Геоинформатика. 2024. № 4. С. 93–118.
Geoinformatika. 2024;(4):93–118.

Искусственный интеллект в прикладных областях знаний

Научная статья
УДК 0048
https://doi.org/10.47148/1609-364X-2024-4-93-118

Обзор и анализ методов объяснительного искусственного интеллекта для решения задач геоэкологического районирования и медицинской профилактики населения

© 2024 г. — Юрий Владиславович Трофимов[1, a)], Алексей Николаевич Аверкин[1, 2, b)], Евгения Наумовна Черемисина[1,3, c)]

[1]Государственный университет «Дубна»; Дубна, Россия

[2]Федеральный исследовательский центр «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН); Москва, Россия

[3]ФГБУ «Всероссийский научно-исследовательский геологический нефтяной институт» (ВНИГНИ); Москва, Россия

[a)]ura_trofim@bk.ru, [b)] averkin2003@inbox.ru, [c)]e.cheremisina@geosys.ru

Аннотация: Данное исследование направлено на изучение применения объяснительного искусственного интеллекта (XAI) в задачах геоэкологического районирования для поддержки устойчивого развития и медицинской профилактики. В условиях возрастающей антропогенной нагрузки на экосистемы и увеличения уровня заболеваемости, вызванного ухудшением экологической обстановки, особое внимание уделено методам раннего выявления рисков. В статье подчеркивается важность XAI для анализа экологических данных с целью снижения воздействия негативных факторов на здоровье населения. Интеграция XAI с географическими информационными системами (ГИС) не только обеспечивает высокую точность геоэкологических прогнозов, но и повышает прозрачность этих прогнозов для экспертов, что способствует принятию более обоснованных решений в области геоэкологии и здравоохранения.

Особое внимание уделено ранней диагностике рисков для здоровья, таких как респираторные и онкологические заболевания, с использованием XAI для анализа экологических данных и медицинских изображений. Объяснительный искусственный интеллект позволяет сделать процесс постановки диагноза более прозрачным и понятным для медицинских специалистов, что повышает доверие к результатам анализа. Внедрение XAI в здравоохранение может не только улучшить точность диагностики, но и оптимизировать ресурсы медицинской системы, перераспределяя их на предупреждение заболеваний.

Обзор существующих систем поддержки принятия решений (СППР) демонстрирует эффективность гибридных моделей, сочетающих нейронные сети и нечеткую логику, для повышения точности и интерпретируемости прогнозов в медицинской сфере. Применение таких моделей открывает новые перспективы для персонализированной медицины, улучшая профилактические стратегии и предоставляя индивидуальные рекомендации на основе комплексного анализа экологических и медицинских данных.

Важной составляющей исследования является территориальное районирование, направленное на управление экологическими рисками и профилактику заболеваний. Такой подход не только снижает нагрузку на системы здравоохранения, но и способствует устойчивому развитию территорий, учитывая влияние экологических и социальных факторов на здоровье населения.

Ключевые слова: ***устойчивое развитие; объяснительный искусственный интеллект (XAI); экологическое районирование; географические информационные системы (ГИС); нейронечеткие модели; профилактика заболеваний; экологические риски; системы поддержки принятия решений; здравоохранение.***

Для цитирования: Трофимов Ю.В., Аверкин А.Н., Черемисина Е.Н. Обзор и анализ методов объяснительного искусственного интеллекта для решения задач геоэкологического районирования и медицинской профилактики населения // Геоинформатика. — 2024. — № 4. — С. 93–118. https://doi.org/10.47148/1609-364X-2024-4-93-118.

Artificial intelligence in applied fields of knowledge

Original article

Review and Analysis of XAI Methods for Addressing Geoecological Zoning and Public Health Prevention Challenges

© 2024 — Yuri V. Trofimov[1, a)], Alexey N. Averkin[1, b)], Eugeniya N. Cheremisina[1, 2 c)]

[1]Dubna University; Dubna, Russia

[2]Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (FRC CSC RAS); Moscow, Russia

[3]All-Russian Research Geological Petroleum Institute (VNIGNI); Moscow, Russia

[a)]ura_trofim@bk.ru, [b)]averkin2003@inbox.ru, [c)]e.cheremisina@geosys.ru

Abstract: This study focuses on the application of Explainable Artificial Intelligence (XAI) in geoecological zoning tasks to support sustainable development and public health prevention. Amid increasing anthropogenic pressures on ecosystems and rising disease rates due to environmental degradation, emphasis is placed on early risk detection methods. The study highlights the

importance of XAI in analyzing ecological data to mitigate the impacts of adverse factors on population health. Integrating XAI with Geographic Information Systems (GIS) not only provides high accuracy in geoecological forecasting but also enhances the transparency of these forecasts for experts, aiding informed decision-making in the fields of geoecology and healthcare.

Special attention is given to early diagnosis of health risks, such as respiratory and oncological diseases, through the use of XAI in analyzing environmental data and medical images. Explainable AI enhances the transparency and understandability of diagnostic processes for medical professionals, fostering trust in analytical outcomes. Implementing XAI in healthcare can not only improve diagnostic accuracy but also optimize healthcare resources by reallocating them toward disease prevention.

A review of existing decision support systems (DSS) demonstrates the efficacy of hybrid models combining neural networks and fuzzy logic to enhance the precision and interpretability of medical forecasts. These models open new prospects for personalized medicine, improving preventive strategies and providing individual recommendations based on comprehensive analyses of environmental and medical data.

A critical aspect of the study is territorial zoning aimed at managing environmental risks and disease prevention. This approach not only reduces the burden on healthcare systems but also promotes sustainable territorial development, taking into account the influence of environmental and social factors on population health.

Key words: *sustainable development; explainable artificial intelligence (XAI); ecological zoning; geographic information systems (GIS); neuro-fuzzy models; disease prevention; environmental risks; decision support systems; healthcare.*

For citation: Trofimov Yu.V., Averkin A.N. Cheremisina E.N. Review and Analysis of XAI Methods for Addressing Geoecological Zoning and Public Health Prevention Challenges. *Geoinformatika.* 2024;(4):93–118. https://doi.org/10.47148/1609-364X-2024-4-93-118. In Russ.

Введение

Устойчивое развитие территорий представляет собой ключевую стратегическую задачу современного общества, в которой сбалансированное взаимодействие между экономическим ростом, социальной стабильностью и сохранением природных ресурсов играет первостепенную роль. Планирование территорий с учетом этих принципов становится важным элементом долгосрочного развития, особенно в условиях роста антропогенных воздействий и изменений климата. Многочисленные ученые подчеркивали важность такого подхода. Еще в начале XX в. В.И. Вернадский заложил основы учения о ноосфере, где устойчивое развитие рассматривалось как часть взаимодействия человека с природой и биосферой, в которой человечество выступает сознательным управляющим элементом. Эти идеи нашли продолжение в трудах других выдающихся ученых, таких как О.Л. Кузнецов и П.Г. Кузнецов, которые долгое время работали над концепцией устойчивого развития, а также ее применения в области стратегического управления территориями и охраны окружающей среды. Значительный вклад в развитие концепции устойчивого развития сделал К.А. Тимирязев, который занимался взаимосвязями между природой и человеческим обществом, уделяя внимание рациональному использованию природных ресурсов и необходимости сохранения экосистем.

Современное развитие общества ставит перед научным сообществом и практиками новые вызовы, связанные с обеспечением устойчивого развития территорий и сохранением здоровья населения. Стратегическое управление территориями в условиях растущего антропогенного воздействия требует инновационных подходов, напрямую связанных с эффективным экологическим менеджментом. В этой связи актуальность приобретает применение современных методов ИИ, в частности, объяснительного искусственного интеллекта (Explainable AI / XAI) для моделирования состояния окружающей среды и прогнозирования ее влияния на здоровье населения.

Устойчивое развитие территорий предполагает баланс между экономическим ростом, социальной стабильностью и экологической безопасностью. Однако достижение этого баланса затруднено из-за сложности и многокомпонентности экологических систем, а также недостаточной прозрачности традиционных методов анализа данных. Внедрение нейросетевых моделей, способных обрабатывать большие объемы разнородных данных и предоставлять объяснимые результаты, открывает новые возможности для геоэкологического районирования и принятия обоснованных управленческих решений. Нейросетевое моделирование окружающей среды позволяет в режиме реального времени отслеживать поступление различных веществ и загрязнителей, а также выявлять скрытые закономерности и тенденции. Использование методов XAI в этом контексте обеспечивает прозрачность и интерпретируемость моделей, что является критически важным для доверия со стороны специалистов и принятия соответствующих мер. Это особенно актуально для медицинской профилактики, где понимание причинно-следственных связей между состоянием окружающей среды и здоровьем населения позволяет разработать эффективные стратегии вмешательства. Более того, стратегическое управление территориями, основанное на данных нейросетевого моделирования, способствует своевременному выявлению экологических рисков и реализации профилактических мероприятий. Это включает в себя не только мониторинг текущего состояния, но и прогнозирование будущих изменений, что позволяет органам управления принимать превентивные меры для снижения негативного воздействия на население и экосистемы.

В данной статье рассмотрено применение XAI для решения задач геоэкологического районирования с целью обеспечения устойчивого развития территорий и медицинской профилактики населе-

ния. Особое внимание уделяется интеграции методов XAI в процессы экологического мониторинга и управления, а также оценке их эффективности на практике. Представленные подходы направлены на улучшение качества принятия решений и повышение эффективности экологического менеджмента в современных условиях.

Концепции интерпретируемого XAI

Современные системы искусственного интеллекта (AI) значительно эволюционируют, переходя от первых поколений экспертных систем к сложным моделям глубокого обучения. Это развитие принесло не только возможности для более точного прогнозирования и анализа данных, но и создало новую проблему — отсутствие прозрачности в принятии решений таких систем. Эта проблема особенно важна в областях, где на основе решений ИИ принимаются критически значимые действия. XAI предлагает решение проблемы объяснимости моделей, концентрируясь на разработке систем, которые принимают решения и одновременно способны объяснять их доступно. Это позволяет пользователям, специалистам и регулирующим органам лучше понимать логику, лежащую в основе выводов ИИ. Данный аспект становится особенно важным в условиях усиления требований к прозрачности ИИ, как это предусмотрено, например, в «Национальной стратегии развития ИИ Российской Федерации» до 2030 г., которая акцентирует внимание на необходимости прозрачности и объяснительности ИИ-систем.

Одними из важных задач XAI являются обеспечение доверия со стороны конечных пользователей и предотвращение ситуаций, когда решения, основанные на сложных моделях, кажутся необоснованными или непрозрачными. В критических контекстах, таких как здравоохранение, прозрачные и объяснимые системы могут значительно улучшить принятие решений, так как специалисты ценят не просто результаты, но и ясность их обоснования. Исследования подчеркивают, что непрозрачные системы ИИ часто вызывают недоверие у пользователей, особенно когда их решения затрагивают критические аспекты жизни, такие как здоровье или безопасность. Хотя модельно-зависимые методы ИИ, такие как глубокие нейронные сети, могут демонстрировать высокую производительность, отсутствие в них объяснимости значительно ограничивает их применение в реальных задачах. Например, согласно исследованию от 2023 г., автора Gillespie и его коллег, недостаток объяснений для предсказаний моделей затрудняет принятие обоснованных решений медицинскими специалистами, что подчеркивает необходимость создания более объяснимых систем ИИ, которые могут предоставлять ясные объяснения своих действий и выводов [1]. В условиях все возрастающих требований к прозрачности и объяснимости ИИ этот аспект становится ключевым для устойчивого применения технологий в здравоохранении и других стратегических областях.

Исторически первые экспертные системы ИИ, такие как системы на основе правил, имели простую структуру и были легко интерпретируемы. Однако с ростом популярности нейронных сетей и других субсимволических методов ИИ, таких как ансамбли моделей, прозрачность значительно снизилась. Модели глубокого обучения, такие как глубокие нейронные сети (DNN), часто рассматриваются как черные ящики, так как их огромное число параметров и сложные структуры затрудняют понимание того, как они принимают решения. Значительный вклад в развитие объяснительного ИИ внес Лотфи Заде, разработавший концепцию нечеткой логики. Нечеткие системы позволяют моделировать человеческое мышление и рассуждения, используя лингвистические переменные и нечеткие правила. Это стало основой для развития нейронечетких систем, объединяющих нейронные сети и нечеткую логику для создания объяснимых моделей. В 2018 г. Агентство передовых оборонных исследовательских проектов (DARPA) запустило программу по развитию объяснительного ИИ, инвестируя значительные ресурсы в создание систем третьего поколения. Целью программы является разработка ИИ, который может интерпретировать и объяснять свои решения, преодолевая ограничения «черного ящика» [2].

Ключевые подходы и методы интерпретации в XAI: от классических, нечетких и гибридных моделей до каскадных моделей нового поколения

ANFIS как гибридный метод прогнозирования и управления в XAI

Первым шагом рассмотрим ANFIS (адаптивную нечеткую систему вывода), которая, хотя и не является классическим инструментом XAI, способна значительно повысить объяснимость сложных моделей за счет своей гибридной архитектуры. Основное преимущество ANFIS заключается в том, что она сочетает возможности нечеткой логики и нейронных сетей. Это делает систему крайне эффективной для решения задач прогнозирования и управления, особенно в тех случаях, когда требуется учитывать многокомпонентные и нелинейные зависимости между параметрами. Несмотря на то, что ANFIS не предназначена исключительно для объяснения решений, как это делается в XAI, ее способность к интерпретации данных через нечеткие правила делает ее важным компонентом для повышения прозрачности решений в сложных системах.

В реальных условиях ANFIS применяется в ряде критически важных областей, где сочетание точности и объяснимости играет ключевую роль. Например, в сфере мониторинга окружающей среды ANFIS используется для точного прогнозирования уровня загрязнения воздуха в городах, где такие предсказания необходимы для обеспечения здоровья насе-

ления и предотвращения негативных последствий для экологии [3]. Также ANFIS активно применяется в задачах водоснабжения, где она помогает прогнозировать качество воды в реках и водоемах, анализируя такие факторы, как количество выпавших осадков, температура воды, а также уровень содержания различных загрязняющих веществ [4]. Благодаря способности ANFIS учитывать множество внешних факторов, ее прогнозы считаются надежными и хорошо применимыми в практике.

Кроме того, ANFIS находит широкое применение в задачах управления энергоэффективностью зданий. В этом контексте его модели используются для прогнозирования энергопотребления в системах отопления, вентиляции и кондиционирования воздуха [5]. Это позволяет эффективно оптимизировать работу систем значительно снижая затраты на энергию и адаптировать модели к изменяющимся условиям.

