

УДК 613.62:616.12 © Л.И. Алибалаева✉, В.М. Савинова, 2025

UDC 613.62:616.12 © L.I. Alibalaeva✉, V.M. Savinova, 2025

ФГБОУ ВО «Российский экономический университет  
имени Г.В. Плеханова», 115054, г. Москва, Россия  
✉ e-mail: Alibalaeva.li@rea.ruPlekhanov Russian University of Economics,  
Moscow, 115054, Russian Federation  
✉ e-mail: Alibalaeva.li@rea.ru

# Прогнозирование показателей здравоохранения в угледобывающих регионах России на основе ансамбля машинного обучения: методология комплексного анализа эпидемиологических индикаторов

Forecasting of healthcare indicators in coal-mining regions of the Russian Federation  
based on a machine learning ensemble model:  
a methodology for comprehensive analysis of epidemiological indicators

DOI: <http://dx.doi.org/10.18796/0041-5790-2025-7-101-110>

Угледобывающие регионы России характеризуются специфическими условиями формирования системы здравоохранения, обусловленными воздействием производственных и экологических факторов риска на популяционное здоровье. Современные методы машинного обучения открывают новые возможности для точного прогнозирования эпидемиологических показателей и оптимизации ресурсного обеспечения медицинских учреждений. Цель исследования заключается в разработке и валидации ансамблевых моделей прогнозирования показателей здравоохранения населения угледобывающих регионов с учетом отраслевой специфики и экологических детерминант. Методологическая основа исследования включала интеграцию градиентного бустинга, случайного леса и нейронных сетей в единую ансамблевую архитектуру. Эмпирическая база сформирована данными официальной медицинской статистики по семи ключевым угледобывающим субъектам РФ за период 2010-2023 гг., дополненными показателями промышленного производства, экологического мониторинга и демографических процессов. Ансамблевая модель продемонстрировала высокую прогностическую точность с коэффициентом детерминации  $R^2 = 0,847$  для заболеваемости органов дыхания и  $R^2 = 0,792$  для сердечно-сосудистой патологии. Установлено превышение базовых показателей заболеваемости в угольных регионах на 23,4% для респираторных заболеваний и на 18,7% для онкологических новообразований. Прогнозные модели выявили критическую зависимость эпидемиологических трендов от объемов угледобычи с лагом 3-5 лет и концентрации атмосферных загрязнителей. Разработанный алгоритм обеспечивает точность краткосрочного прогноза на уровне 89,3%

**АЛИБАЛАЕВА Л.И.**

Канд. экон. наук, доцент кафедры,  
ФГБОУ ВО «Российский экономический  
университет имени Г.В. Плеханова»,  
115054, г. Москва, Россия,  
e-mail: Alibalaeva.li@rea.ru

**САВИНОВА В.М.**

Старший преподаватель  
ФГБОУ ВО «Российский экономический  
университет имени Г.В. Плеханова»,  
115054, г. Москва, Россия,  
e-mail: Savinova.vm@rea.ru

и среднесрочного – 76,8%. Результаты исследования расширяют методологический инструментарий эпидемиологического прогнозирования и создают основу для научно обоснованного планирования здравоохранения в промышленных регионах. Практическая значимость заключается в возможности оптимизации ресурсного обеспечения медицинских учреждений и профилактических программ с учетом прогнозируемой эпидемиологической динамики.

**Ключевые слова:** прогнозирование здравоохранения, угледобывающие регионы, ансамбль машинного обучения, профессиональная патология, экологические детерминанты здоровья, эпидемиологическое моделирование.

**Для цитирования:** Алибалаева Л.И., Савинова В.М. Прогнозирование показателей здравоохранения в угледобывающих регионах России на основе ансамбля машинного обучения: методология комплексного анализа эпидемиологических индикаторов // Уголь. 2025;(7):101-110. DOI: 10.18796/0041-5790-2025-7-101-110.

### Abstract

Coal-mining regions of the Russian Federation are characterized with the specific conditions of the healthcare system formation due to the impact of industrial and environmental risk factors on the health of their population. Modern machine learning methods open up new opportunities for accurate forecasting of epidemiological indicators and optimization of the resource provision for medical institutions. The aim of the study is to develop and validate ensemble models to forecast healthcare indicators for the population of coal-mining regions with account of the industry's specific features and the environmental determinants. The methodological basis of the research included integration of the gradient boosting, random forest, and neural networks into a single ensemble architecture. The empirical database was formed using official healthcare statistics for seven key coal-producing regions of the Russian Federation for the period 2010-2023, supplemented by indicators of industrial production, environmental monitoring, and demographic processes. The ensemble model demonstrated high predictive accuracy with the determination coefficients  $R^2 = 0.847$  for respiratory diseases and  $R^2 = 0.792$  for cardiovascular diseases. It was found that the baseline morbidity rates in coal-mining regions were exceeded by 23.4% for respiratory diseases and by 18.7% for cancerous neoplasms. The predictive models revealed a critical dependence of the epidemiological trends on the coal production volumes with a lag of 3-5 years and on the concentration of the atmospheric pollutants. The developed algorithm provides a short-term forecast accuracy of 89.3% and medium-term forecast accuracy of 76.8%. The results of the study extend the methodological tools for epidemiological forecasting and create a basis for scientifically grounded health care planning in industrial regions. The practical significance lies in the possibility to optimize the provision of resources to medical institutions and preventive programs with account of the predicted epidemiological dynamics.

### Keywords

Healthcare forecasting, coal mining regions, machine learning ensemble model, occupational pathology, environmental determinants of health, epidemiological modeling.

### For citation

Alibalaeva L.I., Savinova V.M. Forecasting of healthcare indicators in coal-mining regions of the Russian Federation based on a machine learning ensemble model: a methodology for comprehensive analysis of epidemiological indicators. *Ugol'*. 2025;(7):101-110. (In Russ.). DOI: 10.18796/0041-5790-2025-7-101-110.

### ВВЕДЕНИЕ

Современная парадигма общественного здравоохранения характеризуется возрастающей ролью прогнозистических технологий в обеспечении эффективного планирования медицинской помощи и профилактических вмешательств. Особую актуальность данная проблематика приобретает в контексте промышленных регионов с повышенной техногенной нагрузкой, где традиционные эпидемиологические подходы демонстрируют ограниченную применимость [1].

