

УДК 519.685+004.04

<https://doi.org/10.21869/2223-1560-2025-29-2-166-185>

Автоматизация процесса прогнозирования данных

О. А. Иващук ¹ ✉, О. Д. Иващук ¹, С. В. Игрунова ²,
Е. В. Нестерова ¹, А. В. Маматов ¹

¹ Белгородский государственный национальный исследовательский университет
ул. Студенческая, д. 14, г. Белгород 308007, Российская Федерация

² Воронежский государственный университет
Университетская пл., д. 1, г. Воронеж 394018, Российская Федерация

✉ e-mail: ivaschuk@bsuedu.ru

Резюме

Цель исследования. Непрерывная модернизация, улучшение методик диагностики требует внедрения инновационных программных решений, современного инструментария, медицинского оснащения и усовершенствованных протоколов обследования, направленных на сбор, хранение, обработку и применение показателей состояния организма, и достоверность медицинской информации. Цель исследования состоит в повышении точности прогнозирования путём разработки оптимальных прогнозных моделей, повышающих надёжность выявленных оценок при диагностике остеопороза с использованием экспертной системы.

Методы. Для достижения поставленных целей использовались параметрические статистические методы для подтверждения обоснованности предположения о нормальной форме распределения всех исследуемых выборочных совокупностей. Основываясь на полученных результатах в качестве наилучшего подхода методики изучения взаимосвязей – корреляционный и регрессионный анализы, для разработки прогностических моделей с определением доверительного интервала предсказания и сопоставления диапазона итогового показателя с вероятностью возникновения болезни.

Результаты. В ходе исследования построены прогнозирующие модели для каждой выделенной группы пациентов совместно с пределами доверительных интервалов прогнозируемых величин ключевых показателей. Определены закономерности взаимозависимости конкретных промежутков целевых факторов и степеней вероятности прогрессирования патологий заболевания. Разработана экспертная система, реализующая полную систему информационного сопровождения, предназначенную для идентификации рисков возникновения заболеваний у пациентов, по предоставленным сведениям, о плотности тканей грудной полости.

Заключение. Разработанная информационная система позволяет осуществлять прогнозирование вероятности болезни для пациентов (мужчин и женщин), возрастные показатели которых выходят за пределы указанного диапазона (10-70 лет), однако достоверность подобных прогнозов неоднозначна, поскольку прогностические модели, составляющие основу функционирования системы, разработаны на базе чей возраст ограничен периодом от 10 до примерно 72 лет.

Ключевые слова: остеопороз; массив данных; регрессионные модели; информационная система; прогнозная модель; алгоритм.

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

© Иващук О. А., Иващук О. Д., Игрунова С. В., Нестерова Е. В., Маматов А. В., 2025

Для цитирования: Автоматизация процесса прогнозирования данных / О. А. Ивашук, О. Д. Ивашук, С. В. Игрунова, Е. В. Нестерова, А. В. Маматов // Известия Юго-Западного государственного университета. 2025; 29(2): 166-185. <https://doi.org/10.21869/2223-1560-2025-29-2-166-185>.

Поступила в редакцию 18.04.2025

Подписана в печать 04.06.2025

Опубликована 23.07.2025

Automation of predicting the risk of osteoporosis

Olga A. Ivashchuk ¹ ✉, Orest D. Ivashchuk ¹, Svetlana V. Igrunova ²,
Elena V. Nesterova ¹, Alexander V. Mamatov ¹

¹ Belgorod State National University
14, Studencheskaya str., Belgorod 308007, Russian Federation

² Voronezh State University
Universitetskaya sq., 1, Voronezh, 394018, Russian Federation

✉ e-mail: ivaschuk@bsuedu.ru

Abstract

Purpose of research. To increase the accuracy of forecasting by developing optimal predictive models that increase the reliability of the identified estimates in the diagnosis of osteoporosis using an expert system.

Methods. In this work, parametric statistical methods were used to confirm the validity of the assumption of the normal distribution of all studied sample populations. Based on the results obtained, correlation and regression analyses are considered as the best approach for studying interrelations, for developing predictive models with determining the confidence interval of prediction and comparing the range of the final indicator with the probability of disease occurrence.

Result. During the study, predictive models were constructed for each identified patient group along with limits of confidence intervals for predicted values of key parameters. Regularities in the mutual dependence between specific target factor ranges and degrees of pathology progression probabilities have been established. An expert information system has been developed, implementing a complete informational support system designed to identify risks of diseases development in patients based on provided data regarding chest tissue density.

Conclusion. The developed information system enables predicting the likelihood of illness for patients (both men and women) whose age exceeds the specified range (10–70 years); however, the reliability of such predictions remains ambiguous since the prognostic models underlying the system's operation are built upon data from individuals aged within the period of approximately 10 to 72 years.

Keywords: osteoporosis; regression models; information system; predictive model; algorithm.

Conflict of interest. The Authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

For citation: Ivashchuk O. A., Ivashchuk O. D., Igrunova S. V., Nesterova E. V., Mamatov A. V. Automation of predicting the risk of osteoporosis // *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta* = *Proceedings of the Southwest State University*. 2025; 29(2): 166-185 (In Russ.). <https://doi.org/10.21869/2223-1560-2025-29-2-166-185>.

Received 18.04.2025

Accepted 04.06.2025

Published 23.07.2025

Введение

Постоянное совершенствование диагностических методов в здравоохране-

нии влечет за собой внедрение новых технологий, инструментов, оборудования и процедур, необходимых для сбора, хранения, извлечения и использова-

ния предикторов о состоянии здоровья и качестве медицинских данных. Это даст возможность сбалансировать полезности полученной информации, ценность данных, затраты ресурсов на диагностику, вычислительную эффективность и надежность вывода при принятии решения.

Здравоохранение отстает с точки зрения цифровизации и инноваций [1,2]. По мере роста числа инструментов системы здравоохранения претерпевают радикальные изменения и технические спецификации, касающиеся индустрии 4.0 [3]. Это в свою очередь мотивирует разработчиков на применение новых технологий, которые оптимизируют процесс принятия медицинских решений и оценки состояния пациента.

Медицинские работники критически оценивают эффективность новых технологий и инструментов, выборочно полагаясь на них там, где их результатам можно с уверенностью доверять, и снизить ошибку человеческого фактора, что позволяет сократить время на постановку диагноза болезни и повысить качество обслуживания.

Наблюдается прогресс в применении новых технологий, таких как больших данных, искусственного интеллекта, телемедицины, блокчейн-платформ применительно к методам оценки и прогноза в области болезней костных и соединительных тканей, в частности остеопороза.

Традиционно врачи сосредоточены на выявлении признаков остеопороза в большей степени у женщин, несмотря на это, остеопороз регулярно диагно-

стируется и у лиц преклонного возраста обоих полов. Остеопороз – это общее нарушение структуры костной ткани, приводящее к повышенной восприимчивости организма к переломам и снижению устойчивости скелета. Среди людей старше 50 лет в течение оставшейся жизни довольно часто случается перелом из-за хрупкости костей.