Не менее значимым является применение ANFIS в промышленности. Одно из исследований продемонстрировало успешную интеграцию ANFIS для управления роботизированными системами на производственных линиях. Это было особенно полезно в условиях, где требовалась высокая степень точности и адаптивности при работе с неопределенностью, что характерно для современных высокоавтоматизированных производств [6]. Гибридный подход ANFIS помогает повысить точность управления и делает систему более гибкой, что важно для сложных динамических процессов.

Одним из ключевых преимуществ использования ANFIS в контексте XAI является его способность преобразовывать сложные нелинейные зависимости между входными параметрами и результатами в легко понимаемые нечеткие правила. Это особенно актуально для таких критически важных задач, как медицинская диагностика и прогнозирование рисков для здоровья населения, обусловленных неблагоприятными факторами окружающей среды. Например, исследования показали, что использование ANFIS для диагностики заболеваний, таких как глазные болезни, диабет, рак и сердечно-сосудистые заболевания, обеспечивает высокую точность прогнозов и прозрачность результатов, что позволяет медицинским специалистам лучше понимать логику работы системы, относится к ней доверительно и принимать более обоснованные решения [7].

Извлечение правил как метод XAI для повышения интерпретируемости нейронных сетей

Наряду с такими гибридными методами, как ANFIS, активно развиваются и другие подходы для обеспечения интерпретируемости моделей ИИ. Одним из наиболее перспективных и востребованных методов Explainable AI (XAI) является извлечение правил из нейронных сетей. Этот подход предоставляет уникальную возможность трансформировать сложные коннекционистские модели глубокого обучения в набор интерпретируемых символических правил. Основное преимущество данного метода заключается в его способности сочетать высокую точность нейронных сетей с прозрачностью символических методов ИИ, что делает его крайне важным инструментом в задачах, где требуются точные прогнозы и понимание логики принятия решений [8].

Методы извлечения правил из нейронных сетей находят широкое применение в различных областях, требующих высокой интерпретируемости решений. Например, этот подход используется в медицинской диагностике, где необходимо предсказать диагноз, а также предоставить медицинским специалистам возможность верифицировать и понять ход работы модели. Для таких специалистов критически важно видеть объяснение предсказаний, чтобы проверять и использовать результаты на практике [9]. Кроме того, в автономных транспортных системах, где необходимо объяснять поведение машины в условиях неопределенности, извлечение правил играет важную роль для повышения доверия к таким системам.

Среди ключевых подходов к извлечению правил из нейронных сетей можно выделить алгоритмы, основанные на нейронечетких системах, и методы, использующие принципы глубокого обучения. Эти подходы обеспечивают преобразование сложных многослойных моделей в интерпретируемые символические представления. Их применение востребовано в сферах, где прозрачность моделей имеет важность, например, в медицинской диагностике и экологическом мониторинге.

Несмотря на множество преимуществ, методы извлечения правил сталкиваются с рядом ограничений. Например, процесс извлечения правил из глубоких нейронных сетей может быть достаточно сложным и требовать значительных вычислительных ресурсов, что может ограничивать их применение в реальном времени [10]. Однако активные исследования в этой области направлены на создание гибридных методов, таких как комбинации нейронных сетей с байесовскими моделями или нечеткой логикой, в частности ANFIS, что помогает снизить вычислительную сложность и повысить объяснимость [11]. Это первостепенно в задачах, где одновременно требуется высокая точность и прозрачность моделей. С математической точки зрения, разработка таких методов предполагает оптимизацию, в которой моделирование направлено не только на достижение минимальных ошибок предсказания, но и на упрощение интерпретации результатов.

Эволюция методов XAI: от экспертных систем до нейронных сетей

Рассмотрение методов неслучайно началось именно с ANFIS и извлечения правил. Эти подходы являются примерами ранних методов объяснимости, которые использовались в экспертных системах — одной из первых форм ИИ. В экспертных системах основное внимание уделялось тому, чтобы решения, принимаемые системой, могли быть объяснены и поняты специалистами. Когда системы на основе правил и нечеткой логики были основными инструментами, объяснимость занимала центральное место. Однако с развитием глубокого обучения и увеличением сложности моделей возникла новая потребность — необходимость объяснять решения «черных ящиков» нейронных сетей. Современные нейронные сети и ансамблевые методы обладают высокой точностью, но их внутренняя работа остается непрозрачной для пользователя. Именно здесь вступают в игру современные методы XAI, разработанные специально для объяснения сложных моделей.

Поэтому, прежде чем переходить к рассмотрению более свежих методов, важно рассмотреть классификацию существующих подходов XAI. Это позволит лучше понять, какие методы лучше подходят для различных задач и моделей. Разные типы методов XAI могут применяться в зависимости от специфики модели, степени ее сложности и требований к прозрачности.

Введение в классификацию методов XAI: от специализированных методов к универсальным решениям

Классификация основана на нескольких ключевых характеристиках. В первую очередь, стоит обратить внимание на зависимость метода от структуры модели: есть методы, которые не требуют доступа к внутренней структуре модели, и методы, для которых такая информация необходима. Следующим критерием также является тип генерируемых объяснений: глобальные или локальные. Глобальные объяснения предоставляют общую картину работы модели, тогда как локальные объяснения фокусируются на интерпретации отдельных предсказаний. Еще один важный аспект — момент генерации объяснений: одни методы разработаны для объяснения уже обученных моделей *(пост-хок методы*), другие же изначально спроектированы как объяснимые.

Модельно-независимые методы. SHAP (Shapley Additive Explanations) один из наиболее известных в задачах объяснения моделей машинного обучения, независимых от их внутренней структуры. Основываясь на теории кооперативных игр, SHAP позволяет количественно оценивать вклад каждого признака в формирование предсказания. Универсальность метода позволяет применять его как к простым моделям, так и к глубоким нейронным сетям, что делает SHAP значимым инструментом для решения задач в междисциплинарных исследованиях. Эффективность этого подхода подтверждена в различных областях, таких как финансовый сектор, медицинская диагностика, эколого-географический анализ и др., где необходима оценка точности и объяснимости прогнозов. SHAP объединяет существующие подходы к объяснению моделей и вводит новые алгоритмы, которые обеспечивают более высокую вычислительную производительность и лучшее соответствие человеческой интуиции по сравнению с ранее разработанными методами [12]. В последние годы особое внимание уделяется адаптации SHAP для решения специфических задач, включая экологическое прогнозирование и оценку загрязнения окружающей среды, где требуется учитывать сложные взаимосвязи факторов. Объяснимость моделей здесь является основой, поскольку она помогает оценивать предсказания и понимать причины, лежащие в их основе. Это помогает вырабатывать научно обоснованные решения, для разработки стратегий управления рисками и обеспечения доверия со стороны экспертов из различных дисциплин [13].

Еще одним универсальным и *модельно-независимым методом* является LIME (Local Interpretable Model-Agnostic Explanations). Этот метод позволяет создавать локальные объяснения для конкретных предсказаний модели. LIME строит объяснимую модель вокруг каждого предсказания, что делает его полезным инструментом для анализа решений сложных моделей, таких как нейронные сети и ансамблевые методы. LIME активно применяется для оценки рисков, связанных с воздействием экологических факторов на здоровье человека, а также в задачах прогнозирования изменений климата и моделирования последствий антропогенной деятельности [14, 15]. Способность LIME работать с моделями «черного ящика» делает его особенно ценным в задачах, где требуется объяснение сложных предсказаний на локальном уровне [16].

Помимо этих методов, можно выделить и другие подходы. Например, метод Anchors создает интерпретируемые «якоря», объясняя, при каких условиях модель принимает конкретное решение. Этот подход получил признание в задачах, где важно обеспечить высокую степень объяснимости без ущерба для точности модели. Другой подход, контрфактические объяснения (Counterfactual Explanations), позволяет пользователю понять, какие изменения во входных данных могли бы привести к другому предсказанию модели. Например, контрфактические объяснения позволяют выявить, какие изменения в показателях пациента (таких, как уровень глюкозы или кровяное давление) могут привести к улучшению прогноза, а методы Anchors обеспечивают прозрачность при объяснимости рекомендаций, основанных на диагностических моделях.

Модельно-зависимые методы требуют доступа к внутренней структуре модели, что позволяет более глубоко анализировать процессы, происходящие внутри модели. Примером таких методов является визуализация активаций в нейронных сетях, данный способ позволяет исследовать, как различные слои нейронной сети влияют на предсказание. Этот метод особенно эффективен в задачах компьютерного зрения и медицинской визуализации, где требуется детальное понимание механизмов принятия решений моделью. Методы визуализации, разработанные для классификационных моделей на основе глубоких сверточных нейронных сетей (ConvNets), успешно применяются в медицине. Например, визуализация активаций позволяет выявить особенности медицинских изображений, такие как аномальные ткани или патологические структуры, которые оказывают влияние на предсказания модели. Первый подход, основанный на генерации изображений, максимизирующих оценку класса, помогает понять, как модель воспринимает диагностические категории, такие как злокачественные и доброкачественные образования. Второй метод, использующий карты значимости, позволяет оценить вклад каждого пикселя изображения в формирование предсказания, что особенно важно для задач диагностики и локализации патологий. Эти методы также могут быть использованы для слабоконтролируемой сегментации медицинских изображений, например, при выделении областей интереса в рентгеновских снимках или МРТ-сканах. Методы визуализации, изначально разработанные для анализа моделей на наборах данных, таких как ImageNet challenge [17], могут быть адаптированы или являться образцом для решения медицинских задач, включая диагностику и локализацию патологий. Анализ связи между визуализацией на основе градиентов и деконволюционными сетями помогает лучше понять механизмы работы моделей, расширяя их применение в медицинской практике [17].

Анализ веса признаков является инструментом, позволяющим определить, какие факторы оказывают наибольшее влияние на предсказания модели. Поскольку этот метод требует доступа к внутренним параметрам и структуре модели, он относится к модельно-зависимым подходам. Анализ веса признаков активно применяется в задачах прогнозирования экологических рисков и оценки климатических изменений. Случайные леса, представляя собой комбинацию древовидных классификаторов, предоставляют возможность оценки важности переменных через внутренние метрики, такие как уменьшение ошибок, сила деревьев и их корреляция. Использование случайного выбора признаков для разделения узлов снижает ошибку обобщения и повышает устойчивость метода к помехам [18]. Для задач регрессии и классификации также широко используется градиентный бустинг деревьев, который позволяет создавать устойчивые и объяснимые модели. Этот метод, опирающийся на оптимизацию в пространстве функций, эффективно справляется с обработкой сложных и шумных данных. Кроме того, алгоритмы, основанные на градиентном бустинге, включают встроенные механизмы интерпретации, которые делают их полезными для понимания предсказаний и анализа влияния признаков, особенно в междисциплинарных исследованиях [19]

Пост-хок методы в свою очередь, предназначены для анализа моделей после их обучения. Эти методы полезны для сложных моделей, таких как глубокие нейронные сети, которые изначально не проектировались как объяснимые. SHAP и LIME также могут быть отнесены к пост-хок методам, так как они позволяют анализировать уже обученные модели, предоставляя как глобальные, так и локальные объяснения [12].

В противоположность пост-хок методам, существуют изначально объяснимые модели, которые были спроектированы с учетом необходимости прозрачности. Примеры таких моделей включают линейные регрессии, деревья решений и нейронечеткие системы [20]. Эти модели являются инструментом объяснения в мультидисциплинарных областях, где требуется высокая точность предсказаний. Например, в экологическом мониторинге они позволяют анализировать данные дистанционного зондирования, включая идентификацию объектов, интерпретацию форм и оценку текстур, что значительно улучшает понимание сложных процессов в окружающей среде. Благодаря своей прозрачной архитектуре, объяснимые модели широко применяются для решения задач управления сточными водами, твердыми коммунальными отходами и создания систем раннего оповещения, что позволяет эффективно реагировать на экологические угрозы. Их преимущества заключаются в способности обеспечивать простые объяснения «по умолчанию», делая результаты доступными для специалистов из разных дисциплин. Такие модели способствуют разработке устойчивых решений, обеспечивая точность, скорость и адаптивность в режиме реального времени [21].

Применение SHAP и LIME как универсальных методов XAI

Возвращение к выделению методов SHAP и LIME в классификации методов XAI, объясняется их уникальной гибкой способностью обеспечивать объяснимость различных моделей «черного ящика», например, глубокие нейронные сети и ансамблевые методы. Оба подхода доказали свою универсальность, применяясь к широкому спектру различных междисциплинарных областей. Рассматривая применение методов SHAP и LIME в контексте универсальных методов, стоит обратить внимание на возможность генерировать как глобальные, так и локальные объяснения. Также гибкость в применении к разным типам данных и моделям. Они

подходят как для работы с линейными моделями (например, линейная регрессия), так и с нелинейными архитектурами (например, случайные леса или градиентный бустинг). Это обеспечивается за счет их алгоритмической адаптивности, SHAP может быть применен к глубоким нейронным сетям для анализа вкладов нейронов, а LIME позволяет объяснять ансамблевые модели путем построения локальных линейных аппроксимаций. Вдобавок оба метода разработаны с учетом подробностей пользователей, не обладающих глубокими техническими знаниями. Говоря об интуитивности и прозрачности обоих методов — современные реализации SHAP, такие как TreeSHAP, оптимизированы для работы с большими наборами данных и сложными моделями, что делает их применимыми в условиях ограниченных вычислительных ресурсов. LIME, в свою очередь, позволяет эффективно строить локальные модели за счет выборки подмножеств данных, что снижает нагрузку на систему.

Говоря об универсальности SHAP, важно отметить его применение в междисциплинарных и мультидисциплинарных областях. Например, в социальной аналитике метод позволяет объяснять модели прогнозирования поведения групп населения, а в финансовом секторе — анализировать факторы, влияющие на кредитные риски или инвестиционные стратегии. Кроме того, SHAP нашел широкое применение в таких дисциплинах, как медицина и экология. В медицине SHAP используется для выявления биомаркеров, анализа факторов риска заболеваний и объяснения моделей прогнозирования клинических исходов, что делает его незаменимым инструментом для персонализированной медицины. Яркий пример использования SHAP в задачах медицинской диагностики — это выявление факторов риска для заболеваний, таких как диабет и сердечно-сосудистые болезни [22]. Эти задачи требуют обьяснимых решений, и SHAP позволяет медицинским специалистам лучше понимать, какие факторы сыграли решающую роль в предсказании модели. Применение SHAP в экологии также подтверждает его универсальность. В прогнозировании уровня загрязнения воздуха и воды, где множество факторов взаимодействуют между собой, SHAP помогает выявить те, которые оказывают наибольшее влияние на конечные результаты модели. Например, в работе Ji Y. et al. SHAP использовался для оценки воздействия загрязняющих веществ на респираторные заболевания. Благодаря методу удалось подробно описать вклад каждого фактора, что стало важным шагом в управлении рисками для здоровья населения и принятии экологических решений [23]. Это подтверждает, что SHAP успешно решает задачи, связанные с анализом сложных экологических систем, где требуется точное понимание вклада различных параметров в конечный результат [15].