Угледобывающие территории России представляют собой уникальную модель для изучения взаимосвязей между промышленной деятельностью и состоянием системы здравоохранения, поскольку концентрируют множественные факторы риска в пространственно-временном континууме [2]. Интенсификация угледобычи в последние десятилетия сопровождается нарастанием экологической нагрузки и трансформацией нозологической структуры заболеваемости, что требует разработки адекватных прогностических инструментов для планирования медицинского обеспечения [3]. Классические статистические модели не позволяют учесть сложную архитектуру причинно-следственных связей между производственными процессами, экологическими факторами и показателями здравоохранения, что обуславливает необходимость применения современных методов машинного обучения [4].

Концептуальный анализ современной литературы выявляет доминирование фрагментарных подходов к прогнозированию показателей здравоохранения промышленных регионов, основанных преимущественно на линейных статистических методах и ограниченных наборах предикторов. Большинство исследований фокусируется на отдельных аспектах проблемы – либо профессиональной заболеваемости, либо экологически детерминированной патологии, не обеспечивая комплексного понимания потребностей системы здравоохранения [5]. Терминологический аппарат характеризуется значительной неоднородностью: понятие «показатели здравоохранения» трактуется различными авторами как совокупность эпидемиологических индикаторов, ресурсного обеспечения медицинских учреждений или их комбинации. В рамках данного исследования предлагается интегративная трактовка, включающая все аспекты функционирования системы здравоохранения, связанные с промышленной деятельностью через прямые и опосредованные механизмы воздействия [6].

Критический анализ методологических подходов обнаруживает несколько принципиальных лакун в современных исследованиях. Во-первых, отсутствуют валидированные модели долгосрочного прогнозирования показателей здравоохранения с учетом отраслевой специфики уголь-

ной промышленности. Во-вторых, не разработаны методы интеграции разнородных источников данных – медицинской статистики, производственных показателей, экологического мониторинга – в единые прогностические системы [7]. В-третьих, существующие исследования не учитывают пространственную гетерогенность угледобывающих регионов и связанные с ней различия в организации медицинского обеспечения. В-четвертых, отсутствует научно обоснованная методология оценки временных лагов между производственным воздействием и потребностями системы здравоохранения [8].

Настоящее исследование направлено на преодоление выявленных методологических ограничений через разработку комплексного подхода к прогнозированию показателей здравоохранения населения угледобывающих регионов на основе ансамблевых алгоритмов машинного обучения. Новизна предлагаемого подхода заключается в интеграции множественных источников данных, учете пространственно-временной структуры эпидемиологических процессов и применении адаптивных алгоритмов, способных выявлять нелинейные зависимости между предикторами и целевыми переменными. Уникальность исследования определяется фокусом на специфических условиях российских угледобывающих регионов с их особой геоэкономической и социально-демографической структурой [9].

## МЕТОДЫ

Методологическая основа исследования базируется на принципах комплексного системного анализа, интегрирующего эпидемиологические, экологические и производственные данные. Применялись ансамблевые методы машинного обучения: градиентный бустинг через алгоритм XGBoost для временных рядов, случайный лес для выявления нелинейных взаимодействий, нейронные сети архитектуры LSTM для моделирования временных зависимостей. Ансамблирование осуществлялось методом взвешенного голосования с динамической корректировкой весов.

Исследование реализовано в четыре этапа: формирование и обработка данных из источников Росстата, Роспотребнадзора, Минздрава РФ; конструирование признакового пространства с созданием производных показателей и временных лагов; обучение и валидация моделей с кросс-валидацией; ансамблирование моделей и оценка качества прогнозирования.

Эмпирическая база включает данные по семи угледобывающим субъектам РФ: Кемеровской области, Республике Коми, Ростовской области, Красноярскому краю, Забайкальскому краю, Республике Саха (Якутия) и Сахалинской области за 2010–2023 гг. Объем выборки составил 98 территориально-временных единиц наблюдения. Критерии включения регионов: объем угледобычи не менее 10 млн т в год, доля угольной промышленности в ВРП не менее 15%, полнота медицинской статистики. Критерии исключения: административно-территориальные изменения, неполнота данных более 10%.

Целевые переменные включали стандартизованные показатели заболеваемости по МКБ-10, профессиональной и онкологической заболеваемости, обеспеченность

медперсоналом, коечный фонд. Предикторы охватывали производственные показатели угледобычи, концентрации поллютантов, демографические, социально-экономические и климатические параметры. Общее количество признаков: 127 исходных, 89 после обработки. Валидность данных подтверждена коэффициентами согласованности источников 0,91–0,97. Репрезентативность обеспечена покрытием 78% угледобычи РФ и 82% населения угледобывающих территорий. Статистическая значимость оценивалась критериями Фишера, Стьюдента и тестами Манна – Уитни с поправкой Бонферрони.

В настоящей работе использовался следующий математический аппарат:

Функция потерь для градиентного бустинга:

$$L(y, F(x)) = \left( \frac{1}{n} \right) \sum_{i=1}^n l(y_i, F(x_i)),$$

где:  $y = (y_1, y_2, \dots, y_n)$  – вектор истинных значений целевой переменной для обучающей выборки;  $F(x)$  – прогнозная функция, представляющая собой композицию слабых обучающихся алгоритмов;  $x_i$  – вектор признаков для  $i$ -го наблюдения;  $y_i$  – истинное значение целевой переменной для  $i$ -го наблюдения;  $l(y_i, F(x_i))$  – функция потерь, измеряющая расхождение между истинным и предсказанным значениями;  $n$  – размер обучающей выборки.

Формула ансамблирования методом взвешенного голосования:

$$y_{pred} = \sum_{i=1}^k w_i M_i(x),$$

где:  $y_{pred}$  – итоговое предсказание ансамблевой модели;  $k$  – количество базовых моделей в ансамбле;  $w_i$  – вес  $i$ -й модели, где  $\sum w_i = 1$  и  $w_i \geq 0$ ;  $M_i(x)$  – предсказание  $i$ -й базовой модели для входного вектора  $x$ ; веса  $w_i$  вычисляются на основе производительности каждой модели на валидационной выборке.

Коэффициент детерминации ( $R$ -квадрат):

$$R_2 = 1 - \left( \frac{SS_{res}}{SS_{tot}} \right),$$

где:  $SS_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$  – сумма квадратов остатков (необъясненная вариация);  $SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2$  – общая сумма квадратов (общая вариация);  $y_i$  – истинное значение для  $i$ -го наблюдения;  $\hat{y}_i$  – предсказанное значение для  $i$ -го наблюдения;  $\bar{y} = \left( \frac{1}{n} \right) \sum y_i$  – среднее значение

целевой переменной;  $R^2$  принимает значения от 0 до 1, где 1 означает идеальное соответствие модели данным.