Объектом исследования стало изучение процессов поддержки принятия решений при оценке вероятности заболевания суставов и хрящей, где наибольший риск здоровью представляет остеопороз. Автоматизация такого рода задач обработки данных обусловлена необходимостью использования больших объемов неоднородных данных для машинного обучения и приложений больших данных. Автоматизированные системы обработки данных, основанные на методах автоматизированного машинного обучения, способны брать необработанные данные и преобразовывать их в полезные функции для задач больших данных, автоматизируя все промежуточные этапы обработки.

Предметом исследования стали подходы, методики способные минимизировать влияние субъективности специалиста при диагностике и принятии решений в неопределённых ситуациях, а также ускорить составление прогноза течения патологического процесса и способствовать улучшению качества оказываемой медицинской помощи.

Группировка данных служит фундаментальным средством интеллектуального анализа информации, используемым для

распределения большого массива данных по отдельным категориям, повышения информативности и оптимизации выявления ключевых зависимостей в объёмах разнородных структурированных и неструктурированных данных. Использование статистических и математических подходов помогает установить разнообразные взаимосвязи между данными на разных уровнях анализа, открывая возможность обнаружить ранее неизвестные тенденции в биопроцессах [4].

Цель исследования состоит в повышении точности прогнозирования путём разработки оптимальных прогнозных моделей, повышающих надёжность выявленных оценок при диагностике остеопороза с использованием экспертной системы.

Материалы и методы

Результирующая оценка, отражающая эффективность любого медицинского исследования, зависит от множества показателей. Малая корреляционная связь результирующей переменной с главными предикторами статически незначима, а это, в свою очередь, требует решать задачу снижения размерности исследуемого набора данных путем выявления переменных, заметно воздействующих на целевую переменную.

Поэтому решается задача устранения избыточных данных за счёт выделения значимых факторов, чьи эффекты сильнее и достовернее всего влияют на итоговые показатели. В источниках [5–6] приведены стандартные подходы к

сокращению многомерности, подлежащих обработке массивов медико-биологической информации.

Анализ показывает, что лишь небольшая доля (около 10%) рассмотренных факторов оказывает реально значимое влияние на полученные результаты, в связи с этим рекомендуется сократить количество факторов, включаемых в прогностические модели, ограничив число предикторов диапазоном 5–6, что соответствует примерно 5–25% от общего количества потенциальных взаимосвязей. Это представляет собой сложную задачу статистического анализа данных.

Для выбора метода снижения размерности необходимо:

- выявить тип распределения массива данных [4–7];
- определить базовые показатели контрольной группы;
- распространение выведенных закономерностей на генеральное множество исследуемых биологических объектов с целью установления ожидаемых значений изучаемых признаков в контрольной группе [8–10];
- моделирование прогнозных значений проверяемых признаков в контрольной группе;
- реализация сравнения разных групп метрик для выявления отличий и сходства между ними;
- определение корреляционной связи между имеющимися в базе данных факторами;
- расчёт коэффициентов регрессии при варьировании входных переменных;

– оценка прогнозного значения конкретного параметра при изменении сопутствующих ему факторов [8-13].

Данные были получены в результате обследования пациентов в Белгородской областной клинической больнице Святителя Иоасафа. Минимальный объем выборки для лиц одного пола составил 270–300 человек, при этом численность каждой из шести формируемых подгрупп должна быть не менее 40 наблюдений. Подобный объем выборки обеспечил достаточную статистическую мощность для оценки характера распределения данных, что позволит обоснованно выбрать методы их обработки в зависимости от соответствия гауссовому, логнормальному или иному типу распределения. Предварительный анализ выявил значимую гетерогенность ($p < 0.05$) возрастной динамики исследуемых биометрических показателей среди мужчин, что потребовало стратификации исходной выборки на две возрастные группы: 10–30 лет и >30 лет. Для последующего анализа каждая возрастная группа была дополнительно разделена на три категории по индексу грудной клетки (Ind): низкий Ind₁ (<1.9), средний Ind₂ (1.9–2.25), высокий Ind₃ (>2.25). Такой подход позволил провести детальный анализ возрастных изменений минеральной плотности костной ткани с использованием оптимизированных регрессионных моделей (линейных и нелинейных) для каждой подвыборки, учитывающих влияние возрастного фактора и морфометрических особенностей грудной клетки.

Поскольку ключевые предикторы (плотность костной ткани первого ребра справа/слева) представляли количественные переменные, то из анализа исключили все порядковые данные. Индекс грудной клетки рассчитывался как отношение сагиттального размера к ширине.

База данных содержала преимущественно ($\geq 99\%$) показатели здоровых лиц, поэтому данные пациентов использовали только для калибровки прогностической модели. Система должна оценивать риск патологии на основе: возрастных референсных интервалов показателей; степени отклонения от средних значений; приближения к нижним границам распределения (ассоциированным с патологией).

Решение задачи снижения размерности полного множества X подразумевает поиск оптимального подмножества M^* из полных данных $M \subseteq X$, удовлетворяющих условию максимального содержания информации $I(M^*)$.

Для разрешения проблемы уменьшения размерности предлагается применить методологию отбора признаков [5-7]. Эта процедура сопряжена с рядом трудностей, ввиду вычислительной сложности и продолжительности операций, необходимых для проверки всех возможных сочетаний переменных. Согласно исследованию [14], выбор факторов в общем виде сводится к выполнению операции $M = F(X)$, позволяющей преобразовать исходное пространство в новую систему предикторов, что известно как «извлечения признаков», при выполнении которой формируется обновленное множество интегрирован-

ных признаков, например, с применением линейного отображения $M = F(X)$.

В процессе извлечения неструктурированные данные множества X преобразуются в более структурированный и удобный формат для повышения качества данных и интерпретируемости модели. Извлечение признаков является подмножеством Z^* проектирования признаков более широкого процесса создания, изменения и выбора признаков в необработанных данных для оптимизации производительности модели. Извлечение признаков подразумевает создание новых признаков, которые по-прежнему захватывают важную информацию из исходных данных, но более эффективным способом.

Существует несколько способов извлечения признаков: метод главных компонент, создающий интегральные предикторы, объединяющие совокупное влияние всех исходных факторов; метод экстремальной группировки, предлагающая заменить выбранную группу признаков общим эквивалентом, аналогичным сумме эффектов указанных признаков [8-13].

Метод главных компонент особенно подходит для извлечения нескольких начальных факторов из большого числа связанных переменных в интегральные предикторы, перед использованием их в другом анализе, таком как множественная регрессия или многомерный дисперсионный анализ.

Анализирует непрерывные переменные путем отбора отдельных предикторов на основе экстремальных оценок вы-

борочного распределения и представления только этих экстремальных оценок для дальнейшего анализа [14].

Выявление значения составленных обобщенных предикторов неприменимо к методу главных компонент, поскольку распределение собственных значений или их отношений зависит от размера выборки, типов данных, размерности шкалы, а исследуемый массив данных имеет различные типы и размерность.

По причине гетерогенности данных тяжело сформировать матрицу коэффициентов Пирсона и оценить необходимое количество подгрупп в массиве [15].