Метод LIME (Local Interpretable Model-Agnostic Explanations) продемонстрировал свою практическую значимость в задачах, где требуется объяснение отдельных предсказаний. LIME активно применяется в медицинской диагностике для анализа индивидуальных случаев, например, при прогнозировании заболеваний или оценке рисков для здоровья, связанных с воздействием экологических факторов [14]. Один из ключевых примеров применения LIME — прогнозирование климатических изменений, где объяснимость локальных предсказаний помогает лучше понимать последствия природных катастроф и управлять ими [13]. LIME позволяет сосредоточиться на объяснении конкретных решений, что делает его полезным для анализа моделей «черного ящика» без ущерба для точности предсказаний. Использование как LIME, так и SHAP повышает объяснимость моделей для разноплановых задач, например при анализе экологических рисков, прогнозировании последствий изменений климата и обосновании медицинских решений [15, 24]. Практические примеры демонстрируют, что SHAP и LIME позволяют строить предсказания и создавать доверительные взаимосвязи с экспертами и конечными пользователями.

Деревья решений и ансамблевые методы в задачах XAI

После рассмотрения методов SHAP и LIME, нужно отдельно рассмотреть другой класс, объяснимых моделей — деревья решений и ансамблевые модели. Деревья решений являются одним из наиболее прозрачных методов машинного обучения, благодаря их наглядности. Они представляют процесс принятия решений в виде древовидной структуры, где каждое условие представляет собой разделение данных на основе одного из признаков, а каждый путь от корня до листа можно интерпретировать как правило «если-то». Такая структура делает деревья решений легко понятными для конечных пользователей, что особенно важно в задачах, требующих прозрачности моделей, таких как экологический анализ или здравоохранение [18]. В экологическом анализе деревья решений часто используются для определения ключевых факторов, влияющих на уровень загрязнения воздуха, воды или почвы. Например, деревья решений успешно применяются для оценки вклада антропогенных факторов, таких как выбросы вредных веществ с промышленных предприятий и транспорта, в загрязнение окружающей среды, что позволяет разработать меры по снижению этих выбросов [25]. Деревья решений используются для прогнозирования распространения загрязняющих веществ в реках или атмосфере, анализируя такие параметры, как скорость ветра, температура и осадки [26]. Применение ансамблевых методов в экологическом мониторинге используются для прогнозирования изменения уровня загрязнения воздуха на основе множества параметров, таких как концентрация

загрязняющих веществ, метеорологические данные и интенсивность дорожного движения. Например, в работе Jiang случайный лес использовался для оценки воздействия промышленных выбросов на качество воздуха в густонаселенных районах [26]. Результаты показали, что такой подход позволяет прогнозировать концентрации загрязняющих веществ и оценивать вклад различных факторов, таких как скорость ветра и температурные градиенты. Аналогично, в задачах оценки качества воды в реках и озерах ансамблевые методы могут использоваться для прогнозирования уровня загрязнения в зависимости от таких факторов, как сельскохозяйственные стоки, количество осадков и сезонные изменения. В работе Boulesteix показано, что использование градиентного бустинга для прогнозирования качества воды обеспечивает высокую точность и позволяет выделить ключевые факторы, влияющие на качество воды в различных временных промежутках [25].

В то время как деревья решений являются легко объяснимыми моделями, ансамблевые методы, такие как случайный лес и градиентный бустинг, часто обладают более высокой точностью, но могут терять объяснимость. Эти методы строят множество деревьев решений и объединяют их результаты, что позволяет улучшить предсказательную способность модели. Однако с ростом числа деревьев и сложности их взаимодействий анализ всей модели становится затруднительным [27]. Тем не менее, существуют подходы, позволяющие вернуть интерпретируемость ансамблевым методам. Одним из наиболее популярных методов является *Permutation Feature Importance*. Этот метод основан на идее, что важность признака можно измерить, наблюдая, как ухудшается качество предсказаний модели при случайной перестановке значений этого признака [19]. Важность признаков в случайном лесе или градиентном бустинге позволяет понять, какие факторы оказывают наибольшее влияние на предсказания модели, что делает его полезным в задачах экологического анализа, например, при оценке факторов загрязнения воздуха и воды [13]. Для повышения интерпретируемости ансамблевых методов разрабатываются различные визуальные и аналитические инструменты. Например, *Partial Dependence Plots* (PDP) позволяют исследовать как один или два признака влияют на предсказание модели, считая остальные признаки постоянными [28]. Это может быть полезно для визуализации воздействия отдельных экологических факторов на уровень загрязнения. PDP широко используются для анализа моделей случайного леса и градиентного бустинга, применяемых в экологических исследованиях, таких как оценка риска наводнений или прогнозирование климатических изменений [29]. Также популярностью пользуются *Individual Conditional Expectation* (ICE) графики, которые представляют собой расширение PDP, показывающее влияние признака на предсказания для каждого наблюдения в отдельности [13]. Это особенно полезно в задачах прогнозирования последствий изменения климата, где каждое наблюдение (например, город или регион) может иметь уникальные характеристики и быть подвержено различным уровням воздействия экологических факторов.

Математические основы интерпретируемости моделей в XAI: от вероятностных методов к нелинейным зависимостям

Рассмотрим математические основы, которые обеспечивают обоснованную и теоретически подкрепленную объяснимость моделей. В предыдущем разделе были рассмотрены различные методы как примеры хорошо известных подходов, обеспечивающих определенный баланс между точностью и объяснимостью. Переход к математическим основам в этом разделе необходим для дальнейшего понимания новизны в области сложных горизонтальных иерархий XAI, которая будет представлена в следующих блоках статьи.

Одной из причин перехода к математическим основам является необходимость анализа того, как базовые теории, применяемые в XAI, формируют фундаментальные подходы к объяснимости сложных моделей, в том числе таких, как ансамблевые методы и нейронные сети. Вероятностные методы, как было ранее отмечено, играют ключевую роль в объяснении неопределенности, что важно при работе с данными, содержащими шум или неполноту. Например, байесовские сети, которые являются вероятностными графическими моделями, позволяют моделировать сложные зависимости между признаками [30]. Байесовский вывод, как один из центральных вероятностных подходов, позволяет обновлять распределения параметров моделей по мере поступления новых данных. Этот процесс важен для разработки моделей, которые могут адаптироваться к изменяющимся условиям, что делает вероятностные методы необходимыми при построении и объяснении сложных систем, таких как ANFIS, о которых упоминалось ранее [30].

Если говорить о теории информации, то невозможно переоценить ее важнейшую роль в понимании того, как модели оценивают неопределенность и выбирают ключевые признаки. Например, энтропия Шеннона используется для измерения неопределенности предсказаний модели и помогает определить, какие признаки имеют наибольшую информативность. Это понятие также активно применяется в методах выбора признаков для деревьев решений, где информационный выигрыш используется для определения важности признаков [31]. Взаимная информация играет ведущую роль в оценке того, насколько конкретный признак связан с целевой переменной, что помогает понять, какие из признаков являются ключевыми для точности предсказаний. Можно сказать, что теория

информации является основой для построения и интерпретации сложных моделей, где важно учитывать информационные взаимодействия между различными признаками [31].

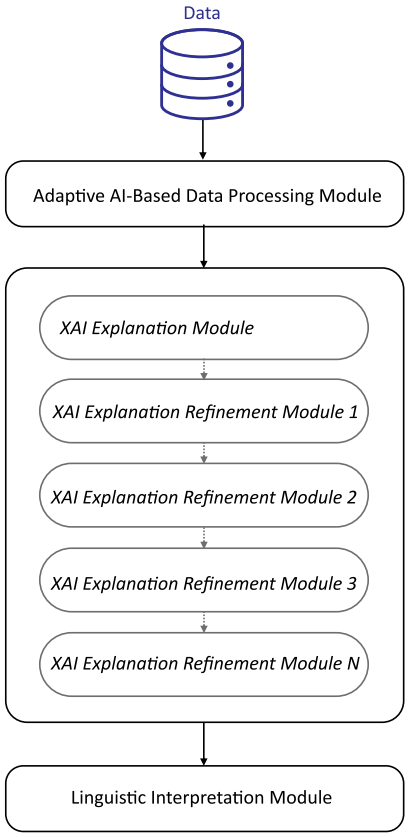
Нелинейные зависимости, которыми пронизаны многие современные модели ИИ, такие как нейронные сети и нечеткие системы, требуют более глубокого анализа для обеспечения их интерпретируемости. В XAI такие зависимости рассматриваются через анализ чувствительности, который позволяет оценить влияние изменений признаков на итоговые предсказания, так как небольшие изменения во входных данных могут значительно повлиять на результат [19]. Анализ чувствительности помогает определить, какие признаки оказывают наибольшее влияние на предсказания модели и используется для улучшения интерпретируемости, особенно в задачах с высокоразмерными данными. Эти принципы тесно связаны с методами визуализации PDP и ICE, рассмотренными ранее, они предоставляют основную помощь для исследования влияния отдельных признаков на результаты предсказаний моделей [29].

При рассмотрении интерпретируемости моделей ИИ и их прогнозов уделяется большое внимание использованию метрических пространств в контексте XAI, поскольку они позволяют измерять схожесть между различными признаками и предсказаниями модели. Метрики пространства используются для оценки того, насколько изменения одного или нескольких признаков влияют на результаты моделей, что критически важно при работе с многомерными данными. Методы визуализации — стратегическая составляющая при рассмотрении любых метрик, опять же PDP и ICE, — позволяют визуализировать влияние каждого признака на предсказания модели при удерживании остальных факторов неизменными. Визуализация взаимодействий между признаками через метрические подходы позволяет исследователям и пользователям лучше понять, какие факторы играют ключевую роль в принятии решений моделью [29].

Иерархическая структура XAI для повышения интерпретируемости моделей

Переходим к следующему этапу рассмотрения объяснимости в ИИ, углубляясь в современные разработки, такие как иерархические структуры объяснительного ИИ (рис. 1). После того, как были рассмотрены базовые математические подходы, важным шагом становится анализ иерархической организации, которая способна повысить качество интерпретируемости моделей, особенно в тех случаях, когда использование одного метода объяснения оказывается недостаточным. В современной практике ИИ, включая такие области, как здравоохранение и экология, где точность решений и доверие к системе критичны, существует необходимость в более комплексных системах интерпретации.

Рис. 1. Иерархическая структура XAI
Fig. 1. Hierarchical Structure of XAI



Иерархическая структура XAI строится на объединении нескольких слоев, каждый из которых выполняет определенную функцию по уточнению и детализации объяснений, генерируемых моделью. В такой системе базовый модуль объяснения инициирует процесс предоставления объяснений, после чего на каждом следующем слое происходит уточнение этих объяснений. Этот процесс можно формализовать через математическую модель, которая учитывает взаимодействие признаков и их весовые коэффициенты в финальном предсказании.

Пусть $f(x)$ — модель, которая возвращает предсказание на основе входных данных x. Чтобы обеспечить интерпретируемость, мы определяем функционал объяснений $E_{(f, x)}$, который возвращает объяснение для данного предсказания. Математически это можно выразить как

$$E_{(f,x)} = \sum_{i}^{n} \left(w_i \times \varphi_{i(x)} \right),$$

где w_i — весовые коэффициенты, характеризующие вклад каждого признака $\varphi_{i(x)}$ в конечное решение. Эти веса могут быть получены с использованием методов, таких как SHAP или LIME.

На первом уровне иерархической структуры производится первичная фильтрация и упрощение объяснений, чтобы сделать их понятными для широкого круга пользователей. Пусть $E_{1(f,x)}$ — результат работы данного уровня, который можно выразить как

$$E_{1(f,x)} = \mathrm{F}(E_{(f,x)}),$$

где F — функция фильтрации и упрощения, которая удаляет из объяснений несущественные или излишне сложные компоненты. На этом этапе задача сводится к выбору таких $\varphi_{i(x)}$, которые наиболее значимы, что можно описать через пороговую функцию:

$$\mathrm{F}\left(E_{(f,x)}\right) = \sum_i^n \left(w_i \times \varphi_{i(x)} \times I\left(w_i > \tau\right)\right),$$

где I — индикаторная функция; τ — пороговое значение значимости признака.

На втором уровне модуль уточнения добавляет контекстные разъяснения, связывая результаты модели с реальными примерами и дополнительными данными. Пусть $E_{2(f,x)}$ — это уточненное объяснение после добавления контекста:

$$E_{2(f,x)} = C\left(E_{1(f,x)}\right), D),$$

где C — функция, добавляющая контекст; D — набор внешних данных или примеров, которые позволяют пользователю лучше понять взаимосвязь между результатом и контекстом применения модели.

Третий уровень уточнения направлен на добавление вероятностных оценок и прогнозирования рисков. Пусть $E_{3(f,x)}$ — объяснение на третьем уровне, включающее оценку рисков:

$$E_{3(f,x)} = P\left(E_{2(f,x)}, R\right),$$

где P — функция прогнозирования, которая использует результат второго уровня и добавляет вероятностные оценки рисков; R — функция, описывающая возможные сценарии риска. Прогнозирование рисков может быть выражено как интеграл вероятности наступления определенного события:

$$P\left(E_{2(f,x)}, R\right) = \int (p(r \mid E_{2(f,x)}))dr),$$

где $p(r|E_{2(f,x)}))$ — условная вероятность наступления риска r при условии текущего состояния $E_{2(f,x)}$.

На последнем этапе структура XAI преобразует уточненные объяснения в лингвистически понятные форматы. Пусть $L(E_{2(f,x)})$ — это интерпретация третьего уровня, адаптированная для конечного пользователя:

$$L(E_{3(f,x)}) = T(E_{3(f,x)}),$$

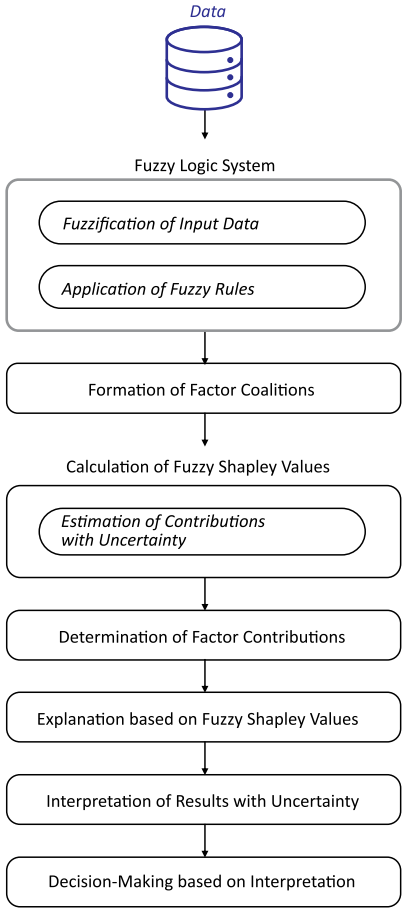
где T — функция лингвистической трансформации, которая формирует текстовое описание на основе математической структуры предыдущих уровней.