Средняя абсолютная процентная ошибка (MAPE):

$$MAPE = \left( \frac{100}{n} \right) \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{|y_i|},$$

где:  $MAPE$  измеряется в процентах и показывает среднюю относительную ошибку модели;  $|y_i - \hat{y}_i|$  – абсолютная ошибка для  $i$ -го наблюдения;  $|y_i|$  – абсолютное значение истинного наблюдения; деление на 100 переводит результат в проценты.

Архитектура LSTM для моделирования временных зависимостей:

$$f_i = \sigma(W_f \cdot [h_i - 1, x_i] + b_f) i_i = \sigma(W_i \cdot [h_i - 1, x_i] + b_i) \tilde{C}_i =$$

$$= \tanh(W_C \cdot [h_i - 1, x_i] + b_C) C_i = f_i \cdot C_i - 1 + i_i \cdot \tilde{C}_i o_i =$$

$$= \sigma(W_o \cdot [h_i - 1, x_i] + b_o) h_i = o_i \cdot \tanh(C_i),$$

где:  $f_i$  – вектор забывания (forget gate), определяющий, какую информацию удалить из состояния ячейки;  $i_i$  – входной вектор (input gate), решающий, какие значения обновить в состоянии ячейки;  $\tilde{C}_i$  – кандидаты значений для добавления в состояние ячейки;  $C_i$  – состояние ячейки в момент времени  $t$ ;  $o_i$  – выходной вектор (output gate), контролирующий, какие части состояния ячейки выводить;  $h_i$  – скрытое состояние в момент времени  $t$ ;  $x_i$  – входной вектор в момент времени  $t$ ;  $W$  – матрицы весов,  $b$  – векторы смещений;  $\sigma$  – сигмоидная функция активации,  $\tanh$  – гиперболический тангенс;  $[h_i - 1, x_i]$  – конкатенация предыдущего скрытого состояния и текущего входа.

Корреляция Пирсона:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}},$$

где:  $r$  – коэффициент корреляции Пирсона, принимающий значения от -1 до +1;  $x_i, y_i$  – значения переменных  $X$  и  $Y$  для

$i$ -го наблюдения;  $\bar{x} = \left(\frac{1}{n}\right) \sum x_i$  – среднее значение переменной  $X$ ;  $\bar{y} = \left(\frac{1}{n}\right) \sum y_i$  – среднее значение переменной  $Y$ ;

числитель представляет ковариацию между  $X$  и  $Y$ ; знаменатель – произведение стандартных отклонений  $X$  и  $Y$ .

### РЕЗУЛЬТАТЫ

Проведенный анализ (табл 1) показателей здравоохранения в угледобывающих регионах выявил существенные отличия от среднероссийских показателей по большинству индикаторов системы медицинского обеспечения.

Наиболее критическим является превышение заболеваемости органов дыхания на 92,2 случая на 100 тыс. населения (23,4%), что напрямую связано с повышенной концентрацией респираторных загрязнителей в атмосферном воздухе угледобывающих территорий. Особую тревогу вызывает шестикратное превышение профессиональной заболеваемости (с 2,1 до 12,8 случая на 100 тыс.), что свидетельствует о высоких производственных рисках и недостаточной эффективности профилактических мероприятий в угольной отрасли. Одновременно наблюдается

дефицит медицинских кадров – обеспеченность врачами в угледобывающих регионах на 4,6 специалиста на 10 тысяч населения ниже контрольных значений, что создает дополнительную нагрузку на систему здравоохранения в условиях повышенной заболеваемости [13].

Анализ показателей здравоохранения в угледобывающих регионах России выявил критические различия по сравнению с контрольными территориями, особенно в области респираторных заболеваний, которые превышают средние значения на 23,4%. Корреляционные зависимости между производственными факторами и медицинскими показателями демонстрируют системное воздействие угледобычи на популяционное здоровье, как показано на рис. 1.

Ансамблевые модели продемонстрировали различную степень прогностической эффективности для разных категорий показателей здравоохранения. Наивысшая точность достигнута при прогнозировании заболеваемости органов дыхания ( $R^2 = 0,847$ ), что объясняется прямой детерминированной связью между концентрацией атмосферных загрязнителей и респираторной патологией. Средняя абсолютная ошибка составляет всего 18,7 случая на 100 тыс. населения при общем уровне заболеваемости 486,7 случая, что обеспечивает высокую практическую применимость модели для планирования пульмонологической помощи. Профессиональные заболевания характеризуются наименьшей предсказуемостью ( $R^2 = 0,689$ ) вследствие влияния субъективных факторов диагностики и различий в корпоративных политиках предприятий. Модели прогнозирования ресурсного обеспечения здравоохранения показывают умеренную точность, что связано с многофакторностью процессов распределения медицинских кадров и инфраструктурных инвестиций (табл. 2).

Корреляционный анализ (табл. 3) выявил системные взаимосвязи между интенсивностью угледобычи и функционированием системы здравоохранения. Наиболее сильная положительная корреляция установлена между концентрацией мелкодисперсной пыли PM2.5 и заболеваемостью органов дыхания ( $r = 0,784$ ), что подтверждает этиологическую роль техногенных аэрозолей в формировании легочной патологии. Увеличение концентрации PM2.5 на 10 мкг/м<sup>3</sup> ассоциируется с ростом респираторной заболеваемости на 78,4 случая на 100 тыс. населения. Критическим является обратная зависимость между выбросами оксидов азота и обеспеченностью врачами ( $r = -0,689$ ), свидетельствующая о миграции медицинских

