://doi.org/10.1371/journal.pone.0231236.
15. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci The Total Environ.* 2020; 729: 138817. https://doi.org/10.1016/j.scitotenv.2020.138817.
16. Roda W.C., Varughese M.B., Han D., Li M.Y. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect Dis Model.* 2020; 5: 271–81. https://doi.org/10.1016/j.idm.2020.03.001.
17. Pandey G., Chaudhary P., Gupta R., Pal S. SEIR and Regression Model based COVID-19 outbreak predictions in India. arXiv:2004.00958. https://doi.org/10.1101/2020.04.01.20049825.
18. Aleshukina A.V., Denisenko V.V., Aleshukin G.S., Goloshva E.V. Application of a mathematical model for predicting the epidemiological situation of COVID-19 in the Rostov Region. COVID19-Preprints.Microbe.ru; 2020 (in Russ.).
19. Yang Z., Zeng Z., Wang K., et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis.* 2020; 12 (3): 165–74. https://doi.org/10.21037/jtd.2020.02.64.
20. Petropoulos F., Makridakis S. Forecasting the novel coronavirus COVID-19. *PLoS One.* 2020; 15 (3): e0231236. https://doi.org/10.1371/journal.pone.0231236.
21. Melik-Huseynov D.V., Karyakin N.N., Blagonravova A.S., et al. Regression models predicting the number of deaths from the new coronavirus infection. *Modern Technologies in Medicine.* 2020; 12 (2): 6–13 (in Russ.). https://doi.org/http://doi.org/10.17691/stm2020.12.2.01.
22. Forecast the global spread of COVID-19. Available at: https://ods.ai/competitions/sberbank-covid19-forecast (accessed 17.06.2021).
23. Prokhorenkova L., Gusev G., Vorobev A., et al. CatBoost: unbiased boosting with categorical features. 32nd Conference on Neural Information Processing Systems. NeurIPS; 2018.
24. StopcoronavirusRF. Available at: https://стопкоронавирус.рф (in Russ.) (accessed 17.06.2021).
25. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Available at: https://github.com/CSSEGISandData/COVID-19 (accessed 17.06.2021).
26. Pavlyshenko B.M. Linear, machine learning and probabilistic approaches for time series analysis. The 1st IEEE International Conference on Data Stream Mining & Processing, 23–27 August 2016, Lviv, Ukraine.
27. Ji C., Zoua X., Hu Y., et al. XG-SF: an XGBoost classifier based on shapelet features for time series classification. *Procedia Comput Sci.* 2019; 147: 24–8. https://doi.org/10.1016/j.procs.2019.01.179.
28. Botchkarev A. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: properties and typology. arXiv:1809.03006. https://doi.org/10.28945/4184.

Сведения об авторах

Гаврилов Денис Владимирович – руководитель медицинского направления ООО «К-Скай» (Петрозаводск, Россия). ORCID ID: https://orcid.org/0000-0002-8745-857X; РИНЦ SPIN-код: 2860-6040.

Абрамов Роман Владимирович – аналитик данных ООО «К-Скай» (Петрозаводск, Россия).

Кирилкина Анна Валерьевна – заместитель главного врача по медицинской части ГБУЗ «Республиканская инфекционная больница» (Петрозаводск, Россия). ORCID ID: https://orcid.org/0000-0002-0400-8750.

Ившин Александр Анатольевич – к.м.н., заведующий кафедрой акушерства и гинекологии, дерматовенерологии ФГБОУ ВО «Петрозаводский государственный университет» (Петрозаводск, Россия). ORCID ID: https://orcid.org/0000-0001-7834-096X; Scopus Author ID: 57222275843; РИНЦ SPIN-код: 8196-6605. E-mail: scipeople@mail.ru.

Новицкий Роман Эдвардович – генеральный директор ООО «К-Скай» (Петрозаводск, Россия). ORCID ID: https://orcid.org/0000-0002-2350-977X; РИНЦ SPIN-код: 8309-1740.

About the authors

Denis V. Gavrilov – Head of Medical Department, K-SkAI LLC (Petrozavodsk, Russia). ORCID ID: https://orcid.org/0000-0002-8745-857X; RSCI SPIN-code: 2860-6040.

Roman V. Abramov – Data Analyst, K-SkAI LLC (Petrozavodsk, Russia).

Anna V. Kirilkina – Deputy Chief Physician for Medicine, Republican Infectious Diseases Hospital (Petrozavodsk, Russia). ORCID ID: https://orcid.org/0000-0002-0400-8750.

Aleksandr A. Ivshin – MD, PhD, Chief of Chair of Obstetrics and Gynecology, Dermatovenerology, Petrozavodsk State University (Petrozavodsk, Russia). ORCID ID: https://orcid.org/0000-0001-7834-096X; Scopus Author ID: 57222275843; RSCI SPIN-code: 8196-6605. E-mail: scipeople@mail.ru.

Roman E. Novitskiy – Director General, K-SkAI LLC (Petrozavodsk, Russia). ORCID ID: https://orcid.org/0000-0002-2350-977X; RSCI SPIN-code: 8309-1740.

Данная интернет-версия статьи была скачана с сайта https://www.pharmacoeconomics.ru. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.