Модель XAI, рассмотренная ранее, позволяет улучшать интерпретируемость моделей ИИ через многоуровневое уточнение объяснений. Однако в ряде ситуаций, особенно в сложных системах с высокой степенью неопределенности, таких как гидрогеологические, геоэкологические или медицинские данные, возникает необходимость работать с неопределенными и неполными данными. В этих случаях классические методы XAI могут не обеспечивать достаточной гибкости. Для решения данной проблемы предлагается использовать гипотетическую многоуровневую структуру, состоящую из слоев нечеткой логики и значений Шепли, которые позволяют учитывать неопределенность и более точно интерпретировать вклады различных факторов в принятие решений (рис. 2).

Согласно гипотезе, применение нечеткой логики в сочетании с методами теории игр, такими как значения Шепли, позволяет более точно моделировать влияние неопределенных факторов и коалиций на процесс принятия решений. Это особенно важно в контексте сложных систем, где традиционные методы анализа могут приводить к потере информации о неопределенности.

Рис. 2. Многоуровневая структура объяснений на основе нечеткой логики и значений Шепли

Fig. 2. Multilevel Explanation Structure Based on Fuzzy Logic and Shapley Values



Математическое обоснование

Первым этапом в нечеткой логике является фаззификация входных данных. Пусть у нас есть набор входных переменных $x = \{x_1, x_2, ..., x_n\}$, которые необходимо фаззифицировать. Для каждой переменной вводится нечеткое множество F, которое описывается функцией принадлежности $\mu_{F(x)}$, где $\mu_{F(x)} \in [0,1]$. Эта функция определяет степень принадлежности элемента x к нечеткому множеству.

Следующий этап — это применение нечетких правил. Пусть у нас есть набор правил $R = \{R_1, R_2, ..., R_m\}$, каждое из которых имеет вид «ЕСЛИ условие ТО результат». Эти правила объединяют фаззифицированные входные данные и формируют нечеткие выходные значения.

На третьем этапе происходит формирование коалиций факторов. Пусть факторные коалиции обозначены как $C = \{C_1, C_2, ..., C_p\}$, где каждая коалиция C_i включает определенный набор факторов. Эти коалиции определяются на основе нечетких правил и их взаимосвязей с входными переменными.

Для оценки вклада каждого фактора в результат используется метод Шепли. Пусть у нас есть множество факторов $F=\{F_1, F_2, ..., F_n\}$, каждый из которых может участвовать в коалиции. Нечеткое значение Шепли для каждого фактора F_i вычисляется как среднее значение его вклада в различные коалиции. Математически это выражается как

$$\Phi_i = \sum_{S \subseteq F \setminus \{i\}} \left[\frac{|S|! \times (|F| - |S| - 1)!}{|F|!} \times \left(v(S \cup \{i\}) - v(S) \right) \right],$$

где $v(S)$ — это функция выигрыша для коалиции S; $|S|$ — число участников в коалиции S. Нечеткость в этом подходе возникает из-за неопределенности в оценке функции выигрыша для различных коалиций.

На основе рассчитанных нечетких значений Шепли можно определить вклад каждого фактора в общий результат модели. Эти значения дают оценку того, какой вклад вносит каждый фактор с учетом неопределенности в данных и правилах. После определения вкладов факторов производится интерпретация результатов с учетом неопределенности. Эта интерпретация может включать в себя оценку степени уверенности в полученных выводах и возможные сценарии при различных комбинациях факторов.

Заключительный экспертный этап заключается в принятии решений на основе интерпретации результатов с учетом нечетких значений Шепли и неопределенности. Эти решения могут включать выбор стратегий управления, оценки рисков и другие действия, основанные на выводах модели.

От концепций XAI к практическому применению в медицине

Изучение принципов XAI позволило глубже понять механизмы интерпретируемости сложных моделей. Научные подходы XAI обеспечивают прозрачность решений, что необходимо для систем, где важны как высокая точность, так и возможность объяснить процесс принятия решений. В здравоохранении XAI помогает анализировать медицинские данные, учитывая неопределенности и влияние множества факторов, что улучшает диагностику и профилактику заболеваний. В территориальном районировании XAI интегрируется с геоинформационными системами (ГИС), чтобы управлять природными и социально-экономическими рисками, делая решения более понятными и обоснованными.

Новые подходы к построению иерархических структур XAI с включением элементов нечеткой логики открывают возможности для работы с такими сложными системами. Иерархический XAI с нечеткостями позволяет более гибко интерпретировать данные в условиях неопределенности, что особенно полезно для анализа сложных многокомпонентных систем, таких как ГИС. Территориальное районирование, включающее разнообразные природные, социальные и экономические факторы, требует именно таких решений, где неопределенности и нечеткости данных могут быть учтены и корректно интерпретированы. Точно такая же ситуация и в здравоохранении, где принятие решений напрямую связано с человеческим здоровьем: инновационные методы позволяют учитывать влияние множества факторов с различной степенью неопределенности, обеспечивая более достоверные и интерпретируемые прогнозы. Применение иерархической структуры XAI, дополненной методами нечеткой логики, может существенно повысить точность диагностики и профилактики, а также улучшить качество и обоснованность принимаемых решений.

В следующих разделах будут рассмотрены конкретные аспекты применения XAI в территориальном районировании и здравоохранении с особым вниманием к разработке интерпретируемых моделей для принятия устойчивых решений и обеспечения общественного здоровья.

Применение XAI в ГИС при территориальном районировании

Территориальное районирование является одним из ключевых инструментов управления земельными ресурсами, охраны окружающей среды и стратегического планирования. Оно направлено на классификацию и разделение территории на зоны с учетом физических, социально-экономических и экологических характеристик для оптимизации использования земли. В последние десятилетия подходы к районированию значительно эволюционировали, особенно благодаря внедрению ГИС и методов дистанционного зондирования, которые позволяют собирать и анализировать большие объемы данных о территории. ГИС прочно

закрепились в современном территориальном анализе благодаря своим возможностям по сбору, хранению, анализу и визуализации пространственных данных. Они находят широкое применение в разнообразных сферах: от городского планирования до экологического мониторинга. Однако, несмотря на широкое использование ГИС, традиционные методы территориального районирования нередко сталкиваются с ограничениями, связанными с недостаточной интерпретируемостью получаемых моделей и предсказаний. Появление XAI открыло новые возможности для преодоления этих ограничений, обеспечивая прозрачность и интерпретируемость моделей машинного обучения. Хотя традиционные модели часто достигают высокой точности, процесс принятия решений остается непрозрачным, что усложняет их использование в областях, где критически важна объяснимость, например, в управлении окружающей средой или государственной политике.

XAI расширяет возможности систем, основанных на ГИС, заполняя пробел между сложными результатами моделирования и процессом принятия решений в территориальном зонировании. В отличие от традиционных моделей «черного ящика», которых в система ГИС достаточно, XAI фокусируется на создании прозрачных моделей, которые могут объяснить свои решения. Это особенно важно для пространственного анализа, где на результаты влияют многочисленные факторы, такие как физико-географические условия, инфраструктурные ограничения и социально-экономические параметры. Применение XAI в географических исследованиях и ГИС повышает точность прогнозов и обеспечивает возможность детального анализа вклада каждого фактора в формирование конечных результатов. Сегодня XAI активно используется в территориальном районировании, особенно в контексте устойчивого развития и охраны здоровья населения. Основная цель применения XAI в территориальном планировании заключается в повышении прозрачности решений, связанных с использованием и управлением природными ресурсами.

Эффективное районирование требует учета многомерных факторов, таких как экологические, социальные и экономические параметры. Интеграция XAI и ГИС создает синергетический эффект, когда модели, используемые для территориального районирования, одновременно предоставляют точные прогнозы и объясняют, как различные факторы, например изменения землепользования, демографические изменения или экологические параметры, влияют на результаты. Это особенно важно в ситуациях, когда необходимо принимать решения, затрагивающие вопросы устойчивого развития и управления природными ресурсами.

Одним из примеров успешной интеграции XAI и ГИС является исследование [32], в котором оценивается предсказательная эффективность адаптивной нейронечеткой системы вывода (ANFIS) с использованием шести различных функций принадлежности (MF) для картирования восприимчивости к оседанию земли (LSSM) в Маранде, северо-запад Ирана. Район подвержен засухам, низкому уровню грунтовых вод и, как следствие, повреждениям от оседания земли. Создана база данных об оседаниях земли на основе полевых обследований. Данные были разделены на три равных набора, где один использовался для тестирования и валидации, а два других — для обучения. В качестве факторов оседания использовались гидрологические и топографические параметры, представленные в виде слоев GIS.

Способы интеграции XAI и ГИС

Рассмотренное выше исследование показало значительные преимущества интеграции методов XAI с ГИС, что позволило улучшить точность и интерпретируемость моделей для территориального районирования и мониторинга. Однако для полноценного исследования возможностей XAI в геоинформационных системах необходимо более глубоко понять различные способы интеграции этих технологий, особенно в контексте многомерных задач. Следующий раздел сфокусирован на более подробном рассмотрении существующих методов интеграции XAI с ГИС, подчеркивая их преимущества и ограничения, а также возможные области их дальнейшего применения. Это особенно актуально в контексте устойчивого развития и пространственного анализа, где правильная интерпретация моделей может способствовать более эффективному управлению природными ресурсами и улучшению качества жизни населения.

Данные из табл. 1 демонстрируют широкий спектр подходов XAI, которые могут быть использованы для объяснения сложных моделей машинного обучения в пространственных задачах. Каждый метод подходит для определенных целей: от простого анализа закономерностей (методы, основанные на правилах и деревьях решений) до сложных интерпретаций в многомерных моделях (SHAP, анализ спектральной релевантности, нейронная проводимость). Эти методы позволяют как локально, так и глобально интерпретировать влияние факторов на прогнозы, что особенно важно для территориального районирования. Таким образом, выбор подходящего подхода XAI зависит от задачи, характеристик данных и уровня детализации, необходимого для анализа. Табл. 1 была переведена и доработана на основе материалов работы Xing и Sieber (2023) [48], чтобы адаптировать ее к задачам пространственно-явных моделей и территориального анализа.

Одним из ключевых аспектов успешной интеграции методов объяснительного XAI с ГИС является выбор подходящего способа интеграции.

Табл. 1. Наиболее популярные алгоритмы и подходы для XAI и их применение в пространственно-явных моделях
Tab. 1. The most popular algorithms and approaches for XAI and their application in spatially explicit models

Методы объяснимого ИИ	Описание	Пример алгоритма, используемого для GeoAI	Источник
Объяснения, основанные на правилах	Использование знаний в предметной области для оценки разницы между данным советом и новой ситуацией	Линейная регрессия	[33]
Древовидные объяснения	Использование деревьев решений для объяснения того, как различные входные данные могут быть использованы для получения соответствующих результатов	Деревья классификации и регрессии	[34]
Объяснения SHAPLEY, основанные на ценностях	Используя значения SHAPLEY, основанные на теории игр, можно определить вклад каждого входного признака в достижение заданного результата	SHAP	[35–38]
Объяснения, основанные на градиенте	Измерение изменений от исходного значения к входному вдоль траектории, например, сравнение слоя с базовым слоем или начальной точкой траектории	Интегрированный градиент	[39]
Объяснения на основе уровней	Атрибуция каждого нейронного слоя рассчитывается исходя из их релевантности, что измеряет их вклад в получение результатов	Поэтапное распространение релевантности (LRP)	[40)
Местные суррогатные модели	Замена функции принятия решений с использованием локальных суррогатных моделей для объяснения процессов принятия решений в DNN	Локальная интерпретируемая модель — независимое объяснение (LIME)	[41–43]
Визуализация и аудиозапись с указанием авторства	Визуальное или звуковое представление атрибуции входных переменных для данного результата, сгенерированного DNN	Карты значимости	[44])
Объяснения на основе графиков	Построение различных графических моделей (включая подграфы) для увязки результатов искусственного интеллекта с заданными входными данными/особенности	Скрытое представление на основе графиков	[45]
Анализ спектральной релевантности	Использование спектральной кластеризации для объединения различных локальных объяснений (например, карт характерных черт) для получения заданных результатов	Анализ спектральной релевантности	[45]
Объяснения, основанные на нейронах	Разложение градиентов по слоям DNN для вычисления вклада нейронов скрытого слоя в получение заданных результатов	Нейронная проводимость	[46]
Противоборствующие объяснения	Объяснение того, как данная модель искусственного интеллекта может быть вынуждена выдавать неверные результаты	Контрафактические примеры	[47]

Этот выбор определяется разнообразием методов XAI и характеристиками пространственных данных, используемых в ГИС. Для лучшего понимания существующих возможностей и их адаптации к конкретным задачам методы интеграции можно классифицировать по ряду важных критериев. В табл. 2 приведена классификация способов интеграции. Эта структура поможет систематизировать существующие подходы и определить наиболее подходящие стратегии для решения задач пространственного анализа и управления землепользованием.