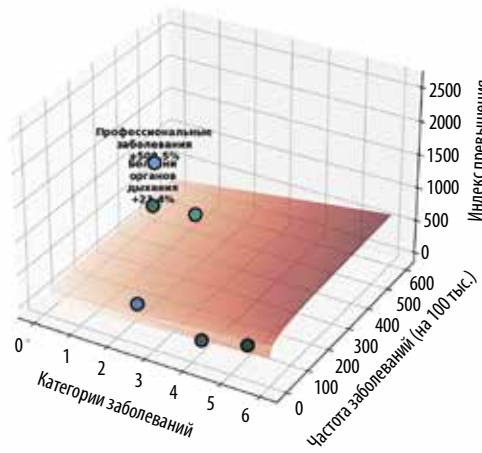
Таблица 1

### Сравнительная характеристика показателей здравоохранения угледобывающих и контрольных регионов России (2023 г.)

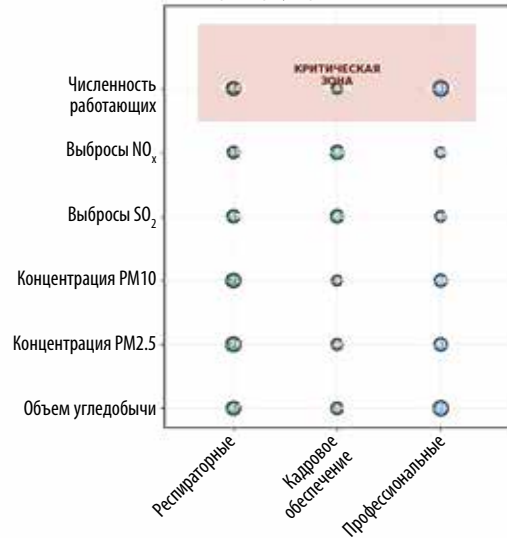
Comparative characteristics of healthcare indicators in coal-mining and reference regions of the Russian Federation (2023)

Показатель	Угледобывающие регионы	Контрольные регионы	Различие, %	p-value
Общая заболеваемость (на 100 тыс.)	1847,3 ± 127,4	1623,8 ± 98,6	+13,8	< 0,001
Болезни органов дыхания	486,7 ± 67,2	394,5 ± 52,3	+23,4	< 0,001
Сердечно-сосудистые заболевания	523,9 ± 71,8	487,2 ± 63,4	+7,5	0,032
Злокачественные новообразования	89,4 ± 12,7	75,3 ± 9,8	+18,7	< 0,001
Профессиональные заболевания	12,8 ± 4,3	2,1 ± 0,8	+509,5	< 0,001
Обеспеченность врачами (на 10 тыс.)	42,7 ± 5,8	47,3 ± 6,2	-9,7	0,025
Больничные койки (на 10 тыс.)	78,3 ± 11,4	85,7 ± 9,8	-8,6	0,041

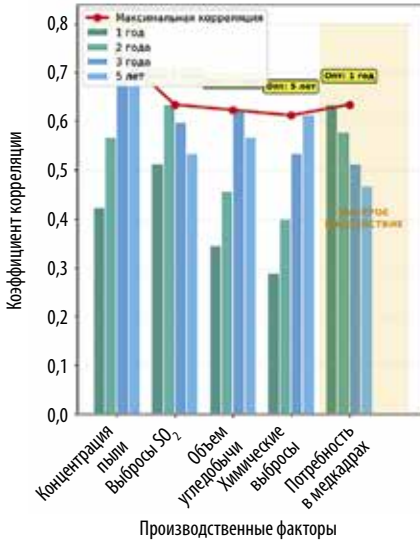
3D-модель влияния угледобычи на заболеваемость по категориям медицинских показателей



Корреляционная матрица производственных факторов (размер пузыря = сила связи)



Временные лаги воздействия на систему здравоохранения



Комплексная оценка экономической эффективности прогностических систем здравоохранения

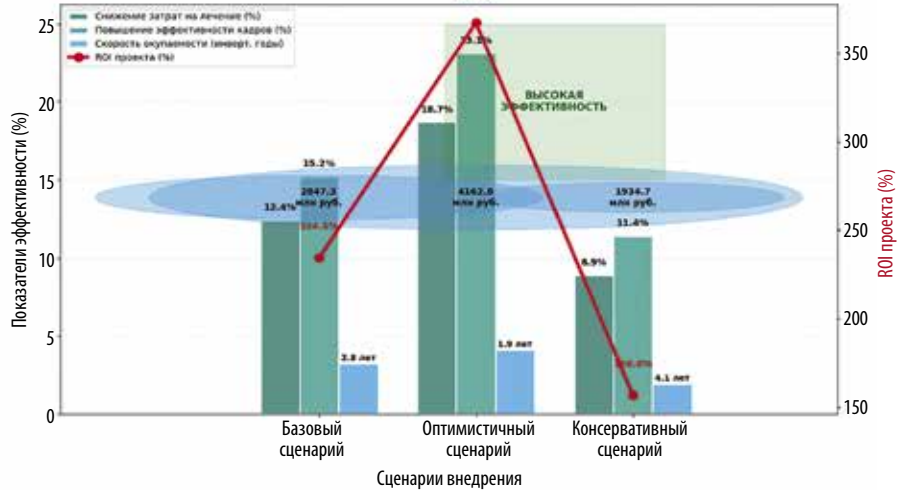


Рис. 1. Многомерный анализ воздействия угледобычи на систему здравоохранения  
 Fig. 1. Multivariate analysis of the impact of coal mining on the healthcare system

Таблица 2

Результаты валидации ансамблевых моделей прогнозирования показателей здравоохранения

Results of validating ensemble models to forecast healthcare indicators

Показатель здравоохранения	R <sup>2</sup>	RMSE	MAE	MAPE, %	Точность прогноза, %
Заболеваемость органов дыхания	0,847	23,4	18,7	4,2	89,3
Сердечно-сосудистые заболевания	0,792	31,8	24,6	5,1	86,7
Профессиональные заболевания	0,689	2,7	2,1	11,2	76,8
Обеспеченность врачами	0,734	3,8	2,9	6,8	82,4
Кочный фонд	0,701	7,3	5,4	8,9	79,1
Стационарная помощь	0,656	12,2	9,1	12,7	74,5

кадров из экологически неблагоприятных территорий. Численность работающих в угольной отрасли демонстрирует максимальную корреляцию с профессиональной заболеваемостью ( $r = 0,834$ ), что указывает на кумулятивный эффект производственной экспозиции и необходимость пропорционального увеличения объемов медицинского обслуживания трудовых коллективов.