Способ экстремальной группировки – отбор отдельных предикторов на основе экстремальных оценок выборочного распределения и представления только этих экстремальных оценок для дальнейшего анализа. Этот метод выборки используется для достижения большей статистической мощности в последующих проверках гипотез.

Применение данного способа предполагает выполнение следующих шагов: 1 шаг – сформировать корреляционную матрицу парных факторов с результирующей функцией; 2 – разбить на группы исходный массив X ; 3 – выбрать внутри группы из максимальных коэффициентов корреляции максимальные; 4 – факторный коэффициент корреляции между группами выбирается минимальный; 5 – формируется матрица факторов в один столбец, соответствующий равнодействующему предиктору.

Задача селекции признаков решается с помощью задач перебора с помощью задач перебора всех сформированных возможных подмножеств и сравнения их распределений с помощью методов: методы накопленных частот, а также метод Шеннона и метод Кульбака [16,17]. Все эти методы не зависят от числа градаций, они эффективны [18].

Для разработки модели с помощью метода накопленных частот определяется наиболее значимая информативность признаков диагностических данных о заболеваниях, равная максимальному абсолютному значению разности факторов X , принадлежащих различным классам.

Все эти методы часто используются в медицинских информационных системах [18].

Проверка нормальности распределения вариационных рядов проводилась для обоснования применимости параметрических методов анализа данных и выделения подгрупп. Для этого были учтены следующие критерии: значения критерия «хи-квадрат», определяющие согласованность эмпирического распределения с теоретическим нормальным законом; форма гистограмм, отражающая визуальное распределение наблюдений относительно гипотезы о нормальном законе; показатели, такие как коэффициент асимметрии и эксцесс, позволяющие количественно оценить степень отклонения реального распределения от нормального.

Таким образом, сочетание перечисленных подходов позволяет объективно

судить о возможности использования параметрических методов и деления данных на группы.

Это дало основание остановиться на методах корреляционного и регрессионного анализа как наиболее подходящих способах исследования и выдвинуть предположение, что оптимальными формами создаваемых прогнозных моделей станут линейные или полиномиальные регрессионные модели с последующим расчётом доверительных интервалов для прогнозируемых значений факторов – как среднестатистических, так и персональных.

Результаты и их обсуждение

В результате анализа динамики развития остеопороза у мужчин и женщин по имеющимся БД выявлено, что ключевыми атрибутами, определяющими вероятность возникновения и развитие этого заболевания, выступают следующие показатели плотности: оптическая плотность клювовидного отростка лопатки справа (M_1) и оптическая плотность клювовидного отростка лопатки слева (M_2), оптическая плотность бугорка первого ребра справа (M_3), оптическая плотность бугорка первого ребра слева (M_4), оптическая плотность переднего конца костной части первого ребра справа (M_5) и оптическая плотность переднего конца костной части первого ребра слева (M_6). Фиктивные переменные: возрастной фактор пациента (Z_1); гендерная принадлежность (Z_2); конфигурационные особенности грудной клетки (Z_3).

В рамках разработки информационной системы были определены ключевые стадии общего исследовательского процесса: формулировка правил построения базы данных для анализа; выбор приоритетных переменных; установление минимально необходимого объема экспериментальной информации; спецификация формата и границ итоговых обработанных данных; разработка методов представления выводов информационной системы.

Исходя из полученных данных принято решение, что условием качественного научного анализа и разработки эффективных предсказательных моделей является предварительная классификация исследуемых лиц по критериям половой принадлежности, возрастной группы и специфики телосложения.

Начальное исследование приведённых медицинских баз данных (проводилось в рамках магистерской работы под руководством Ивашук О.Д. на кафедре прикладной математики и компьютерного моделирования НИУ «БелГУ») подтвердило недостаточность текущих объёмов выборок и потребовало их серьёзного увеличения. Учитывая, что материалы обо всех прошедших обследовании (мужчины и женщины) делится на три условные группы по индексу грудной клетки, в каждой из этих групп необходимо создать два набора данных, соответствующих указанным возрастным интервалам.

Во избежание потери репрезентативности общая численность обследованных каждого пола должна оставаться на уров-

не не ниже 270–300 человек, а размеры каждой из шести подгрупп обязаны соответствовать показателю не менее 40 человек. Такая стратегия обеспечивает точную оценку формы распределения признаков и правильность выбора методов количественного анализа в зависимости от распределения (нормального или логнормального).

Следующим этапом, понимая, что установленные ранее значащие критерии, подобные M3 и M4, характеризуются численными показателями, из завершённого массива данных ликвидированы все ранговые признаки, существующие изначально, а индекс грудной клетки перешёл к записи действительным числом, выводимым как коэффициент отношения её вертикального измерения к горизонтальному.

Исходя из цели исследования анализируя базу данных, в которой большая часть (более 99%) содержала сведения о здоровых субъектах, следовательно, параметры пациентов с патологией имели лишь второстепенное значение и служили лишь для примерной калибровки автоматизированной системы прогнозирования заболеваемости, основанной на характеристиках здоровых людей. В связи с этим использовались вариативные границы факторов, отклонения которых от среднестатистических значений зависели бы от возрастной группы.

Результаты многоуровневого анализа позволили обосновать необходимость применения методов корреляционного и регрессионного анализа. Наиболее при-

емлемыми стали регрессионные модели – линейные и полиномиальные, сопровождаемые установлением доверительных интервалов для предикторов как в группах, так и индивидуально.

Этап предварительных анализов выявил разнонаправленность тенденций изменения предикторов в возрастных группах: подростковый и молодой возраст (10-30 лет), средний и пожилой возраст (после 30 лет).

Разработка моделей прогнозирования вероятности развития заболевания

Гистограммный анализ распределения подтвердил нормальность данных, что обосновало применение параметрических методов (корреляционный и регрессионный анализ). Наилучшие прогностические модели получены в виде линейных/полиномиальных регрессий с расчетом доверительных интервалов для средних и индивидуальных значений.

Модель прогнозирования вероятности развития заболевания у мужчин отражает заболевание от возраста. Далее независимо для каждой подгруппы проводился анализ, включающий построе-

ние индивидуальных регрессионных моделей, направленных на точное отображение зависимости изменения плотности структур грудной клетки от процесса старения.

Например, при работе с базой данных, содержащей показатели плотности области бугорка первого ребра справа у мужчин младше 30 лет с показателем Ind_1 менее 1.9, было задействовано 45 пациентов. Аппроксимация данных проведена полиномиальной моделью (1) в Matlab R22a, в качестве метода перекрестной проверки брался Kfold.

$$y = -2.5975x + 395.63, \quad (1)$$

где y – показатель оптической плотности M_3 у мужчин младше 30 лет; x – возраст пациента.

Оценка качества регрессионной модели выполнена программой Statistica 14.0. Качество регрессионной модели показано на рис. 1.

Регрессионная модель характеризуется достаточно высоким качеством, является так же, как и регрессионные коэффициенты – достоверной.