Классификация способов интеграции XAI и ГИС позволяет выделить ключевые направления развития и применения этих технологий в области пространственного анализа и территориального районирования. Применение глобальной и локаль-

Табл. 2. Способы интеграции методов XAI и ГИС
Tab. 2. Methods of Integrating XAI and GIS

Критерий	Тип	Описание
По уровню интерпретируемости	Глобальная интерпретируемость	Фокусируется на объяснении общих закономерностей на глобальном уровне. Пример: SHAP
	Локальная интерпретируемость	Объясняет решения для конкретных участков территории. Пример: LIME
По типу используемых моделей	Модельно-зависимая интеграция	Требует доступа к структуре модели для глубокого анализа. Пример: визуализация активаций в нейронных сетях
	Модельно-независимая интеграция	Работает независимо от структуры модели. Пример: SHAP и LIME
По типу данных	Интеграция с экологическими данными	Анализ экологических факторов, таких как загрязнение воздуха, воды и лесной покров
	Интеграция с социально-экономическими данными	Анализ демографических и инфраструктурных данных для планирования
По цели применения	Пространственный анализ риска	Использование XAI для объяснения рисков, связанных с экологией, изменением климата и распространением болезней
	Оптимизация землепользования	Объяснение того, как географические и экологические факторы влияют на землепользование
По используемой архитектуре XAI	Иерархические структуры XAI	Многоуровневая интерпретация моделей, начиная с вклада признаков и до их детализации
	Гибридные методы XAI	Комбинация методов интерпретации для улучшения анализа сложных пространственных моделей
По степени автоматизации	Полуавтоматические интеграции	Частичное участие экспертов для верификации данных
	Полностью автоматизированные интеграции	Полная автоматизация процесса интерпретации для масштабного планирования

ной интерпретируемости, работа с различными типами моделей и данных, а также использование гибридных и иерархических структур XAI открывают новые возможности для создания более точных и прозрачных прогнозов. Учитывая степень автоматизации и возможности комбинирования методов, можно значительно повысить эффективность решений в области управления природными ресурсами, планирования землепользования и охраны окружающей среды. После представленной классификации способов интеграции XAI и ГИС стоит обратить особое внимание на два метода, которые получили наибольшее распространение в задачах пространственного анализа

Одним из ключевых методов XAI, применяемых в задачах пространственного анализа, является SHAP. В контексте территориального планирования и районирования SHAP позволяет детально объяснить, как географические, экологические и социально-экономические факторы влияют на результаты модели. Это может включать такие аспекты, как плотность растительного покрова, топография, близость к водоемам или транспортной инфраструктуре. Интеграция SHAP с ГИС обеспечивает пространственную визуализацию этих факторов, что позволяет принимать решения, основанные на глубоких данных о взаимодействии территориальных особенностей и прогнозируемых результатов. Примером успешного применения SHAP в ГИС является исследование, проведенное в столичном регионе Сеула (SMA), где модель XGBoost использовалась для прогнозирования расширения городов. В ходе исследования SHAP применялся для объяснения того, как такие факторы, как растительный покров, рельеф и расстояние до водоемов, влияли на прогнозы о развитии городских территорий. Использование SHAP дало возможность количественно оценить вклад каждого фактора в модель, что позволило градостроителям принимать более взвешенные решения. Этот подход оказался важным для разработки стратегий устойчивого развития, поскольку позволил сбалансировать интересы городского роста с необходимостью сохранения природных ресурсов [49]. Метод SHAP не только обеспечивает глобальные интерпретации — общие закономерности, влияющие на все географические территории, — но и предоставляет локальные интерпретации, которые фокусируются на конкретных районах. Например, SHAP помогает объяснить, почему в одном районе предпочтительны определенные виды землепользования, а в другом — нет. Это особенно важно для политиков и градостроителей, которым необходимо учитывать местные особенности и принимать решения на микроуровне, поддерживая долгосрочное развитие территории.

Другим важным методом XAI, применяемым в пространственном анализе, является LIME. В отличие от SHAP, который предоставляет глобальные интерпретации, LIME фокусируется на объяснении отдельных прогнозов путем создания интерпретируемых моделей на локальном уровне. Это особенно важно в тех случаях, когда необходимо детализировать и понять прогнозы для конкретных участков территорий. Например, если модель машинного обучения предсказывает, что определенная зона подходит для промышленного использования, LIME может объяснить, какие именно факторы — близость к транспортной инфраструктуре или наличие природных ресурсов, повлияли на такое решение. В исследовании, проведенном в холмистых районах Южной Кореи, LIME использовался для объяснения прогнозов модели относительно пригодности земель для городского развития. Модель оценивала такие факторы, как уклон территории, плотность растительного покрова и гидрологические особенности, и давала прогноз о пригодности земель для городского строительства [50]. Применение LIME позволило детализировать, какие из этих факторов оказывали наибольшее влияние на прогноз для каждого участка территории, что дало возможность градостроителям корректировать планы в соответствии с реальными условиями и требованиями экологической устойчивости. В этом контексте важно также учесть, что подходы XAI, такие как SHAP и LIME, адаптируемы для различных применений в ГИС, расширяя их потенциал для решения задач, связанных с пространственным планированием, управлением природными ресурсами и экологической оценкой.

Важнейшая особенность интеграции методов XAI с ГИС заключается в том, что XAI не только объясняет, как именно модель пришла к тем или иным результатам, но и позволяет визуализировать пространственное распределение значимых факторов. Это значительно облегчает принятие решений, связанных с землепользованием, городским планированием и природопользованием. Применение методов XAI в ГИС также способствует повышению доверия к моделям, поскольку результаты могут быть объяснены на понятном для пользователей уровне с учетом географических особенностей территорий.

Потенциал мультикаскадной иерархии XAI на примере ГИС

Следующий этап в развитии XAI — это внедрение *мультикаскадной иерархии XAI*, которая позволяет структурировать интерпретации на различных уровнях сложности, начиная от базовых факторов и заканчивая комплексными многомерными взаимосвязями. Такой подход особенно эффективен в применении к ГИС, где пространственная изменчивость и взаимозависимость факторов требуют более глубокой и точной интерпретации. В этом разделе будет подробно рассмотрен потенциал данной иерархической структуры на примере ее применения в ГИС для пространственного планирования и анализа территорий. Возможность объяснить, почему модель предполагает определенные правила районирования, способствует подотчетности, что является ключевым аспектом общественного доверия к городскому и экологическому планированию. Методы XAI также позволяют улучшить коммуникацию с заинтересованными сторонами. В решениях по городскому планированию часто участвуют многочисленные заинтересованные стороны, включая управленцев, градостроителей, экологические агентства и широкую общественность. Когда модели районирования можно объяснить прозрачно, становится легче вовлекать заинтересованные стороны в содержательные дискуссии о землепользовании, распределении ресурсов и устойчивости. Это способствует более демократичному процессу принятия решений, в котором все стороны могут понять причины, лежащие в основе предложений по районированию и планированию, и внести свой вклад в достижение более сбалансированных результатов.

Мультикаскадная иерархия XAI обеспечивает более глубокое понимание процессов принятия решений, особенно в условиях неопределенности и многокритериальности. Рассмотрение интеграции модулей объяснимости в сложные системы, такие как ГИС, демонстрирует значительные преимущества. ГИС позволяют реализовывать пространственные решения на основе анализа множества факторов. Однако современные ГИС-продукты, как правило, не включают модули объяснимости на горизонтальных уровнях, только для входных данных, что затрудняет понимание логики системы. Интеграция таких модулей в ГИС обеспечит экспертам возможность видеть всю иерархию решений, принимаемых системой, что особенно важно для прикладных областей, связанных с экологией, здравоохранением и территориальным планированием.

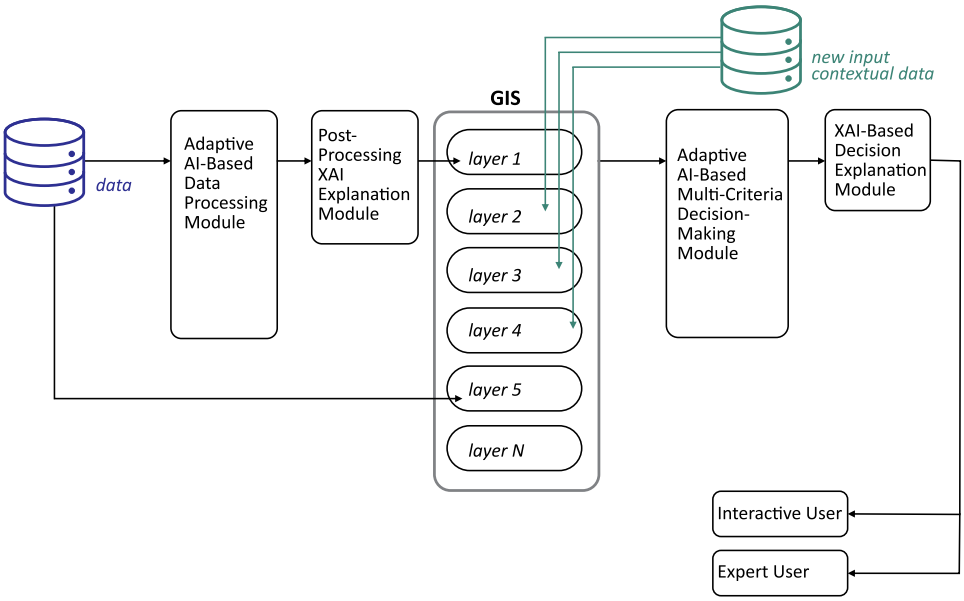
На представленной схеме (рис. 3) изображена архитектура такой мультикаскадной иерархической системы XAI, где на каждом уровне происходит взаимодействие между ИИ и ГИС для обеспечения объяснимости на каждом этапе обработки данных и принятия решений. Это позволяет получить итоговое решение и понять, какие данные и факторы на каждом этапе повлияли на конечный результат.

Теперь, переходя к математической модели, рассмотрим, как каждый из этих модулей взаимодействует с входными данными и как обеспечивается объяснимость на каждом уровне системы.

Математическое описание мультикаскадной иерархической системы XAI для многокритериального принятия решений

Мультикаскадная иерархическая система XAI представляет собой интегрированный подход к интерпретации данных и принятию решений в

Рис. 3. Схема использования мультикаскадной иерархии XAI в ГИС
Fig. 3. Scheme of Use of XAI Multicascade Hierarchy in GIS



сложных системах. В этой системе обработка данных происходит через несколько этапов, начиная с адаптивного модуля ИИ, который обрабатывает входные данные, и заканчивая многокритериальным модулем принятия решений с последующей интерпретацией результатов.

На первоначальном этапе входные данные поступают в адаптивный ИИ-модуль, который выполняет первичную обработку. Модель ИИ анализирует данные, извлекая ключевые признаки и готовя их для последующих этапов. Рассмотрим подробнее: пусть X — исходные данные, а результат обработки на первом этапе обозначим как $f(X)$.

После первичной обработки данных система использует модуль объяснения XAI. Этот модуль интерпретирует результаты работы ИИ-модели, предоставляя объяснения по полученным данным. Пусть $E(f(X))$ — функция, которая объясняет результаты работы адаптивной ИИ-модели. Она делает результаты интерпретируемыми для последующих каскадов, включая ГИС.

Обработанные данные вместе с новыми контекстуальными данными, поступающими от ГИС, передаются в несколько слоев анализа. Эти слои ГИС обрабатывают пространственную информацию, которая может дополняться новыми данными, поступающими извне (например, изменения в картографических или геоэкологических данных). Модуль XAI может применяться как до, так и после анализа этих данных в зависимости от их характера. Если объяснения требуются на уровне каждого слоя, то можно использовать локальные XAI модули:

$$L_n = E(GIS_n(X)),$$

где GIS_n — слой ГИС; L_n — объяснение для данного слоя n.

После обработки всех слоев ГИС результат передается в адаптивный ИИ-модуль многокритериального принятия решений. Этот модуль агрегирует информацию из всех слоев ГИС и учитывает контекстуальные данные для принятия решений. Пусть функция принятия решения описывается как

$$D = M(f(X), GIS_1(X), GIS_2(X), \ldots, GIS_n(X)),$$

где M — функция многокритериального принятия решений, основанная на данных, обработанных ИИ и ГИС.

После принятия решения модуль XAI снова используется для объяснения конечного решения, предоставленного многокритериальной системой. Этот модуль предназначен для того, чтобы сделать решения понятными как для экспертов, так и для пользователей, работающих с системой. Пусть итоговое объяснение будет $E(D)$, где D — итоговое решение, принятое системой.

Финальные объяснения передаются экспертам и пользователям, которые могут взаимодействовать с системой и принимать обоснованные решения на основе предоставленных данных и выводов системы.

Математическая модель

1. Первичная обработка данных:

$$f(X) = f(\{x_1, x_2, \ldots, x_n\}),$$

где $X = \{x_1, x_2, \ldots, x_n\}$ — входные данные; $f(X)$ — результат первичной обработки ИИ.

2. Объяснение первичной обработки:

$$E\left(f\left(X\right)\right) = \sum_{i=1}^{n} w_i \times \varphi_i\left(f\left(X\right)\right),$$

где w_i — весовые коэффициенты, характеризующие вклад каждого признака $\varphi_i(f(X))$ в финальный результат первичной обработки.

3. Обработка данных на уровнях ГИС:

$$L_n = E\left(GIS_n\left(X\right)\right) = \sum_{i=1}^{k} w_i^n \times \varphi_i^n\left(GIS_n\left(X\right)\right),$$

где φ_i^n — функция пространственного анализа на уровне n.

4. Объяснение на каждом уровне:

$$E_n = \sum_{i=1}^{k} w_i^n \cdot \varphi_i^n\left(GIS_n\left(X\right)\right),$$

где w_i^n — весовой коэффициент на уровне n; $\varphi_i^n(GIS_n(X))$ — функции анализа для каждого слоя GIS на уровне n.

5. Итоговое многокритериальное решение:

$$D = M(f(X), GIS_1(X), GIS_2(X), \ldots, GIS_n(X)),$$

где D — результат многоуровневого анализа, основанный на функции M, которая агрегирует данные исходной модели $f(X)$ и различных слоев $GIS_n(X)$ пространственных данных.

6. Объяснение итогового решения:

$$E\left(D\right) = \sum_{j=1}^{m} v_j \cdot \psi_j\left(D\right),$$

где v_j — весовые коэффициенты; $\psi_j(D)$ — функции, определяющие вклад на уровне j в результат $E(D)$.

Важно понимать, как эта система обеспечивает поддержку принятия решений для различных категорий пользователей. После описания математической модели мультикаскадной иерархии XAI мы можем более наглядно представить, как это взаимодействие структурируется на практике. Рис. 4 представляет взаимодействие между иерархической системой XAI принятия решений и двумя основными категориями пользователей (интерактивным пользователем и экспертом).

• Интерактивный пользователь взаимодействует с системой, просматривая данные и получая объяснения результатов, что помогает ему лучше понять логику принятия решений системы XAI.

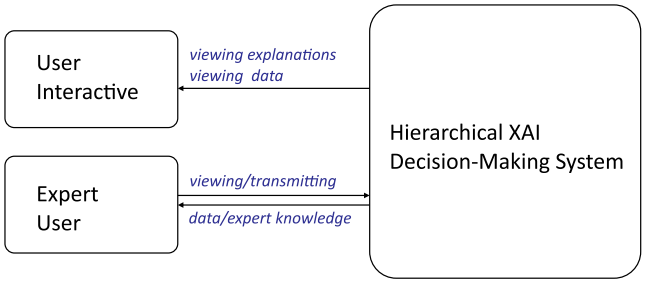
• Эксперт передает экспертные знания и данные в систему, а также имеет возможность просматривать выводы системы для валидации или интерпретации результатов. Экспертный вклад особенно важен на этапе настройки и адаптации системы для конкретных областей применения.