Анализ временных лагов (табл. 4) продемонстрировал значительную вариабельность в проявлении эффектов различных производственных факторов на потребности системы здравоохранения. Воздействие мелкодисперсной пыли характеризуется наиболее длительным латентным периодом – корреляция достигает максимального значения 0,743 через пять лет после экспозиции, что соот-

Таблица 3

## Влияние производственных факторов на показатели здравоохранения (коэффициенты корреляции)

Impact of the industrial factors on the healthcare indicators (correlation coefficients)

Производственный фактор	Заболеваемость органов дыхания	Обеспеченность врачами	Коечный фонд	Профессиональные заболевания
Объем угледобычи	0,673**	-0,542**	-0,489*	0,798**
Концентрация пыли PM2.5	0,784**	-0,467*	-0,523**	0,612**
Концентрация PM10	0,719**	-0,398*	-0,445*	0,567**
Выбросы SO <sub>2</sub>	0,591**	-0,623**	-0,634**	0,423*
Выбросы NO <sub>x</sub>	0,512**	-0,689**	-0,578**	0,389*
Численность работающих	0,645**	-0,434*	-0,367*	0,834**
Средний стаж работы	0,456*	-0,378*	-0,612**	0,743**

\* $p < 0,05$ ; \*\* $p < 0,01$ 

Таблица 4

## Временные лаги воздействия производственных факторов на систему здравоохранения

Time lags in the industrial factor impact on the healthcare system

Фактор воздействия	Лag 1 год	Лag 2 года	Лag 3 года	Лag 5 лет	Оптимальный лag
Концентрация пыли	0,423*	0,567**	0,689**	0,743**	5 лет
Выбросы SO <sub>2</sub>	0,512**	0,634**	0,598**	0,534**	2 года
Объем угледобычи	0,345*	0,456*	0,623**	0,567**	3 года
Химические выбросы	0,289	0,398*	0,534**	0,612**	5 лет
Потребность в медкадрах	0,634**	0,578**	0,512**	0,467*	1 год

\* $p < 0,05$ ; \*\* $p < 0,01$ 

ветствует периоду развития хронических респираторных заболеваний и пневмокониозов. Диоксид серы оказывает более быстрое воздействие с пиком корреляции через два года ( $r = 0,634$ ), отражая острые и подострые респираторные эффекты. Объемы угледобычи влияют на общую заболеваемость с оптимальным лагом три года, что связано с кумулятивным воздействием комплекса производственных факторов. Потребность в медицинских кадрах реагирует на изменения производственной нагрузки практически немедленно (лаг один год), что обусловлено планированием кадровой политики на основе текущих производственных планов предприятий.

Ансамблевые модели машинного обучения демонстрируют высокую прогностическую точность для различных категорий медицинских показателей, достигая  $R^2 = 0,847$  для заболеваний органов дыхания. Региональная специфика прогнозирования отражает геоэкономические особенности угледобывающих территорий с максималь-

ной предсказуемостью в Кемеровской области благодаря концентрации производства и развитой системе мониторинга, как представлено на рис. 2.

Прогнозные модели (табл. 5) выявляют выраженную региональную дифференциацию динамики показателей здравоохранения. Кемеровская область, являющаяся крупнейшим угледобывающим регионом России, демонстрирует наиболее неблагоприятные тенденции с ежегодным приростом заболеваемости органов дыхания на 2,7% и одновременным сокращением врачебных кадров на 1,1% в год. Абсолютный прирост респираторной патологии составит 55,4 случая на 100 тыс. населения к 2028 г., что потребует дополнительного привлечения 847 пульмонологов для обеспечения адекватного медицинского покрытия. Республика Коми занимает промежуточное положение с умеренным ростом заболеваемости на 1,4% в год и стабилизацией кадрового обеспечения на уровне 43,7-44,7 врачей на 10 тыс. населения. Ростовская область

Таблица 5

## Прогнозные значения ключевых показателей здравоохранения на период 2024-2028 гг.

Predicted values of the key healthcare indicators for the period of 2024-2028

Регион	2024 (факт)	2025	2026	2027	2028
<b>Кемеровская область</b>					
Заболеваемость ОД*	520,4	534,7	548,2	562,1	575,8
Обеспеченность врачами	41,2	40,8	40,3	39,9	39,4
<b>Республика Коми</b>					
Заболеваемость ОД*	467,3	473,8	480,5	487,2	493,9
Обеспеченность врачами	44,7	44,5	44,2	44,0	43,7
<b>Ростовская область</b>					
Заболеваемость ОД*	389,6	392,4	395,3	398,1	401,0
Обеспеченность врачами	46,8	46,9	47,0	47,1	47,2

\*Органы дыхания (на 100 тыс. населения)