Регрессионная статистика				
Множественный R	0.72			
R-квадрат	0.52			
Нормированный R-квадрат	0.51			
Стандартная ошибка	30.18			
Наблюдения	66			
Значимость F	1.09451E-11			
		Коэффициенты	Р-Значение	
		Y-пересечение	395.6330	6.26445E-34
		Переменная X 1	-2.5975	1.09451E-11

Рис. 1. Показатели качества регрессионной модели зависимости плотности области бугорка первого ребра справа у мужчин старше 30 лет с $Ind_1 < 1.9$

Fig. 1. Quality indicators of the regression model depending on the density of the tubercle region of the first rib on the right in men over 30 years old with $Ind_1 < 1.9$

Модель позволила рассчитать среднее прогнозное значение показателя M_3 у здоровых людей, а также определить верхний и нижний пределы допустимого разброса ожидаемых значений средней и индивидуальной плотности с доверительной вероятностью 0.95 [9].

Расчет границ доверительных интервалов для ожидаемого среднего значения $\bar{y}$ и для предсказываемого моделью индивидуального значения $\bar{y}_i$ проводился по формулам:

$$\bar{y} = \tilde{y} \pm t_{0.95;n-2} * m_{\bar{y}}, \quad (2)$$

$$\bar{y}_i = \tilde{y}_i \pm t_{0.95;n-2} * m_{y_i}, \quad (3)$$

где $t_{0.95;n-2}$ – критическое значение t -распределения Стьюдента ($p=0.95$, $df=n-2$);

$$m_{\bar{y}} = S_{\text{ост}} * \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{n * \sigma_x^2}} \text{ и } m_{y_i} = S_{\text{ост}} * \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{n * \sigma_x^2}}$$

стандартные ошибки (СО) вычисленного среднего и индивидуального значения результирующего фактора, $S_{\text{ост}}$ – остаточная (необъясненная) дисперсия результирующего признака, а $\sigma_x^2 = \bar{x}^2 - \bar{x}^2$ – дисперсия признака x .

При получении наиболее оптимальной регрессионной модели в виде полинома 2-ой степени проводилось ее приведение к линейной зависимости $y = \tilde{a}_0 + \tilde{a}_1 \tilde{x}$ путем замены

$$\tilde{x} = (a * x + b)^2, \quad (4)$$

что позволило применить для оценки доверительных границ выражения (2) и (3) с использованием в качестве значений фактора величины, рассчитанные по выражению (4).

Поиск оптимальных регрессионных моделей проводился в прикладном пакете Statistica 14.0, с автоматическим расчетом стандартных ошибок прогнозируемого среднего значения $m_{\bar{y}_i}$. Сравнение полученных значений СО с вычисленными по формуле (2) подтвердило корректность расчета границ доверительных интервалов, что позволило использовать их для формирования диапазонов низкой, средней и высокой вероятностью развития заболевания, которое было проведено с учетом мнений лечащих специалистов в данной области медицины.

Анализ прогноза показателя M_3 с учётными границами доверительных интервалов для средних и индивидуальных значений на уровне 0.95 у мужчин в возрасте свыше 32 года с разными значениями индекса грудной клетки (<1.9, 1.9-2.25, >2.25) демонстрирует регрессионную модель на рис. 2 с Ind_1 менее 1.9.

Значения плотности индекса грудной клетки Ind_1 размещены по оси абсцисс, а возраст обследуемых – по оси ординат.

Согласно регрессионной модели (1), построена таблица уровней показателя оптической плотности M_3 у мужчин старше 30 лет с $Ind_1 < 1.9$, одна из её частей представлена в табл. 1.

Подобным образом созданы регрессионные модели для остальных категорий и возрастных групп по индексу грудной клетки (Ind).

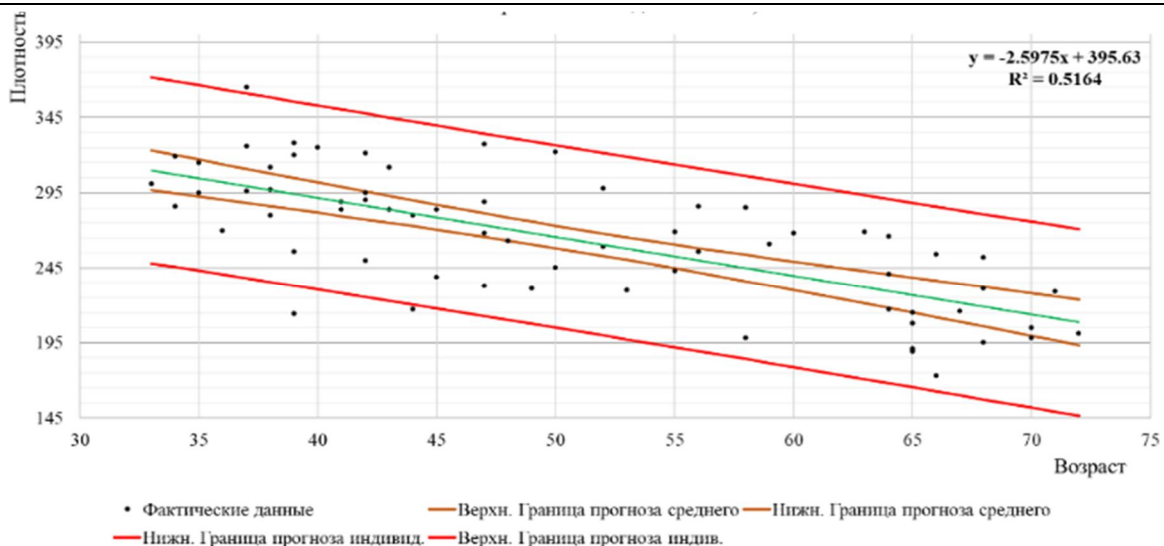


Рис. 2. График регрессионной зависимости показателя оптической плотности M_3 у мужчин старше 30 лет

Fig. 2. A graph of the regression dependence of the optical density index M_3 in men over 30 years of age

Таблица 1. Уровни вероятности заболеть по значению плотности основной группы 1

Table 1. Probability levels of getting sick according to the density of the main group 1

Возраст / Age	Нижняя граница среднего / Lower limit of average	Нижняя граница индивид / Lower limit individual	Вероятность заболеть низкая / Low likelihood of disease	Вероятность заболеть средняя / Medium likelihood of disease	Вероятность заболеть высокая / High likelihood of disease
33	301	249	301-283	283-266	< 266
34	298	246	298-281	281-264	< 264
35	295	244	295-278	278-261	< 261
36	293	241	293-275	275-258	< 258
37	290	238	290-273	273-256	< 256
38	287	236	287-270	270-253	< 253
39	285	233	285-268	268-250	< 250
40	282	231	282-265	265-248	< 248
41	279	228	279-262	262-245	< 245
42	277	225	277-260	260-243	< 243
43	274	223	274-257	257-240	< 240
44	271	220	271-254	254-237	< 237
45	269	218	269-252	252-235	< 235
46	266	215	266-249	249-232	< 232
47	263	212	263-246	246-229	< 229
48	261	210	261-244	244-227	< 227
49	258	207	258-241	241-224	< 224
50	256	205	256-239	239-222	< 222

На рис. 3 показаны прогнозные модели для женщин старше 35 лет. Данные показателя, характеризующего оптическую плотность ребра M_3 и M_4 , обработаны аналогичным способом (учитывая

различия в возрастной структуре и диапазоне индексов грудной клетки), график регрессионных зависимостей вместе с доверительными интервалами представлен визуально на рис. 3.