Рассмотрим этот процесс с помощью математических представлений, которые позволят глубже понять, как система XAI структурирует доступ и взаимодействие пользователей с различными слоями данных и объяснений.

Иерархическая система XAI принимает входные данные X, которые могут включать как структурированные, так и неструктурированные данные. Модель $f(X)$ применяет алгоритмы ИИ для анализа данных и генерации решений. Пусть R — это результат работы системы:

Рис. 4. Схема реализации доступа в проектируемую ГИС с модулями XAI

Fig. 4. Scheme of Access Implementation in the Designed GIS with XAI Modules



$$R = f(X),$$

где R — это вывод или предсказание системы, основанное на анализе входных данных X.

После того, как система приняла решение, включается модуль объяснения, который интерпретирует результат.

Пусть $E(f(X))$ — функция, которая объясняет результаты системы:

$$E\left(f\left(X\right)\right) = \sum_{i=1}^{n} w_i \cdot \phi_i\left(X\right),$$

где w_i — это весовые коэффициенты, отражающие вклад каждого признака $\phi_i(X)$ в итоговое решение.

Интерактивный пользователь получает доступ к данным и результатам объяснений системы. Пусть U_I обозначает взаимодействие интерактивного пользователя с системой. Тогда информация, доступная пользователю, может быть описана как функция данных и объяснений:

$$U_I = (X, E(f(X))).$$

Интерактивный пользователь может использовать полученные объяснения для более глубокого понимания решения, принятого системой.

Эксперт не только получает результаты и объяснения, но также передает свои данные или знания для корректировки или улучшения системы. Пусть U_E обозначает взаимодействие эксперта с системой. Экспертный вклад может быть описан как

$$U_E = (X_E, E(f(X)) + f(X_E)),$$

где X_E, E — дополнительные данные, введенные экспертом для уточнения работы системы.

Если эксперт предоставляет новые данные или уточнения, система может пересчитать решение с учетом экспертного вклада. Это можно описать как обновленную функцию:

$$R' = f(X + X_E),$$

где X_E — это данные или знания, введенные экспертом. Обновленное решение R' также может быть

объяснено модулем XAI, что приводит к новой интерпретации для интерактивного пользователя и эксперта:

$$E\left(f\left(X+X_E\right)\right)=\sum_{i=1}^{n+m} w_i^{'}\cdot\varphi_i\left(X+X_E\right),$$

где $w_i^{'}$ — новые весовые коэффициенты, учитывающие экспертные данные.

Эта математическая модель описывает основные этапы взаимодействия между пользователями и иерархической системой XAI. Интерактивные пользователи получают доступ к данным и объяснениям, что повышает доверие к системе, тогда как эксперты могут передавать свои знания и данные, влияя на процесс принятия решений и обеспечивая адаптацию системы под конкретные задачи.

Будущие направления развития интегрированных систем XAI и ГИС в задачах районирования и достижениях целей ЦУР

По мере того, как области ГИС и XAI продолжают развиваться, появляются новые возможности для их применения в территориальном районировании. Одной из перспективных областей является интеграция потоков данных в реальном времени с ГИС-моделями с поддержкой XAI. Включая динамические данные, такие как схемы дорожного движения, уровень загрязнения или экономическая активность, в модели районирования и планирования, градостроители могут принимать более адаптивные решения, отражающие текущие условия, а не полагаться исключительно на исторические данные. Кроме того, достижения в области моделей глубокого обучения в сочетании с методами XAI обещают повысить точность и прозрачность прогнозов районирования. К потенциальным направлениям для исследований можно отнести многоагентные системы (MAS) в сочетании с XAI для приложений районирования. MAS может моделировать поведение и взаимодействие различных агентов (например, граждан, предприятий и государственных учреждений) в заданном пространстве. В сочетании с XAI эти системы могут предоставить прозрачные объяснения того, как различные параметры могут повлиять на различные заинтересованные стороны, помогая планировщикам сбалансировать конкурирующие интересы и создать более справедливую городскую среду.

Интеграция XAI и ГИС представляет собой смену парадигмы в области территориального планирования и районирования, предлагая как повышенную точность прогнозов моделей, так и повышенную прозрачность для лиц, принимающих решения, облегчая государственным менеджерам и заинтересованным сторонам совместную работу над устойчивыми, информированными стратегиями землепользования. По мере дальнейшего развития этих технологий их применение в территориальном районировании, несомненно, получит еще большее распространение, открывая новые возможности для решения сложных социально-экономических и экологических проблем, связанных с ростом городов.

Перспективы внедрения методов XAI в отечественные геоинформационные системы

Современные геоинформационные системы активно развиваются, предоставляя широкие возможности для пространственного анализа, прогнозирования и визуализации данных. Однако с увеличением сложности моделей и объема данных возникает проблема объяснимости результатов. Методы объяснимого искусственного интеллекта могут значительно повысить прозрачность работы аналитических алгоритмов, что особенно важно для задач, связанных с управлением природными ресурсами, территориальным планированием и экологическим мониторингом.

Одним из удачных примеров отечественных разработок является программно-технологический комплекс ГИС INTEGRO, созданный ФГБУ «ВНИГНИ» в рамках инициатив по импортозамещению [51]. Эта система уже сегодня предоставляет мощный функционал для пространственного анализа, включая построение цифровых карт, аналитическую обработку данных и создание трехмерных моделей [52]. Однако ее многомодульная архитектура позволяет рассматривать возможность дальнейшего расширения функционала за счет интеграции модулей объяснимости, основанных на принципах XAI.

Потенциальные направления для внедрения XAI в ГИС INTEGRO включают:

1. Локальную интерпретируемость (например, LIME) для объяснения прогнозов в конкретных зонах территориального анализа.

2. Глобальную интерпретируемость (SHAP) для анализа вклада различных факторов в результаты моделей на уровне всего региона.

3. Адаптивную нейронечеткую аппроксимацию (ANFIS) как новый модуль интерполяции данных, позволяющий учитывать неопределенности и множественные критерии.

Методы нечеткой логики, такие как использование функций принадлежности и дефазификация, позволяют учитывать высокую степень неопределенности данных, которая часто возникает при работе с пространственными данными. Например, между точками наблюдений (скважинами) существуют зоны, где данных нет, но их можно моделировать, используя законы логики и анализ данных. Применение ANFIS в ГИС INTEGRO поможет:

• Учитывать нелинейные связи и пространственную вариативность параметров.

• Строить многовариантные модели, которые будут учитывать различные сценарии, такие как высокий или средний риск размещения объектов.

• Создавать карты пространственного распределения достоверности, которые помогут анализировать зоны с высоким уровнем неопределенности.

Преимущества интеграции XAI и нечеткого моделирования:

Включение таких модулей в функционал ГИС INTEGRO позволит:

• Повысить прозрачность аналитических процессов и доверие к моделям со стороны пользователей.

• Обеспечить возможность анализа вклада различных факторов в прогнозы, что особенно важно для стратегического планирования и управления природными ресурсами.

• Расширить применимость системы для задач экологического моделирования, таких как прогнозирование экологических рисков и выбор территорий для размещения полигонов твердых коммунальных отходов (ТКО).

Примером практического применения XAI в отечественных ГИС может стать объяснение влияния геофизических и геологических параметров на прогнозы о пригодности территорий для землепользования или определения зон экологического риска. Такие модули позволят пользователям видеть, как конкретные параметры (например, данные о грунтовых водах, растительности или инфраструктуре) влияют на результаты анализа. Это важно для задач устойчивого развития и управления территорией. Внедрение модулей XAI в ГИС INTEGRO, разработанную как отечественная альтернатива зарубежным системам, позволит повысить доверие пользователей к результатам аналитических прогнозов и откроет новые возможности для ее использования в стратегическом планировании и управлении природными ресурсами. Подобный шаг соответствует мировым трендам и продемонстрирует конкурентоспособность российской геоинформационной отрасли на международном уровне.

Медицинская профилактика через устойчивое территориальное планирование и районирование

Медицинская профилактика играет центральную роль в достижении целей устойчивого развития. Эти цели охватывают многопараметрические аспекты экологии, экономики и социологии. В таких условиях ключевой задачей становится обеспечение здоровья и благополучия всех возрастных групп. Для этого требуется радикальное переосмысление существующих подходов к здравоохранению, особенно в условиях глобальных экологических изменений и демографического роста. Профилактика заболеваний и обеспечение долгосрочной устойчивости систем здравоохранения становятся важными элементами в этом процессе, что способствует достижению целей устойчивого развития, включая улучшение здоровья населения и снижение нагрузки на природные ресурсы. Одной из центральных задач устойчивого развития является поддержание баланса между экономическими интересами, социальным благополучием и экологической безопасностью. В этом контексте здравоохранение приобретает особую значимость, поскольку напрямую влияет как на качество жизни населения, так и на способность общества адаптироваться к изменениям окружающей среды. Превентивные меры, направленные на предотвращение заболеваний, могут существенно снизить нагрузку на системы здравоохранения, затраты на лечение, а также защитить природные ресурсы. В этой связи стратегическое планирование территорий с учетом профилактических мер становится важным инструментом достижения целей устойчивого развития. Использование таких инструментов, как XAI и ГИС, позволяет улучшить точность анализа и выявить критические зоны, требующие особого внимания.

Включение экологических факторов в процесс планирования управлением территории позволяет снизить нагрузку на системы здравоохранения и предотвратить рост числа заболеваний, связанных с экологическим загрязнением. Таким образом, применение XAI и ГИС становится важным инструментом для реализации стратегий профилактики заболеваний на основе данных об окружающей среде.

Ранняя диагностика как ключевой элемент медицинской профилактики на основе XAI

В рамках устойчивого развития медицинская профилактика становится ключевым элементом для обеспечения долгосрочного благополучия населения. Это обусловлено необходимостью создания эффективных систем здравоохранения, которые способны справляться с нарастающей нагрузкой, вызванной ухудшением экологической ситуации, урбанизацией и старением населения. Одним из основополагающих принципов устойчивого здравоохранения является ранняя диагностика, направленная на предотвращение заболеваний и минимизацию их последствий. В данном контексте XAI играет важную роль, обеспечивая прозрачность и обоснованность решений, что способствует снижению нагрузки на системы здравоохранения и экономику в целом. Несмотря на сравнительно молодой возраст области XAI, ее потенциал в здравоохранении уже продемонстрировал свои огромные возможности. За последние годы наблюдается значительный рост интереса к использованию XAI для медицинской диагностики, особенно в контексте раннего выявления заболеваний. Необходимость в прозрачных решениях обусловлена их способностью предоставлять высокоточные прогнозы и объ-

яснять врачам, каким образом были получены эти результаты. Это, в свою очередь, повышает доверие к системам ИИ.

Ранняя диагностика имеет стратегическое значение в профилактике заболеваний, особенно тех, которые могут прогрессировать и привести к тяжелым последствиям, таких как сердечно-сосудистые, онкологические и диабет. Например, при анализе ОКТ-снимков модуль внимания в сверточных нейронных сетях позволяет улучшить распознавание патологий сетчатки глаза, таких как диабетический макулярный отек и возрастная макулярная дегенерация, что ускоряет диагностику и снижает вероятность ошибок [53]. Важной особенностью применения XAI в этой области является возможность объяснять врачам, какие именно данные повлияли на решение модели, что значительно повышает доверие к результатам и способствует интеграции таких технологий в практическую медицину. Помимо использования CNN, особое внимание уделяется гибридным методам, которые совмещают преимущества классических методов компьютерного зрения с искусственными нейронными сетями и нечеткой логикой. Примером служит разработка гибридной системы для диагностики диабетической ретинопатии, сочетающей интерпретируемость XAI с гибкостью и адаптацией к различным стадиям заболевания [54]. Такой подход позволяет врачам получать не только диагнозы, но и информацию о данных, на основании которых они были сделаны, что значительно повышает прозрачность и надежность системы. Методы XAI, такие как Grad-CAM и LIME, помогают преодолевать проблему интерпретируемости современных систем глубокого обучения. Эти инструменты позволяют визуализировать значимые области медицинских изображений, облегчая восприятие результатов медицинскими специалистами и усиливая доверие к ИИ [55]. Кроме того, нейронечеткие сети (НС) представляют собой перспективный инструмент для классификации медицинских изображений, особенно в условиях неопределенности данных. Например, в работе А.Н. Аверкина рассматривается применение НС для анализа офтальмологических изображений, где высокая точность диагностики может зависеть от минимальных отклонений в данных [56]. Эти системы позволяют сочетать мощные обучающие возможности нейронных сетей с гибкостью нечеткой логики, создавая тем самым более адаптируемые и интерпретируемые решения.

Кроме того, активно развиваются системы поддержки принятия решений (СППР), которые применяют методы XAI и становятся основой для персонализированной медицины и концепции здравоохранения 5.0. Примером является разработка модульной системы, которая позволяет интегрировать XAI в СППР, сохраняя высокий уровень доверия со стороны врачей и обеспечивая доступность диагностики на ранних стадиях [57]. Эти решения открывают новые возможности для повышения качества медицинских услуг и эффективной профилактики заболеваний.

Таким образом, применение XAI в медицинской профилактике и диагностике открывает новые горизонты для улучшения качества здравоохранения. Технологии объяснительного ИИ, такие как XAI, позволяют сочетать высокую точность прогнозов с прозрачностью и интерпретируемостью результатов, что особенно важно для врачей и пациентов. Гибридные методы, сочетающие классические подходы с нейронными сетями и нечеткой логикой, уже сегодня демонстрируют высокую эффективность в ранней диагностике и поддержке принятия решений. Интеграция этих технологий в практическое здравоохранение способствует созданию более устойчивых и эффективных систем, готовых к вызовам будущего.

Заключение

В статье представлены новые подходы, способствующие интеграции XAI с географическими информационными системами (ГИС) в задачах территориального районирования и медицинской профилактики. Разработанные методы, такие как нечеткая интерпретация Шепли и мультикаскадная иерархия XAI, открывают дополнительные возможности для более точного анализа и интерпретации пространственных данных. Эти инновации улучшают качество принимаемых решений, делая их более обоснованными и понятными для широкого круга пользователей, включая экспертов и специалистов по территориальному планированию. В условиях современного мира, где задачи территориального районирования играют важнейшую роль для устойчивого развития, предложенные решения обеспечивают повышение точности прогнозов и прозрачности процессов принятия решений. Применение мультикаскадной иерархии XAI имеет значительный потенциал в анализе сложных систем, таких как пространственные модели и сценарии устойчивого развития территорий. Этот подход позволяет учитывать многослойные факторы, влияющие на здоровье населения и состояние окружающей среды, что важно для долгосрочного планирования и профилактики заболеваний. Интеграция XAI и ГИС помогает принимать решения, которые базируются на точных данных и сопровождаются объяснениями, понятными для всех участников процесса. Это способствует более эффективному управлению природными ресурсами и обеспечивает более устойчивое развитие городских и сельских территорий.