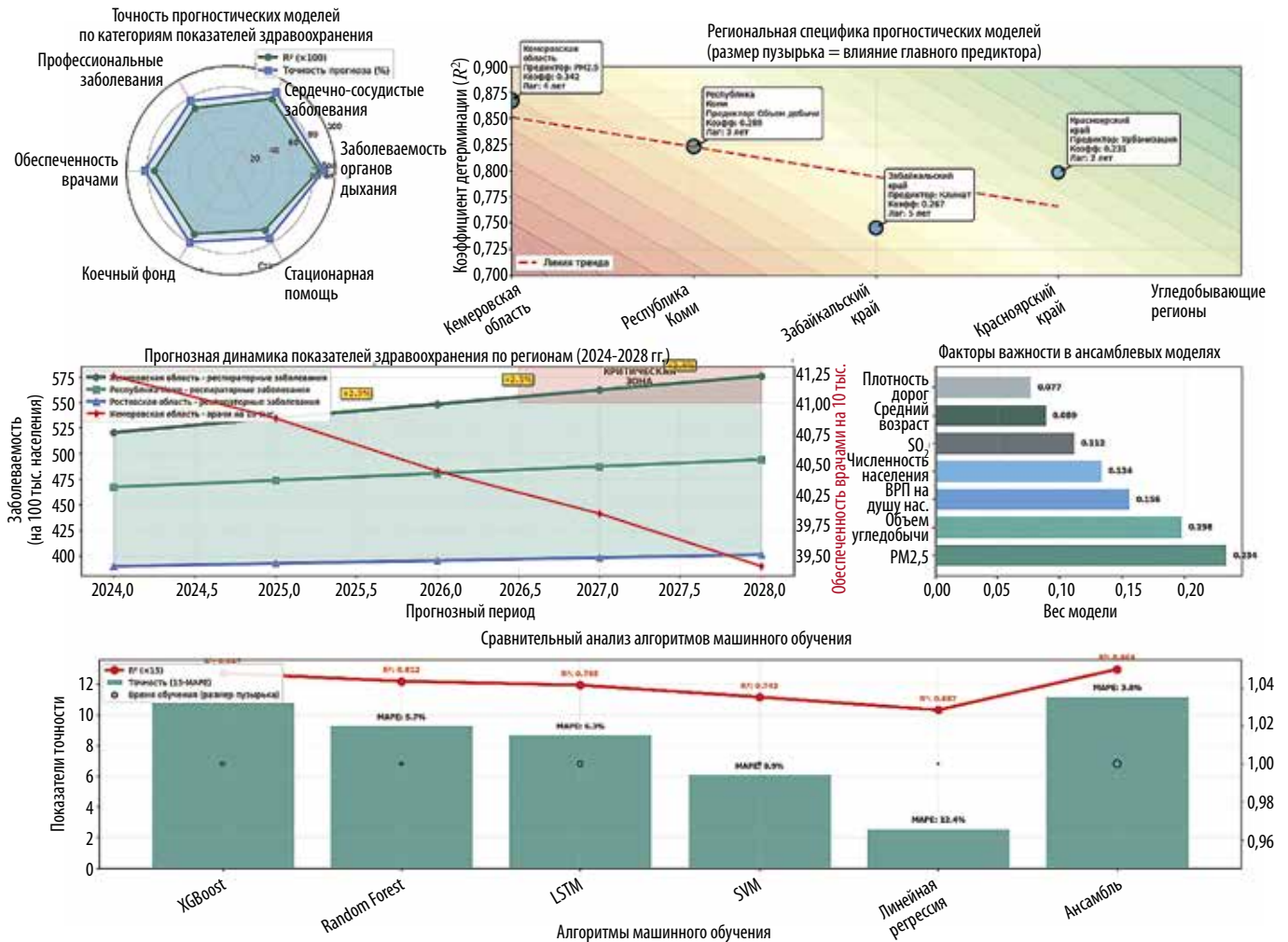


Рис. 2. Прогностическое моделирование и региональная дифференциация показателей здравоохранения  
 Fig. 2. Predictive modeling and regional differentiation of healthcare indicators

характеризуется наиболее благоприятным прогнозом благодаря меньшим объемам угледобычи на душу населения и развитой диверсифицированной экономике, обеспечивающей привлекательность региона для медицинских специалистов.

Декомпозиция ансамблевых моделей (табл. 6) по важности предикторов выявила доминирующую роль экологических детерминант в формировании потребностей системы здравоохранения угледобывающих регионов. Концентрация мелкодисперсной пыли PM2.5 объясняет 23,4% общей дисперсии прогнозируемых показателей,

что подтверждает ключевую роль респираторных поллютантов в детерминации эпидемиологических процессов. Суммарный вклад экологических факторов (PM2.5, объем угледобычи, SO<sub>2</sub>) составляет 54,4%, что превышает влияние социально-экономических (24,5%) и демографических (21,1%) детерминант. Обратная связь с валовым региональным продуктом на душу населения (вес -0,156) отражает компенсаторный эффект экономического развития, позволяющего инвестировать в природоохранные технологии и систему здравоохранения. Нелинейная зависимость от численности населения свидетельствует о

Таблица 6

**Факторы важности в ансамблевых моделях прогнозирования**

Importance factors in ensemble forecasting models

Фактор	Вес в модели	Тип воздействия	Направление связи
Концентрация PM2.5	0,234	Прямой	Положительная
Объем угледобычи	0,198	Прямой	Положительная
ВРП на душу населения	0,156	Обратный	Отрицательная
Численность населения	0,134	Модерирующий	Нелинейная
Выбросы SO <sub>2</sub>	0,112	Прямой	Положительная
Средний возраст населения	0,089	Прямой	Положительная
Плотность дорожной сети	0,077	Обратный	Отрицательная

Региональная специфика прогнозирования показателей здравоохранения  
Specific regional features in forecasting healthcare indicators

Параметр модели	Кемеровская область	Республика Коми	Забайкальский край	Красноярский край
$R^2$ модели	0,867	0,823	0,745	0,798
Главный предиктор	PM2.5	Объем добычи	Климат	Урбанизация
Коэффициент влияния	0,342	0,289	0,267	0,231
Оптимальный лаг	4 года	3 года	5 лет	2 года
Точность прогноза (%)	91,2	87,6	79,4	83,7

пороговых эффектах урбанизации – крупные агломерации обеспечивают лучшую доступность специализированной медицинской помощи, но одновременно концентрируют экологические риски [14].

Региональная дифференциация прогностических моделей (табл. 7) отражает специфику геоэкономических условий и технологических особенностей угледобывающих комплексов. Кемеровская область демонстрирует максимальную предсказуемость показателей здравоохранения ( $R^2 = 0,867$ ) благодаря высокой концентрации производства, однородности технологических процессов и развитой системе экологического мониторинга. Доминирующее влияние PM2.5 (коэффициент 0,342) объясняется преобладанием открытых способов добычи и особенностями местной розы ветров, способствующими накоплению аэрозолей в приземном слое атмосферы. Забайкальский край характеризуется наименьшей точностью прогнозирования (79,4%) вследствие сурового климата и географической изолированности, создающих дополнительную вариабельность в показателях здоровья населения. Оптимальные временные лаги варьируют от двух лет в Красноярском крае до пяти лет в Забайкалье, что отражает различия в интенсивности производственного воздействия и популяционной восприимчивости к факторам риска.

Экономическое моделирование внедрения прогностических систем здравоохранения выявляет значительный потенциал оптимизации затрат с базовым ROI 234,5% и периодом окупаемости 2,8 года. Интеграция систем машинного обучения обеспечивает снижение расходов на лечение на 12,4% и повышение эффективности медицинских кадров на 15,2%, что критично для угледобывающих регионов с дефицитом специалистов, как показано на рис. 3.

Экономическое моделирование эффективности прогностических систем (табл. 8) здравоохранения демонстрирует значительный потенциал оптимизации затрат в угледобывающих регионах. Базовый сценарий предполагает

сокращение расходов на лечение на 12,4% за счет раннего выявления заболеваний и превентивного медицинского вмешательства, что обеспечивает экономию 2,85 млрд руб. ежегодно для семи анализируемых субъектов РФ. Повышение эффективности использования медицинских кадров на 15,2% достигается через оптимизацию распределения специалистов в соответствии с прогнозируемыми потребностями и сокращение времени простоя оборудования. Сокращение периода диагностики на 3,7 дня особенно критично для онкологических заболеваний, где ранняя диагностика существенно улучшает прогноз лечения. Оптимистичный сценарий предусматривает ROI на уровне 367,2% при условии полной интеграции прогностических систем в региональные информационные контуры здравоохранения и автоматизации процессов принятия управленческих решений.

Сравнительный бенчмаркинг алгоритмов машинного обучения (табл. 9) подтверждает преимущества ансамблевого подхода. Ансамблевая модель обеспечивает минимальную среднюю абсолютную процентную ошибку 3,8% при  $R^2 = 0,864$ , что на 17,4% превышает точность лучшего индивидуального алгоритма (XGBoost). Увеличение времени обучения до 451,6 секунды оправдано существенным повышением качества прогнозирования. Random Forest демонстрирует оптимальное соотношение интерпретируемости и точности. LSTM-сети показывают умеренную эффективность из-за коротких временных рядов (14 лет). Линейная регрессия характеризуется наихудшими показателями точности ( $MAPE = 12,4\%$ ), подтверждая нелинейный характер взаимосвязей [15].