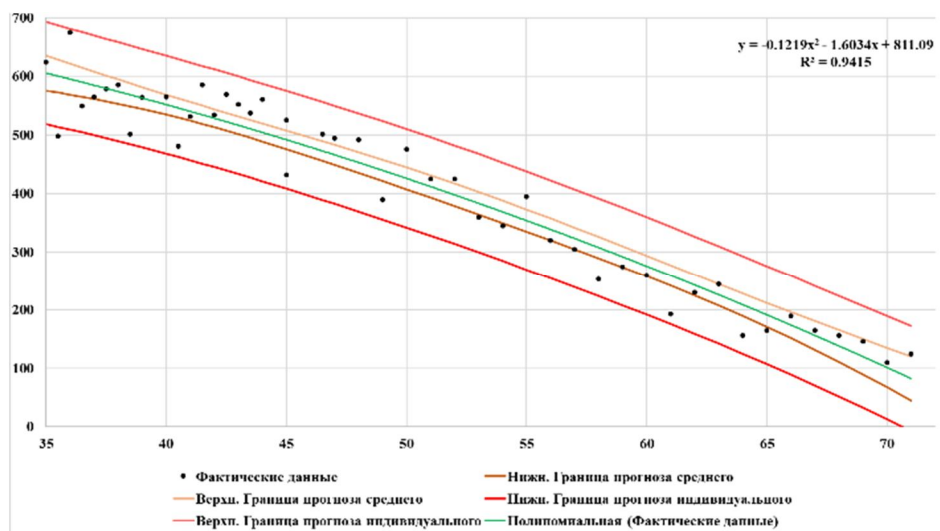


Рис. 3. Диапазоны значений показателя M_3 , соответствующие низкой, средней и высокой вероятности патологии у женщин старше 35 лет

Fig. 3. Ranges of M_3 values corresponding to low, medium and high probability of pathology in women over 35 years of age

Вычисляются прогностические показатели средней оптической плотности M_3 у пациенток без заболеваний. С помощью данной модели рассчитаны прогнозные значения среднего показателя оптической плотности M_3 у здоровых пациенток, а также определены верхние и нижние границы диапазона для среднего и индивидуального значения этого показателя с $r=0.95$, что отражено в табл. 2.

На основе полученных данных для каждой возрастной группы женщин от 36 до 72 лет выделены три диапазона значений показателя оптической плотности M_3 , которым соответствует низкая, средняя и высокая вероятность заболевания.

Для проектирования алгоритмов информационной системы, определяющих риск развития болезни для любого воз-

раста, даже при отсутствии изначальных данных, стали функции, аппроксимирующие рассчитанные значения плотности факторов, задающих границы риска заболеваний. Разработка универсальной структуры и алгоритмизации процессов информационной системы завершила следующий этап реализации задачи.

Алгоритм функционирования автоматизированной системы, включающий согласованные совместно с медицинскими специалистами показатели классификации пациентов в определённые категории, для каждой из которых была разработана соответствующая регрессионная модель, позволяющая прогнозировать вероятность возникновения патологии различной степени выраженности заболевания показан на рис. 4.

Таблица 2. Соотношение оптической плотности M_3 и риска заболевания у женщин старше 35 лет с $\text{Ind}_1 < 1.1$ **Table 2.** The ratio of optical density M_3 and disease risk in women over 35 years of age с $\text{Ind}_1 < 1.1$

Возраст / Age	Нижняя граница среднего / Lower limit of average	Нижняя гра- ница индивид / Lower limit individual	Вероятность за- болеть низкая / Low likelihood of disease	Вероятность за- болеть средняя / Medium likelihood of disease	Вероятность за- болеть высокая / High likelihood of disease
36	571	510	571-551	551-530	530
37	562	500	562-541	541-520	520
38	552	489	552-531	531-510	510
39	542	479	542-521	521-500	500
40	532	468	532-511	511-489	489
41	521	456	521-500	500-478	478
42	510	445	510-489	489-467	467
43	499	433	499-477	477-455	455
44	487	421	487-465	465-443	443
45	475	409	475-453	453-431	431
46	463	396	463-441	441-418	418
47	451	383	451-428	428-406	406
48	438	370	438-415	415-393	393
49	424	357	424-402	402-379	379
50	411	343	411-388	388-366	366
51	397	329	397-374	374-352	352
52	383	315	383-360	360-337	337
53	368	300	368-345	345-323	323
54	353	286	353-331	331-308	308
55	338	271	338-315	315-293	293
56	322	256	322-300	300-278	278
57	306	240	306-284	284-262	262
58	290	224	290-268	268-246	246
59	274	208	274-252	252-230	230
60	257	192	257-235	235-213	213
61	240	175	240-218	218-197	197
62	222	158	222-201	201-180	180
63	204	141	204-183	183-162	162
64	186	124	186-165	165-145	145
65	168	106	168-147	147-127	127
66	149	88	149-129	129-108	108
67	130	70	130-110	110-90	90
68	110	52	110-91	91-71	71
69	91	33	91-71	71-52	52
70	70	14	70-52	52-33	33