Кроме того, такие методы помогают не только при планировании территорий, но и в профилактике заболеваний. За счет пространственного анализа можно более точно выявлять зоны риска и разрабатывать стратегии по снижению нагрузки на системы здравоохранения. Включение в модели

экологических, инфраструктурных и социально-экономических факторов, а также их объяснение с помощью XAI улучшает качество прогнозов и повышает доверие к принятым решениям. Мультикаскадная иерархия XAI создает условия для глубинного анализа, который может быть применен в широком спектре задач: от мониторинга состояния окружающей среды до разработки профилактических мер в системе здравоохранения.

Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № 124092700007-4 «Применение объяснительного искусственного интеллекта для интерпретации алгоритмов машинного обучения».

Список источников

1. *Gillespie N., Lockey S., Curtis C., Poo, J., Akbari A.* Trust in Artificial Intelligence: A Global Study [Электронный ресурс] / The University of Queensland, KPMG Australia. – 2023. – 82 p. – Режим доступа: https://www.aiunplugged.io/wp-content/uploads/2023/10/Trust-in-Artificial-Intelligence.pdf (дата обращения: 15.11.2024). DOI:10.14264/00d3c94.

2. *Аверкин А.Н.* Объяснимый искусственный интеллект как часть искусственного интеллекта третьего поколения // Речевые технологии. – 2023. – №1. – С. 4–10.

3. *Yonar A., Yonar H.* Modeling air pollution by integrating ANFIS and metaheuristic algorithms // Modelling Earth Systems and Environment. – 2023. – Vol. 9. – pp. 1621–1631. DOI: 10.1007/s40808-022-01573-6.

4. *Azad A., Karami H., Farzin S., Saeedian A., Kashi H., Sayyahi F.* Prediction of Water Quality Parameters Using ANFIS Optimized by Intelligence Algorithms (Case Study: Gorganrood River) // KSCE Journal of Civil Engineering. – 2018. – Vol. 22. – pp. 2206–2213. DOI: 10.1007/s12205-017-1703-6.

5. *Balasubramanian C., Lal Raja Singh R.* ANFIS-BCMO technique for energy management and consumption of energy forecasting in smart grid with internet of things // Journal of Intelligent and Fuzzy Systems. – 2022. – Vol. 43. –Iss. 6. – P. 7577–7593. DOI: 10.3233/JIFS-221040.

6. *Zacharia P.T.* An Adaptive Neuro-fuzzy Inference System for Robot Handling Fabrics with Curved Edges towards Sewing // Journal of Intelligent and Robotic Systems. – 2010. – Vol. 58. – pp. 193–209. DOI: 10.1007/s10846-009-9362-6.

7. *Hosseini M.S., Zekri M.* Review of Medical Image Classification using the Adaptive Neuro-Fuzzy Inference System // Journal of Medical Signals and Sensors. – 2012. – T. 2. – Iss. 1. – pp. 49–60.

8. *Andrews R., Diederich J., Tickle A.B.* Survey and critique of techniques for extracting rules from trained artificial neural networks // Knowledge-Based Systems. – 1995. – Vol. 8. – Iss. 6. – pp. 373–389. DOI: 10.1016/0950-7051(96)81920-4.

9. *Xue G., Chang Q., Wang J., Zhang K., Pal N.R.* An Adaptive Neuro-Fuzzy System with Integrated Feature Selection and Rule Extraction for High-Dimensional Classification Problems // IEEE Transactions on Fuzzy Systems. – 2023. – Vol. 31. – No. 7. – pp. 2167–2181. DOI: 10.1109/TFUZZ.2022.3220950.

10. *Craven M.W., Shavlik J.W.* Extracting Tree-Structured Representations of Trained Networks // NIPS'95: Proceedings of the 8th International Conference on Neural Information Processing Systems. – Cambridge, MA : MIT Press, 1995. – pp. 24–30.

11. *D'Avila Garcez A., Lamb L.C.* Neurosymbolic AI: The 3rd Wave // Artificial Intelligence Review. – 2023. – Vol. 56. – pp. 12387–12406. DOI: 10.1007/s10462-023-10448-w.

12. *Lundberg S.M., Lee S.I.* A Unified Approach to Interpreting Model Predictions // NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. – Red Hook : Curran Associates, 2017. – pp. 4765–4774.

13. *Molnar C.* Interpretable Machine Learning: A Guide for Making Black Box Models Explainable [Электронный ресурс]. – 2020 – Режим доступа: https://christophm.github.io/interpretable-ml-book/index.html (дата обращения: 03.12.2024).

14. *Mazumder A., Lyons N., Dubey A., Pandey A., Santra A.* XAI-Increment: A Novel Approach Leveraging LIME Explanations for Improved Incremental Learning [Электронный ресурс] // arXiv:2211.01413v1 [cs.LG]. – 2022. – Режим доступа: https://https://arxiv.org/abs/2211.01413v1 (дата обращения: 06.12.2024 г.). DOI: 10.48550/arXiv.2211.01413v1.

15. *Arrieta A.B., Díaz-Rodríguez N., Del Ser J., Bennetot A., Tabik S., Barbado A., Garcia S., Gil-Lopez S., Molina D., Benjamins R., Chatila R., Herrera F.* Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI // Information Fusion. – 2020. – Vol. 58. – pp. 82–115. DOI: 10.1016/j.inffus.2019.12.012.

16. *Ribeiro M.T., Singh S., Guestrin C.* 'Why Should I Trust You?': Explaining the Predictions of Any Classifier // KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – New York : Association for Computing Machinery, 2016. DOI: 10.1145/2939672.2939778.

17. *Simonyan K., Vedaldi A., Zisserman A.* Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps [Электронный ресурс] // arXiv:1312.6034 [cs.CV]. – 2014. – Режим доступа: https://arxiv.org/abs/1312.6034 (дата обращения: 04.12.2024 г.). DOI: 10.48550/arXiv.1312.6034.

18. *Breiman L.* Random Forests // Machine Learning. – 2001. – Vol. 45. – pp. 5–32. DOI: 10.1023/A:1010933404324.

19. *Friedman J.H.* Greedy Function Approximation: A Gradient Boosting Machine // Annals of Statistics. – 2001. – Vol. 29. – No. 5. – pp. 1189–1232. DOI: 10.1214/aos/1013203451.

20. *Rudin C.* Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead // Nature Machine Intelligence. – 2019. – Vol. 1. – pp. 206–215. DOI: 10.1038/s42256-019-0048-x.

21. *Wani A.K., Rahayu F., Ben Amor I., Quadir M., Murianingrum M., Parnidi P., Ayub A., Supriyadi S., Sakiroh S., Saefudin S., Kumar A., Latifah E.* Environmental Resilience through Artificial Intelligence: Innovations in Monitoring and Management // Environmental Science and Pollution Research. – 2024. – Vol. 31. – pp. 18379–18395. DOI: 10.1007/s11356-024-32404-z.

22. *Athanasiou M., Sfrintzeri K., Zarkogianni K., Thanopoulou A.C., Nikita K.S.* An explainable XGBoost-based approach towards assessing the risk of cardiovascular disease in patients with Type 2 Diabetes Mellitus [Электронный ресурс] // arXiv:2009.06629. – 2020. – Режим доступа: https://arxiv.org/abs/2009.06629 (дата обращения 04.12.2024). DOI: 10.48550/arXiv.2009.06629.

23. *Ji Y., Zhi X., Wu Y., Zhang Y., Yang Y., Peng T., Ji L.* Regression analysis of air pollution and pediatric respiratory diseases based on interpretable machine learning // Frontiers in Earth Science. – 2023. – Vol. 11. – 1105140. DOI: 10.3389/feart.2023.1105140.

24. *Janzing D., Minorics L., Blöbaum P.* Feature relevance quantification in explainable AI: A causality problem [Электронный ресурс] // arXiv:1910.13413. – 2019. – Режим доступа: https://arxiv.org/abs/1910.13413 (дата обращения 04.12.2024). DOI: 10.48550/arXiv.1910.13413.

25. *Boulesteix A.L., Bender A., Bermejo J.L., Strobl C.* Random forest Gini importance favours SNPs with large minor allele frequency: impact, sources and recommendations // Briefings in Bioinformatics. – 2012. – Vol. 13. – Iss. 3. – pp. 292–304. DOI: 10.1093/bib/bbr053.

26. *Jiang T., Chen B., Nie Z., Zhehao R., Xu B., Tang S.* Estimation of hourly full-coverage $PM_{2.5}$ concentrations at 1-km resolution in China using a two-stage random forest model // Atmospheric Research. – 2021.– Vol. 248 – 105146. DOI: 10.1016/j.atmosres.2020.105146.

27. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction / 2nd ed. – New York : Springer, 2009. – 745 p. DOI: 10.1007/978-0-387-84858-7.

28. *Greenwell B.M.* pdp: An R Package for Constructing Partial Dependence Plots // The R Journal. – 2017. – Vol. 9. – Iss. 1. – pp. 421–436. DOI: 10.32614/RJ-2017-016.

29. *Goldstein A., Kapelner A., Bleich J., Pitkinet E.* Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation // Journal of Computational and Graphical Statistics. – 2015. – Vol. 24. – Iss. 1. – pp. 44–65. DOI: 10.1080/10618600.2014.907095.

30. *Ciaramella A., Tagliaferri R., Pedrycz W.* The genetic development of ordinal sums // Fuzzy Sets and Systems. – 2005. – Vol. 151. – Iss. 2. – pp. 303–325. DOI: 10.1016/j.fss.2004.07.003.

31. *Cover T.M., Thomas J.A.* Elements of Information Theory. – Hoboken : John Wiley & Sons, 2006. – 748 p. DOI:10.1002/047174882X.

32. *Ghorbanzadeh O., Rostamzadeh H., Blaschke T., Gholaminia K., Aryal J.* A new GIS-based data mining technique using an adaptive neuro-fuzzy inference system (ANFIS) and k-fold cross-validation approach for land subsidence susceptibility mapping // Natural Hazards. – 2018. – Vol. 94. – pp. 497–517. DOI: 10.1007/s11069-018-3449-y.

33. *Veronesi F., Korfiati A., Buffat R., Raubal M.* Assessing accuracy and geographical transferability of machine learning algorithms for wind speed modelling // Societal Geo-innovation. AGILE 2017. Lecture Notes in Geoinformation and Cartography / Ed. A. Bregt, T. Sarjakoski, R. van Lammeren, F. Rip. – Cham: Springer, 2017. – pp. 297–310. DOI: 10.1007/978-3-319-56759-4_17.

34. *Lee Y., Kim D.-Y.* The decision tree for longer-stay hotel guest: the relationship between hotel booking determinants and geographical distance // International Journal of Contemporary Hospitality Management. – 2021. – Vol. 33. – Iss. 6. – pp. 2264–2282. DOI: 10.1108/IJCHM-06-2020-0594.

35. *Li Z.* Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost // Computers, Environment and Urban Systems. – 2022. – Vol. 96. – 101845. DOI: 10.1016/j.compenvurbsys.2022.101845.

36. *Yang C., Chen M., Yuan Q.* The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: An exploratory analysis // Accident Analysis & Prevention. – 2021. – Vol. 158. – 106153. DOI: 10.1016/j.aap.2021.106153.

37. *Simini F., Barlacchi G., Luca M., Pappalardo L.* A Deep Gravity model for mobility flows generation // Nature Communications. – 2021. – Vol. 12. – 6576. DOI: 10.1038/s41467-021-26752-4.

38. *Wang F., Wang Y., Zhang K., Hu M., Weng Q., Zhang H.* Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation // Environmental Research. – 2021. – Vol. 202. – 111660. DOI: 10.1016/j.envres.2021.111660.

39. *Luo R., Xing J., Chen L., Pan Z., Cai X., Li Z., Wang J., Ford A.* Glassboxing deep learning to enhance aircraft detection from SAR imagery // Remote Sensing. – 2021. – Vol. 13. – No. 18. – 3650. DOI: 10.3390/rs13183650.

40. *Sonnewald M., Lguensat R.* Revealing the impact of global heating on North Atlantic circulation using transparent machine learning // Journal of Advances in Modeling Earth Systems. – 2021. – Vol. 13. – Iss. 8. – e2021MS002496. DOI: 10.1029/2021MS002496.

41. *Temenos A., Tzortzis I. N., Kaselimi M., Rallis I., Doulamis A., Doulamis N.* Novel insights in spatial epidemiology utilizing explainable AI (XAI) and remote sensing // Remote Sensing. – 2022. – Vol. 14. – No. 13. – 3074. DOI: 10.3390/rs14133074.

42. *Amiri S.S., Mottahedi S., Lee E.R., Hoque S.* Peeking inside the black-box: Explainable machine learning applied to household transportation energy consumption // Computers, Environment and Urban Systems. – 2021. – Vol. 88. – 101647. DOI: 10.1016/j.compenvurbsys.2021.101647.

43. *Parmar J., Das P., Dave S.M.* A machine learning approach for modelling parking duration in urban land-use // Physica A: Statistical Mechanics and its Applications. – 2021. – Vol. 572. – 125873. DOI: 10.1016/j.physa.2021.125873.

44. *Peng Y., Cui B., Yin H., Zhang Y., Du P.* Automatic SAR change detection based on visual saliency and multihierarchical fuzzy clustering // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. – 2022. – Vol. 15. – pp. 7755–7769. DOI: 10.1109/JSTARS.2022.3199017.

45. *Chang B., Jang G., Kim S., Kang J.* Learning graph-based geographical latent representation for point-of-interest recommendation // CIKM'20: Proceedings of the 29th ACM International Conference on Information and Knowledge Management. – New York: Association for Computing Machinery, 2020. – pp. 135–144. DOI: 10.1145/3340531.3411905.

46. *Wang X., Gu Y., Liu H.* A transfer learning method for the protection of geographical indication in China using an electronic nose for the identification of Xihu Longjing tea // IEEE Sensors Journal. – 2021. – Vol. 21. – Iss. 6. – pp. 8065–8077. DOI: 10.1109/JSEN.2020.3048534.

47. *Chen L., Xu Z., Li Q., Peng J., Wang S., Li H.* An empirical study of adversarial examples on remote sensing image scene classification // IEEE Transactions on Geoscience and Remote Sensing. – 2021. – Vol. 59. – Iss. 9. – pp. 7419–7433. DOI: 10.1109/TGRS.2021.3051641.