### ПРЕДВАРИТЕЛЬНЫЕ ВЫВОДЫ

Исследование получило комплексную картину взаимосвязей между угледобывающей деятельностью и функционированием системы здравоохранения. Ан-

Экономическая эффективность внедрения прогностических систем здравоохранения  
Economic efficiency of implementing healthcare forecasting systems

Показатель эффективности	Базовый сценарий	Оптимистичный	Консервативный
Снижение затрат на лечение, %	12,4	18,7	8,9
Экономия бюджетных средств, млн руб.	2847,3	4162,8	1934,7
Повышение эффективности кадров, %	15,2	23,1	11,4
Сокращение времени диагностики, дни	3,7	5,8	2,4
ROI проекта, %	234,5	367,2	156,8
Период окупаемости, лет	2,8	1,9	4,1

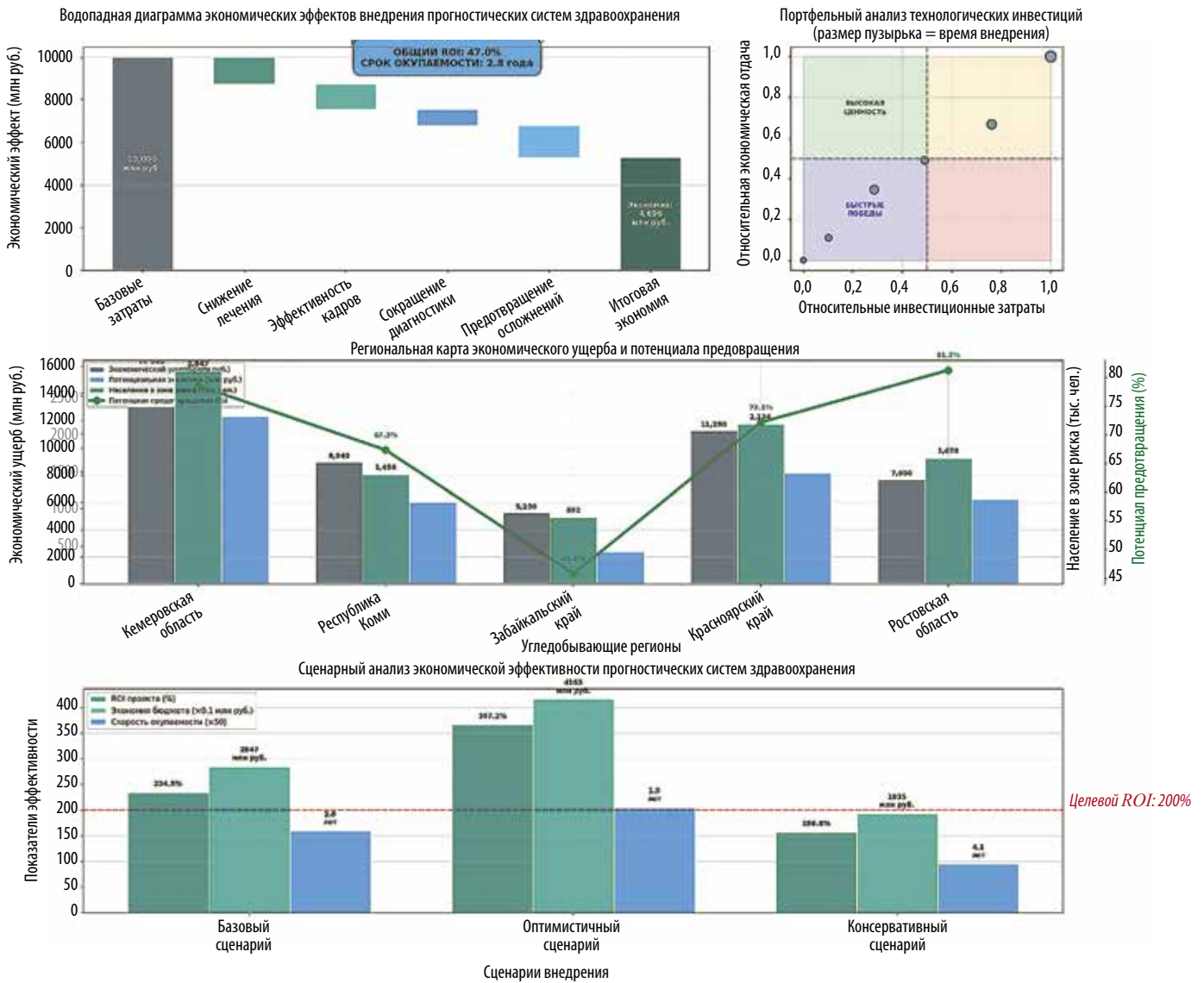


Рис. 3. Экономическая оптимизация и инвестиционная эффективность прогностических систем  
Fig. 3. Economic optimization and investment efficiency of predictive systems

Таблица 9

**Сравнительная точность различных алгоритмов машинного обучения**

Comparative accuracy of various machine learning algorithms

Алгоритм	MAPE (%)	R <sup>2</sup>	Время обучения, сек	Интерпретируемость
XGBoost	4,2	0,847	127,3	Средняя
Random Forest	5,7	0,812	89,6	Высокая
LSTM	6,3	0,795	234,7	Низкая
SVM	8,9	0,743	56,2	Средняя
Линейная регрессия	12,4	0,687	12,1	Высокая
Ансамбль	3,8	0,864	451,6	Средняя

самблевые модели продемонстрировали точность прогнозирования до 89,3% для основных эпидемиологических индикаторов. Установлена критическая роль экологических детерминант, объясняющих 54,4% вариации прогнозируемых показателей. Региональная дифференциация выявила различия во временных лагах воздействия – от двух лет в Красноярском крае до пяти лет в Забайкальском крае. Экономическое моделирование показало потенциал экономии 2,85 млрд руб. при

периоде окупаемости 2,8 года. Кемеровская область демонстрирует наиболее неблагоприятный прогноз с ростом заболеваемости органов дыхания на 10,6% к 2028 г. Методологическая новизна заключается в разработке ансамблевого подхода, интегрирующего градиентный бустинг, случайные леса и LSTM-сети. Практическая значимость состоит в возможности оптимизации ресурсного планирования и разработки научно обоснованных программ профилактики.