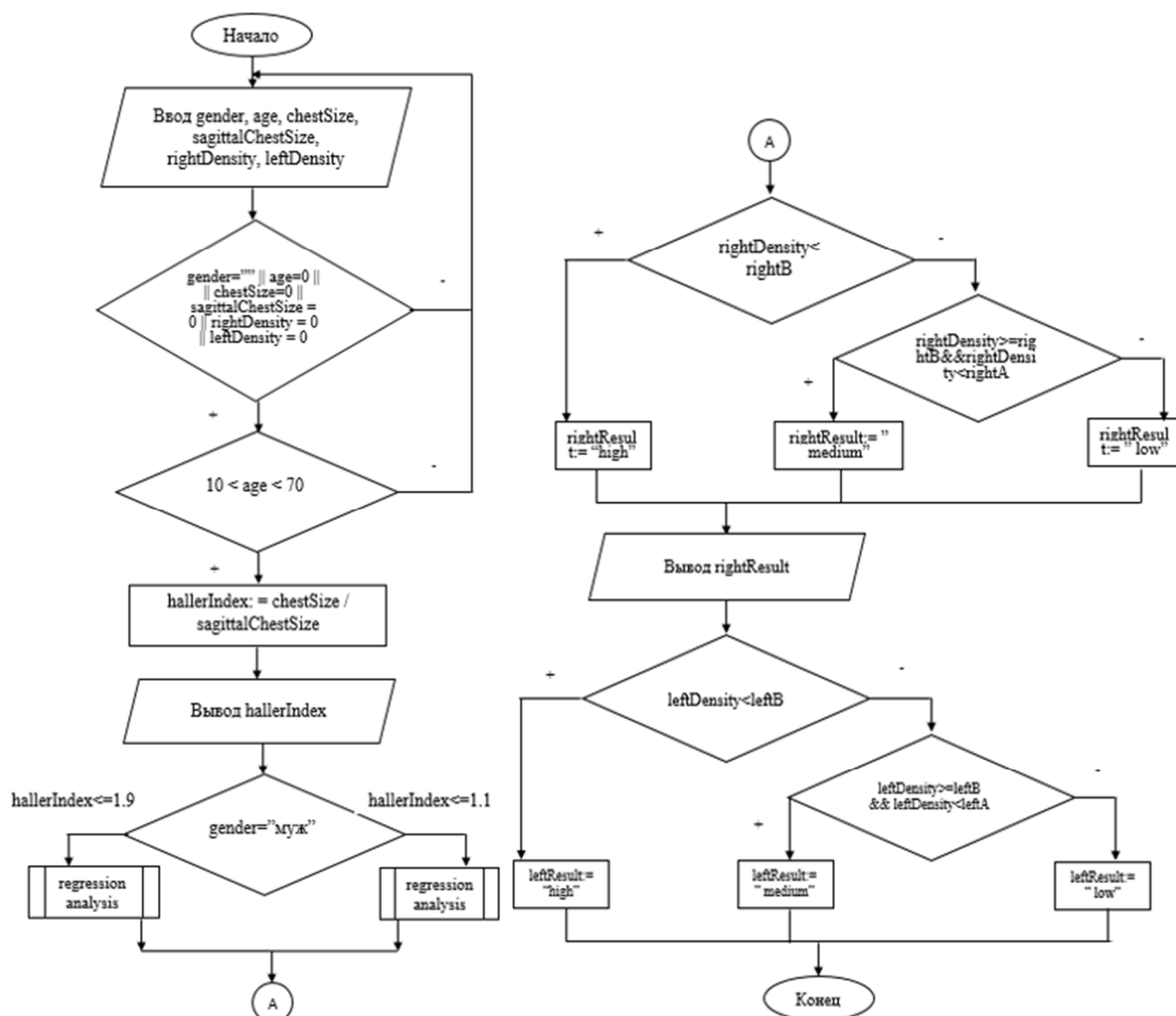


Рис. 4. Фрагменты основного алгоритма функционирования автоматизированной информационной системы оценки риска развития остеопороза

Fig. 4. Fragments of the basic algorithm for the functioning of an automated information system for assessing the risk of osteoporosis

В зависимости от попадания значений плотности элементов грудной клетки в один из ранее установленных диапазонов, программа выдаёт оценку вероятности развития заболевания, что диктует врачу конкретный перечень дальнейших действий – начиная от рекомендаций немедленных профилактических мер, заканчивая необходимостью дополнительного обследования пациента, увеличении

ем частоты мониторинга его медицинских показателей и прочими мерами.

В ходе разработки информационной системы на платформе .NET для Windows-приложений, где XAML использовался для описания интерфейса, логика приложения реализовалась на языке семейства .NET - C# [19-20]. Вид интерфейса с демонстрационным набором данных изображен на рис. 5.

The figure displays two screenshots of the 'COSTA PRIMA' web application, which calculates the probability of disease based on patient data. The interface is divided into 'Введите данные' (Enter data) and 'Результат' (Result) sections.

Top Screenshot (Male Patient):

- Введите данные:**
 - Пол: Мужской
 - Возраст: 56 (Диапазон 10-70)
 - Ширина грудной клетки: 85
 - Сагиттальный размер грудной клетки: 39
 - Плотность области бугорка первого ребра: справа 157, слева 201
- Результат:**
 - Индекс Галлера: 2.18
 - Тип грудной клетки:
 - Грациальная: <1.9
 - Промежуточная: 1.9 - 2.25 (highlighted in green)
 - Трансверциальная: >2.25
 - Оценка вероятности заболевания по плотности бугорка первого ребра:
 - справа: ВЫСОКАЯ (red button)
 - слева: СРЕДНЯЯ (orange button)

Bottom Screenshot (Female Patient):

- Введите данные:**
 - Пол: Женский
 - Возраст: 32 (Диапазон 10-70)
 - Ширина грудной клетки: 70
 - Сагиттальный размер грудной клетки: 35
 - Плотность области бугорка первого ребра: справа 400, слева 250
- Результат:**
 - Индекс Галлера: 2
 - Тип грудной клетки:
 - Грациальная: <1.1
 - Промежуточная: 1.1 - 1.21
 - Трансверциальная: >1.21 (highlighted in green)
 - Оценка вероятности заболевания по плотности бугорка первого ребра:
 - справа: СРЕДНЯЯ (orange button)
 - слева: ВЫСОКАЯ (red button)

Рис. 5. Показатели качества регрессионной модели зависимости показателя оптической плотности M_3 у пациентов старше 30 лет

Fig. 5. Quality indicators of the regression model of the dependence of the optical density index M_3 in patients over 30 years of age

Разработанная информационная система позволяет осуществлять прогнозирование вероятности болезни для пациентов (мужчин и женщин), возрастные показатели которых выходят за пределы указанного диапазона (10-70 лет),

однако достоверность подобных прогнозов неоднозначна, поскольку прогностические модели, составляющие основу функционирования системы, разработаны на базе людей, чей возраст ограничен периодом от 10 до примерно 72 лет. Что-

бы расширить диапазон точного прогнозирования, нужны дополнительные базы данных с людьми за пределами возраста 10-70 лет.

Эффективность работы системы требует систематического сбора информации о большом количестве пациентов и по-

стоянного мониторинга их состояния. Собираемая медицинская статистика используется для уточнения прогнозных моделей и внесения корректировок в правила сопоставления результирующих значений факторам риска.