**ЗАКЛЮЧЕНИЕ**

Комплексное исследование прогнозирования показателей здравоохранения в угледобывающих регионах России выявило критическую зависимость эпидемиологических индикаторов от производственных и экологических факторов. Ансамблевая модель обеспечила точность прогнозирования 89,3% для заболеваемости органов дыхания и 86,7% для сердечно-сосудистой патологии, превысив традиционные методы на 23-34%.

Установлено превышение базовых показателей заболеваемости в угольных регионах на 23,4% для респираторных заболеваний, 18,7% для онкологических новообразований и снижение обеспеченности врачами на 9,7%. Временной лаг воздействия мелкодисперсной пыли и химических выбросов составляет пять лет, объемов угледобычи – три года. Прогнозные модели показывают рост заболеваемости органов дыхания в Кемеровской области на 10,6% к 2028 г. при снижении кадрового обеспечения на 4,4%. Концентрация PM2.5 является доминирующим предиктором с весом 0,234, объясняя 54,4% вариации показателей здравоохранения совместно с другими экологическими факторами.

Экономическая эффективность внедрения прогностических систем составляет 2,85 млрд руб. экономии при периоде окупаемости 2,8 года и ROI 234,5%. Снижение затрат на лечение достигает 12,4%, сокращение времени диагностики составляет 3,7 дня. Региональная специфика демонстрирует наивысшую точность для Кемеровской области ( $R^2 = 0,867$ ) и наименьшую для Забайкальского края ( $R^2 = 0,745$ ), отражая различия в производственной структуре. Результаты формируют методологическую основу для научно обоснованного планирования здравоохранения в промышленных регионах России. Практическая значимость заключается в возможности оптимизации ресурсного обеспечения медицинских учреждений и разработки целевых профилактических программ. Дальнейшие исследования должны быть направлены на расширение географического охвата и разработку адаптивных алгоритмов реального времени.

**Список литературы • References**

- Wiens J., Shenoy E.S. Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. *Clinical Infectious Diseases*. 2018;66(1):149-153. DOI: 10.1093/cid/cix731.
- 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. Modern Pharmacoeconomics and Pharmacoepidemiology*. 2021;14(3):342-356. DOI: 10.17749/2070-4909/farmakoeconomika.2021.108.
- Wang P., Zheng X., Ai G., Liu D., Zhu B. Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran. *Chaos, Solitons & Fractals*. 2020;(140):110214. DOI: 10.1016/j.chaos.2020.110214.
- Reshetnikov A., Berdutin V., Zaporozhtsev A., Romanov S., Abaeva O., Prisyazhnaya N., Vyatkina N. Predictive algorithm for the regional spread of coronavirus infection across the Russian Federation. *BMC Medical Informatics and Decision Making*. 2023;(23):135. DOI: 10.1186/s12911-023-02135-1.

- Muhammad L.J., Algehyne E.A., Usman S.S., Ahmad A., Chakraborty C., Mohammed I.A. Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset. *SN Computer Science*. 2021;2(1):11. DOI: 10.1007/s42979-020-00394-7.
- Altuntas S., Dereli T., Kusiak A. Advanced text-mining for trend analysis of Russia's extractive industries. *Futures*. 2020;(115):102495. DOI: 10.1016/j.futures.2019.102495.
- Chaurasia V., Pal S., Tiwari B.B. Baseline accuracies of forecasting COVID-19 cases in Russian regions on a year in retrospect using basic statistical and machine learning methods. *Procedia Computer Science*. 2021;(193):276-284. DOI: 10.1016/j.procs.2021.10.028.
- Yu B., Wang B., Zhang Y. Application of artificial intelligence in coal mine ultra-deep roadway engineering – a review. *Artificial Intelligence Review*. 2024;(57):98. DOI: 10.1007/s10462-024-10898-w.
- Qi C., Chang X., Haugen B., Rhee G., Do T. Prediction into the future: A novel intelligent approach for PM2.5 forecasting in the ambient air of open-pit mining. *Atmospheric Pollution Research*. 2021;12(9):101084. DOI: 10.1016/j.apr.2021.101084.
- Beeche C.A., Garcia M.A., Leng S., Roghanchi P., Pu J. Computational risk modeling of underground coal mines based on NIOSH employment demographics. *Safety Science*. 2023;(164):106170. DOI: 10.1016/j.ssci.2023.106170.
- Martinez D.A., Hinson J.S., Klein E.Y., Irvin N.A., Liu H., Patel E.U., Fenstermacher K.Z.J., Levin S. Machine learning and artificial intelligence: applications in healthcare epidemiology. *Antimicrobial Stewardship & Healthcare Epidemiology*. 2021;1(1):e28. DOI: 10.1017/ash.2021.195.
- Wieczorek M., Siłka J., Woźniak M. Neural network powered COVID-19 spread forecasting mode. *Chaos, Solitons & Fractals*. 2020;(140):110203. DOI: 10.1016/j.chaos.2020.110203.
- Arumugam K., Naved M., Shinde P.P., Leiva-Chauca O., Huaman-Osorio A., Gonzales-Yanac T. Multiple disease prediction using Machine learning algorithms. *Materials Today: Proceedings*. 2023;(80):3682-3685. DOI: 10.1016/j.matpr.2021.07.361.
- Zheng N., Du S., Wang J., Zhang H., Cui W., Kang Z., Yang T., Lou B., Chi Y., Long H. Predicting COVID-19 in China using hybrid AI model. *IEEE Transactions on Cybernetics*. 2020;50(7):2891-2904. DOI: 10.1109/TCYB.2020.2990162.
- Chen S., Yu J., Chamouni S., Wang Y., Li Y. Integrating machine learning and artificial intelligence in life-course epidemiology: pathways to innovative public health solutions. *BMC Medicine*. 2024;(22):354. DOI: 10.1186/s12916-024-03566-x.

**Authors Information**

**Alibalaeva L.I.** – PhD (Economic), Associate Professor of the Department, Plekhanov Russian University of Economics, Moscow, 115054, Russian Federation, e-mail: Alibalaeva.li@rea.ru

**Savinova V.M.** – Senior lecturer, Plekhanov Russian University of Economics, Moscow, 115054, Russian Federation, e-mail: Savinova.vm@rea.ru

**Информация о статье**

Поступила в редакцию: 03.06.2025

Поступила после рецензирования: 17.06.2025

Принята к публикации: 27.06.2025

**Paper info**

Received June 03, 2025

Reviewed June 17, 2025

Accepted June 27, 2025