Список литературы

1. Golinelli D. Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literatur // J Med Internet Res. 2020. Vol. 22, iss. 11. P.12020. DOI: 10.2196/22280.
2. Modelling the continental-scale spread of Schmallenberg virus in Europe: approaches and challenges / S. Gubbins, J. Richardson, M. Baylis, A.J. Wilson, J.C. Abrahantes // Prev Vet Med. 2014; 116(4):404-11. <https://doi.org/10.1016/j.prevetmed.2014.02.004>.
3. Интеллектуальная система оценки качества руды / О.Д. Ивашук, Е.В. Нестерова, С.В. Игрунова, О.О. Ивашук, В.И.Федоров, А.Ю. Родионов // Искусственный интеллект и принятие решений. 2023. №4. С. 94-102. <https://doi.org/10.14357/20718594230409>
4. Zwiener I., Blettner M., Hommel G. Survival analysis: part 15 of a series on evaluation of scientific publications // Dtsch Arztebl Int. Germany. Cologne: Deutscher Ärzte-Verlag. 2011; 108(10): 163-9. <https://doi.org/10.3238/arztebl.2010.0163>.
5. Statistics with Confidence: Confidence Intervals and Statistical Guidelines / D. G. Altman, D. Machin, T. N. Bryant, M.J. Gardner. London: British Medical Journal Publications, 2000. P. 240. doi: 10.1007/s10654-016-0149-3.
6. Prel J.-B., Hommel G., Röhrig B., Blettner M. onfidence Interval or PValue Part 4 of a Series on Evaluation of Scientific Publications // Deutsches Ärzteblatt International. 2009. Vol. 106, № 19. P. 335–339. <https://doi.org/10.3238/arztebl.2009.0335>
7. Гржибовский А.М., Иванов С.В., Горбатова М.А. Однофакторный линейный регрессионный анализ с использованием программного обеспечения Statistica и SPSS // Наука и Здравоохранение. 2017. №2. С. 5-33. <https://doi.org/10.34689/SH.2017.19.2.001>
8. Bewick V., Cheek L., Ball J. Statistics review 14: Logistic regression // Crit Care. 2005; 9 (1): 112–118. <https://doi.org/10.1186/cc3045>
9. Анализ данных методов снижения размерности входных данных / С.Д. Ерохин, В.В. Борисенко, Л.Д. Мартишин, А.С. Фадеев // Т-Comm: Телекоммуникации и транспорт. 2022. №1. С. 30-37. <https://doi.org/10.36724/2072-8735-2022-16-1-30-37>
10. Гржибовский А. М., Иванов С. В., Горбатова М. А. Сравнение количественных данных двух независимых выборок с использованием программного обеспечения Statistica и SPSS: параметрические и непараметрические критерии // Наука и Здравоохранение. 2016. №2. С. 5-28. DOI: 10.34689/SH.2016.18.2.001.

11. Базилевский М. П. Многокритериальный подход к построению модели парно-множественной линейной регрессии // Известия Саратовского университета. Новая серия. Серия Математика. Механика. Информатика. 2021. №1. С. 88-99. <https://doi.org/10.18500/1816-9791-2021-21-1-88-99>
12. Karel G.M. Moons, Douglas G. Altman, Johannes B. Reitsma, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration // Ann Intern Med. 2015;162:W1-W73. . <https://doi.org/10.7326/M14-0698>.
13. Быкова В. В., Катаева А. В. Методы и средства анализа информативности признаков при обработке медицинских данных // Программные продукты и системы. 2016. №2 (114). С. 172-178. <https://doi.org/10.15827/0236-235X.114.172-178>
14. Ajana S., Acar N., Bretillon L., et al. Benefits of dimension reduction in penalized regression methods for high-dimensional grouped data: a case study in low sample size // Bioinformatics. London: Oxford Academic. 2019. 35(19):3628–34. <https://doi.org/10.1093/bioinformatics/btz135>
15. Ялаев Б. И. [и др.] Разработка прогностических клинико-генетических моделей риска развития первичного остеопороза с использованием нейросетевого обучения // Проблемы эндокринологии. 2024. Т.70, №6. <https://doi.org/10.14341/probl13421>
16. Нейросетевые модули с виртуальными потоками для классификации прогнозирования функционального состояния сложных систем / А.В. Киселев, Т.В. Петрова, С.В. Дегтярев, А.Ф. Рыбочкин, С.А. Филист, О.В. Шаталова, В.Н. Мишустин // Известия Юго-Западного государственного университета. 2018. Т. 22, № 4(79). С. 123-134. <https://doi.org/10.21869/2223-1560-2018-22-4-123-134>.
17. Липатова А.В., Потапченко Т.Д. Разработка аналитической системы оценки возникновения рисков здоровью населения на базе алгоритмов машинного обучения // Искусственный интеллект и машинное обучение. Cifra. Компьютерные науки и информатика. 2025. № 1 (5), январь. <https://doi.org/10.60797/COMP.2025.5.3>
18. Forecasting the environmental situation at the purification plants of the enterprise based on fuzzy logic / O. D. Ivashchuk, E. V. Nesterova, S. V. Igrunova, E. V. Kaliuzhnaya, I. V. Udovenko // Journal of Physics: Conference Series, IV International Conference on Applied Physics, Information Technologies and Engineering (APITECH-IV 2022) 2022. Vol. 2388. DOI 10.1088/1742-6596/2388/1/012039.
19. Иващук О.Д. [и др.]. Прогнозирование изменения производительности мельниц обогатительной фабрики при изменении гранулометрического состава питающей руды // Современные наукоемкие технологии. 2023. № 11. С. 33-38. <https://doi.org/10.17513/snt.3981>. URL: <https://top-technologies.ru/ru/article/view?id=39817>. 7
20. Иващук О. А. [и др.]. Разработка моделей прогнозирования эффективности работы валковой дробилки высокого давления на основе регрессионного анализа //СТИН. 2020. №. 6. С. 37-40.

21. Мунс КГМ [и др.]. Прозрачная отчетность о многофакторной предсказательной модели для индивидуального прогнозирования или диагностики (TRIPOD): пояснение и уточнение // Цифровая диагностика. 2022. 3.3. С. 232-322. <https://doi.org/10.15690/vsp.v22i2.2557>

References

1. Golinelli D. Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literatur. *J Med Internet Res.* 2020; 22(11): 12020. DOI: 10.2196/22280..

2. Gubbins S., Richardson J., Baylis M., Wilson A.J., Abrahantes J.C. Modelling the continental-scale spread of Schmallenberg virus in Europe: approaches and challenges. *Prev Vet Med.* 2014; 116(4):404-11. <https://doi.org/10.1016/j.prevetmed.2014.02.004>.

3. Ivashchuk O.D., et al. Intelligent system for assessing the quality of ore. *Iskusstvennyi intellekt i prinyatie reshenii = Artificial intelligence and decision making.* 2023; 4: 94-102 (In Russ.).

4. Zwiener I., Blettner M., Hommel G. Survival analysis: part 15 of a series on evaluation of scientific publications. *Dtsch Arztebl Int. Germany. Cologne: Deutscher Ärzte-Verlag.* 2011; 108(10):163-9. <https://doi.org/10.3238/arztebl.2010.0163>.

5. Altman D. G., Machin D., Bryant T. N., Gardner M.J. Statistics with Confidence: Confidence Intervals and Statistical Guidelines. London: British Medical Journal Publications; 2000. P. 240. doi: 10.1007/s10654-016-0149-3.

6. Prel J.-B., Hommel G., Röhrig B., Blettner M. onfidence Interval or PValue Part 4 of a Series on Evaluation of Scientific Publications. *Deutsches Ärzteblatt International.* 2009; 106 (19): 335–339. <https://doi.org/10.3238/arztebl.2009.0335>

7. Grjibovski A.M., Ivanov S.V., Gorbatoва M.A. Univariante regression analysis using statistica and spss software. *Nauka i Zdravookhranenie = Science & Healthcare Semey.* 2017; (2): 5-33. (In Russ.). <https://doi.org/10.34689/SH.2017.19.2.001>.]

8. Bewick V., Cheek L., Ball J. Statistics review 14: Logistic regression. *Crit Care.* 2005; 9 (1): 112–118. <https://doi.org/10.1186/cc3045>

9. Erokhin S.D., Borisenko B.B., Martishin L.D., Fadeev A.S. Analysis of existing methods to reduce the dimensionality of input data. *T-Comm: Telekommunikatsii i transport = T-Comm.* 2022; 16 (1): 30-35. (In Russ.). <https://doi.org/10.36724/2072-8735-2022-16-1-30-37>.

10. Grjibovski A. M., Ivanov S. V., Gorbatoва M. A. Analysis of quantitative data in two independent samples using Statistica and SPSS software: parametric and non-parametric tests. *Nauka i Zdravookhranenie = Science & Healthcare.* 2016; 2: 5-28. (In Russ.).

11. Bazilevskiy M.P. Multi-criteria approach to pair-multiple linear regression models constructing. *Izvestiya of Saratov University. Mathematics. Mechanics. Informatics.* 2021; 1: 88-99 (In Russ.).

12. Karel G.M. Moons, Douglas G. Altman, Johannes B. Reitsma, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med.* 2015;162:W1-W73. <https://doi.org/10.7326/M14-0698>.
13. Bykova V.V., Kataeva A.V. Methods and tools for analysing informative features when processing medical data. *Programmnye produkty i sistemy = Software & Systems.* 2016; (2):172-178. (In Russ.).
14. Ajana S., Acar N., Bretillon L., et al. Benefits of dimension reduction in penalized regression methods for high-dimensional grouped data: a case study in low sample size. *Bioinformatics. London: Oxford Academic.* 2019; 35(19):3628–34. <https://doi.org/10.1093/bioinformatics/btz135>
15. Yalaev B.I., Novikov A.V., Minniakhmetov I.R., Khusainova R.I. Development of prognostic clinical and genetic models of the risk of low bone mineral density using neural network training. *Problemy endokrinologii = Problems of Endocrinology.* 2024; 70(6). (In Russ.). <https://doi.org/10.14341/probl13421>
16. Kiselev A.V., Petrova T.V., Degtyaryov S.V., Rybochkin A.F., Filist S.A., Shatalova O.V., Mishustin V. N. Hybrid Deciding Modules with Virtual Streams for Classification and Prediction of Functional State of Complex Systems. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University.* 2018; 22(4): 123-134 (In Russ.). <https://doi.org/10.21869/2223-1560-2018-22-4-123-134>.
17. Lipatova A.V., Potapchenko T.D. Razrabotka analiticheskoy sistemy ocenki vozniknoveniya riskov zdorov'yu naseleniya na baze algoritmov mashinnogo obucheniya. *Iskusstvennyy intellekt i mashinnoe obuchenie. Cifra. Komp'yuternye nauki i informatika = Cifra. Komp'yuternye nauki i informatika.* 2025; (1) (In Russ.). <https://doi.org/10.60797/COMP.2025.5.3>.
18. Ivashchuk O.D., Nesterova E.V., Igrunova S.V., Kaliuzhnaya E.V., Udovenko I.V. Forecasting the environmental situation at the purification plants of the enterprise based on fuzzy logic. *Journal of Physics: Conference Series, IV International Conference on Applied Physics, Information Technologies and Engineering (APITECH-IV 2022).* 2022; 2388. DOI 10.1088/1742-6596/2388/1/012039.
19. Ivashchuk O.D., et al. Prognozirovanie izmeneniya proizvoditel'nosti mel'nic obogatitel'noj fabрики pri izmenenii granulometricheskogo sostava pitayushchej rudy. *Sovremennye naukoemkie tekhnologii.* 2023; (11): 33-38 (In Russ.). <https://doi.org/10.17513/snt.3981>. URL: <https://top-technologies.ru/ru/article/view?id=39817>. 7
20. Ivashchuk O. A., et al. Razrabotka modelej prognozirovaniya effektivnosti raboty valkovoj drobilki vysokogo davleniya na osnove regressionnogo analiza. *STIN.* 2020; (6): 37-40 (In Russ.).
21. Muns KGM, et al. Transparent reporting on a multifactorial predictive model for individual forecasting or diagnosis (TRIPOD): explanation and clarification. *Cifrovaya diagnostika.* 2022; 3.3: 232-322. (In Russ.). <https://doi.org/10.15690/vsp.v22i2.2557>

Информация об авторах / Information about the Authors

Ивашук Ольга Александровна, доктор технических наук, профессор, кафедра информационных и робототехнических систем, Белгородский государственный национальный исследовательский университет, г. Белгород, Российская Федерация, e-mail: ivaschuk@bsuedu.ru, ORCID: <http://orcid.org/0000-0002-9383-9141>

Ивашук Орест Дмитриевич, кандидат технических наук, доцент, кафедра прикладной математики и компьютерного моделирования, Белгородский государственный национальный исследовательский университет, г. Белгород, Российская Федерация, e-mail: ivaschuk_o@bsuedu.ru, ORCID: <http://orcid.org/0000-0002-8261-3702>

Игрунова Светлана Васильевна, кандидат социологических наук, доцент, доцент кафедры математического моделирования, Воронежский государственный университет, г. Воронеж, Российская Федерация, e-mail: igrunovasv@mail.ru, ORCID: <http://orcid.org/0000-0002-8780-4700>

Нестерова Елена Викторовна, кандидат экономических наук, доцент, кафедра информационных и робототехнических систем, Белгородский государственный национальный исследовательский университет, г. Белгород, Российская Федерация, e-mail: nesterova@bsuedu.ru, ORCID: <http://orcid.org/0000-0002-2442-5527>

Маматов Александр Васильевич, доктор технических наук, доцент, кафедра информационных и робототехнических систем, Белгородский государственный национальный исследовательский университет, г. Белгород, Российская Федерация, e-mail: mamatovav@bsuedu.ru, ORCID: <http://orcid.org/0000-0001-8933-3159>

Olga A. Ivashchuk, Dr. of Sci. (Engineering), Professor, Information and Robotic Systems Department, Belgorod State National Research University, Belgorod, Russian Federation, e-mail: ivaschuk@bsuedu.ru, ORCID: <http://orcid.org/0000-0002-9383-9141>

Orest D. Ivashchuk, Cand. of Sci. (Engineering), Associate Professor, Information and Robotic Systems Department, Belgorod State National Research University, Belgorod, Russian Federation, e-mail: ivaschuk_o@bsuedu.ru, ORCID: <http://orcid.org/0000-0002-8261-3702>

Svetlana V. Igrunova, Cand. of Sci. (Sociological), Associate Professor, Information and Robotic Systems Department, Voronezh State University, Voronezh, Russian Federation, e-mail: igrunovasv@mail.ru, ORCID: <http://orcid.org/0000-0002-8780-4700>

Elena V. Nesterova, Cand. of Sci. (Economic), Associate Professor Information and Robotic Systems Department, Belgorod State National Research University, Belgorod, Russian Federation, e-mail: nesterova@bsuedu.ru, ORCID: <http://orcid.org/0000-0002-2442-5527>

Alexander V. Mamatov, Dr. of Sci. (Engineering), Associate Professor, Information and Robotic Systems Department, Belgorod State National Research University, Belgorod, Russian Federation, e-mail: mamatovav@bsuedu.ru, ORCID: <http://orcid.org/0000-0001-8933-3159>