48. *Xing J., Sieber R.* The challenges of integrating explainable artificial intelligence into GeoAI // Transactions in GIS. – 2023. – Vol. 27. – Iss. 3. – pp. 626–645. DOI: 10.1111/tgis.13045.

49. *Kim M., Kim D., Jin D., Kim G.* Application of Explainable Artificial Intelligence (XAI) in Urban Growth Modeling: A Case Study of Seoul Metropolitan Area, Korea // Land. – 2023. – Vol. 12. – Iss. 2 – 420. DOI: 10.3390/land12020420.

50. *Choi C., Choi J., Kim C., Lee D.* The Smart City Evolution in South Korea: Findings from Big Data Analytics // The Journal of Asian Finance, Economics and Business. – 2020. – Vol. 7. – No. 1. – pp. 301–311. DOI: 10.13106/jafeb.2020.vol7.no1.301.

51. *Черемисина Е.Н., Финкельштейн М.Я., Любимова А.В.* ГИС INTEGRO – импортозамещающий программно-технологический комплекс для решения геолого-геофизических задач // Геоинформатика. – 2018. – № 3. – С. 8–17.

52. *Черемисина Е.Н., Финкельштейн М.Я., Деев К.В., Большаков Е.М*. ГИС INTEGRO. Состояние и перспективы развития в условиях импортозамещения // Геология нефти и газа. – 2021. – № 3. – С. 31–40. DOI: 10.31087/0016-7894-2021-3-31-40.

53. *Волков Е.Н.* Применение модуля внимания в сверточной нейронной сети в задаче анализа ОКТ-снимков // V Международная конференция по нейронным сетям и нейротехнологиям (NeuroNT'2024) : сборник докладов конференции (Санкт-Петербург, 20 июня 2024 года). – СПб : СПбГЭТУ «ЛЭТИ», 2024. – С. 99–102.

54. *Волков Е.Н., Аверкин А.Н., Ярушев С.А.* К вопросу о разработке интерпретируемого программного решения для диагностики ретинобластомы // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем : материалы Всероссийской конференции с международным участием (Москва, 8–12 апреля 2024 г.). – Москва: РУДН, 2024. – С. 468–471.

55. Averkin A.N., Volkov E.N., Yarushev S.A. Explainable Artificial Intelligence in Deep Learning Neural Nets-Based Digital Images Analysis // Journal of Computer and Systems Sciences International. – 2024. – Vol. 63. – pp. 175–203. DOI: 10.1134/S1064230724700138.

56. *Аверкин А.Н., Волков Е.Н., Ярушев С.А.* Возможности применения нейро-нечетких сетей для задачи классификации офтальмологических изображений // Двадцать первая Национальная конференция по искусственному интеллекту с международным участием (КИИ-2023) : труды конференции (Смоленск, 16–20 октября 2023 г.). – Т. 2. – Смоленск: Принт-Экспресс, 2023. – С. 6–18.

57. *Ярушев С.А., Аверкин А.Н., Ануров А.О.* Разработка модульного решения для приложений-сервисов в персонализированной медицине и здравоохранении 5.0 // Мягкие измерения и вычисления. – 2024. – Т. 79. – № 6. – С. 68–78. DOI: 10.36871/2618-9976.2024.06.007.

References

1. *Gillespie N., Lockey S., Curtis C., Poo, J., Akbari A.* Trust in Artificial Intelligence: A Global Study. The University of Queensland, KPMG Australia. 2023. 82 p. Available at: https://www.aiunplugged.io/wp-content/uploads/2023/10/Trust-in-Artificial-Intelligence.pdf (accessed 15.11.2024). DOI:10.14264/00d3c94.

2. *Averkin A.N.* Explainable Artificial Intelligence as Part of Third-Generation Artificial Intelligence. *Speech Technology*. 2023;(1):4–10.

3. *Yonar A., Yonar H.* Modeling air pollution by integrating ANFIS and metaheuristic algorithms. *Modelling Earth Systems and Environment*. 2023;9:1621–1631. DOI: 10.1007/s40808-022-01573-6.

4. *Azad A., Karami H., Farzin S., Saeedian A., Kashi H., Sayyahi F.* Prediction of Water Quality Parameters Using ANFIS Optimized by Intelligence Algorithms (Case Study: Gorganrood River). *KSCE Journal of Civil Engineering*. 2018;22:2206–2213. DOI: 10.1007/s12205-017-1703-6.

5. *Balasubramanian C., Lal Raja Singh R.* ANFIS-BCMO technique for energy management and consumption of energy forecasting in smart grid with internet of things. *Journal of Intelligent and Fuzzy Systems*. 2022;43(6):7577–7593. DOI: 10.3233/JIFS-221040.

6. *Zacharia P.T.* An Adaptive Neuro-fuzzy Inference System for Robot Handling Fabrics with Curved Edges towards Sewing. *Journal of Intelligent and Robotic Systems*. 2010;58:193–209. DOI: 10.1007/s10846-009-9362-6.

7. *Hosseini M.S., Zekri M.* Review of Medical Image Classification using the Adaptive Neuro-Fuzzy Inference System. *Journal of Medical Signals and Sensors*. 2012;2(1):49–60.

8. *Andrews R., Diederich J., Tickle A.B.* Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-Based Systems*. 1995;8(6):373–389. DOI: 10.1016/0950-7051(96)81920-4.

9. *Xue G., Chang Q., Wang J., Zhang K., Pal N.R.* An Adaptive Neuro-Fuzzy System with Integrated Feature Selection and Rule Extraction for High-Dimensional Classification Problems. *IEEE Transactions on Fuzzy Systems*. 2023;31(7):2167–2181. DOI: 10.1109/TFUZZ.2022.3220950.

10. *Craven M.W., Shavlik J.W.* Extracting Tree-Structured Representations of Trained Networks. In: NIPS'95: Proceedings of the 8th International Conference on Neural Information Processing Systems. Cambridge, MA: MIT Press; 1995. pp. 24–30.

11. *D'Avila Garcez A., Lamb L.C.* Neurosymbolic AI: The 3rd Wave. *Artificial Intelligence Review*. 2023;56:12387–12406. DOI: 10.1007/s10462-023-10448-w.

12. *Lundberg S.M., Lee S.I.* A Unified Approach to Interpreting Model Predictions. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates; 2017. pp. 4765–4774. .

13. *Molnar C.* Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2020. Available at: https://christophm.github.io/interpretable-ml-book/index.html (accessed 10.10.2024).

14. *Mazumder A., Lyons N., Dubey A., Pandey A., Santra A.* XAI-Increment: A Novel Approach Leveraging LIME Explanations for Improved Incremental Learning. arXiv:2211.01413v1 [cs.LG]. 2022. Available at: https://arxiv.org/abs/2211.01413v1 (accessed 06.12.2024). DOI: 10.48550/arXiv.2211.01413v1.

15. *Arrieta A.B., Díaz-Rodríguez N., Del Ser J., Bennetot A., Tabik S., Barbado A., Garcia S., Gil-Lopez S., Molina D., Benjamins R., Chatila R., Herrera F.* Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion*. 2020;58:82–115. DOI: 10.1016/j.inffus.2019.12.012.

16. *Ribeiro M.T., Singh S., Guestrin C.* 'Why Should I Trust You?': Explaining the Predictions of Any Classifier. In: KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery; 2016. DOI: 10.1145/2939672.2939778.

17. *Simonyan K., Vedaldi A., Zisserman A.* Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv:1312.6034 [cs.CV]. 2014. Available at: https://arxiv.org/abs/1312.6034 (accessed 04.12.2024). DOI: 10.48550/arXiv.1312.6034.

18. *Breiman L.* Random Forests. *Machine Learning*. 2001;45:5–32. DOI: 10.1023/A:1010933404324.

19. *Friedman J.H.* Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*. 2001;29(5):1189–1232. DOI: 10.1214/aos/1013203451.

20. *Rudin C.* Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*. 2019;1:206–215. DOI: 10.1038/s42256-019-0048-x.

21. *Wani A.K., Rahayu F., Ben Amor I., Quadir M., Murianingrum M., Parnidi P., Ayub A., Supriyadi S., Sakiroh S., Saefudin S., Kumar A., Latifah E.* Environmental Resilience through Artificial Intelligence: Innovations in Monitoring and Management. *Environmental Science and Pollution Research*. 2024;31:18379–18395. DOI: 10.1007/s11356-024-32404-z.

22. *Athanasiou M., Sfrintzeri K., Zarkogianni K., Thanopoulou A.C., Nikita K.S.* An explainable XGBoost-based approach towards assessing the risk of cardiovascular disease in patients with Type 2 Diabetes Mellitus. arXiv:2009.06629. 2020. Available at: https://arxiv.org/abs/2009.06629 (accessed 04.12.2024). DOI: 10.48550/arXiv.2009.06629.

23. *Ji Y., Zhi X., Wu Y., Zhang Y., Yang Y., Peng T., Ji L.* Regression analysis of air pollution and pediatric respiratory diseases based on interpretable machine learning. *Frontiers in Earth Science*. 2023;11:1105140. DOI: 10.3389/feart.2023.1105140.

24. *Janzing D., Minorics L., Blöbaum P.* Feature relevance quantification in explainable AI: A causality problem. arXiv:1910.13413. 2019. Available at: https://arxiv.org/abs/1910.13413 (accessed 04.12.2024). DOI: 10.48550/arXiv.1910.13413.

25. *Boulesteix A.L., Bender A., Bermejo J.L., Strobl C.* Random Forest Gini Importance Favours SNPs with Large Minor Allele Frequency: Impact, Sources and Recommendations. *Briefings in Bioinformatics*. 2012;13(3):292–304. DOI: 10.1093/bib/bbr053.

26. *Jiang T., Chen B., Nie Z., Zhehao R., Xu B., Tang S.* Estimation of hourly full-coverage $PM_{2.5}$ concentrations at 1-km resolution in China using a two-stage random forest model. *Atmospheric Research*. 2021;248:105146. DOI: 10.1016/j.atmosres.2020.105146.

27. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer; 2009. 745 p. DOI: 10.1007/978-0-387-84858-7.

28. *Greenwell B.M.* pdp: An R Package for Constructing Partial Dependence Plots. *The R Journal*. 2017;9(1):421–436. DOI: 10.32614/RJ-2017-016.

29. *Goldstein A., Kapelner A., Bleich J., Pitkinet E.* Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics.* 2015;24(1):44–65. DOI: 10.1080/10618600.2014.907095.

30. *Ciaramella A., Tagliaferri R., Pedrycz W.* The genetic development of ordinal sums. *Fuzzy Sets and Systems*. 2005;151(2):303–325. DOI: 10.1016/j.fss.2004.07.003.

31. *Cover T.M., Thomas J.A.* Elements of Information Theory. Hoboken: John Wiley & Sons; 2006. 748 p. DOI:10.1002/047174882X.

32. *Ghorbanzadeh O., Rostamzadeh H., Blaschke T., Gholaminia K., Aryal J.* A new GIS-based data mining technique using an adaptive neuro-fuzzy inference system (ANFIS) and k-fold cross-validation approach for land subsidence susceptibility mapping. *Natural Hazards*. 2018;94:497–517. DOI: 10.1007/s11069-018-3449-y.

33. *Veronesi F., Korfiati A., Buffat R., Raubal M.* Assessing accuracy and geographical transferability of machine learning algorithms for wind speed modelling. In: Societal Geo-innovation. AGILE 2017. Lecture Notes in Geoinformation and Cartography. Bregt A., Sarjakoski T., van Lammeren R., Rip F. Cham: Springer; 2017. pp. 297–310. DOI: 10.1007/978-3-319-56759-4_17.

34. *Lee Y., Kim D.-Y.* The decision tree for longer-stay hotel guest: the relationship between hotel booking determinants and geographical distance. *International Journal of Contemporary Hospitality Management.* 2021;33(6):2264–2282. DOI: 10.1108/IJCHM-06-2020-0594.

35. *Li Z.* Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems.* 2022;96:101845. DOI: 10.1016/j.compenvurbsys.2022.101845.

36. *Yang C., Chen M., Yuan Q.* The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: An exploratory analysis. *Accident Analysis & Prevention*. 2021;158:106153. DOI: 10.1016/j.aap.2021.106153

37. *Simini F., Barlacchi G., Luca M., Pappalardo L.* A Deep Gravity model for mobility flows generation. *Nature Communications*. 2021;12:6576. DOI: 10.1038/s41467-021-26752-4.

38. *Wang F., Wang Y., Zhang K., Hu M., Weng Q., Zhang H.* Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation. *Environmental Research*. 2021;202:111660. DOI: 10.1016/j.envres.2021.111660.

39. *Luo R., Xing J., Chen L., Pan Z., Cai X., Li Z., Wang J., Ford A.* Glassboxing deep learning to enhance aircraft detection from SAR imagery. *Remote Sensing*. 2021;13(18):3650. DOI: 10.3390/rs13183650.

40. *Sonnewald M., Lguensat R.* Revealing the impact of global heating on North Atlantic circulation using transparent machine learning. *Journal of Advances in Modeling Earth Systems*. 2021;13(8):e2021MS002496. DOI: 10.1029/2021MS002496.

41. *Temenos A., Tzortzis I. N., Kaselimi M., Rallis I., Doulamis A., Doulamis N*. Novel insights in spatial epidemiology utilizing explainable AI (XAI) and remote sensing. *Remote Sensing*. 2022;14(13):3074. DOI: 10.3390/rs14133074.

42. *Amiri S.S., Mottahedi S., Lee E.R., Hoque S.* Peeking inside the black-box: Explainable machine learning applied to household transportation energy consumption. *Computers, Environment and Urban Systems*. 2021;88;101647. DOI: 10.1016/j.compenvurbsys.2021.101647.

43. *Parmar J., Das P., Dave S.M.* A machine learning approach for modelling parking duration in urban land-use. *Physica A: Statistical Mechanics and its Applications*. 2021;572:125873. DOI: 10.1016/j.physa.2021.125873.

44. *Peng Y., Cui B., Yin H., Zhang Y., Du P.* Automatic SAR change detection based on visual saliency and multihierarchical fuzzy clustering. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2022;15:7755–7769. DOI: 10.1109/JSTARS.2022.3199017.

45. *Chang B., Jang G., Kim S., Kang J.* Learning graph-based geographical latent representation for point-of-interest recommendation. In: CIKM'20: Proceedings of the 29th ACM International Conference on Information and Knowledge Management. New York: Association for Computing Machinery; 2020. pp. 135–144. DOI: 10.1145/3340531.3411905.

46. *Wang X., Gu Y., Liu H.* A transfer learning method for the protection of geographical indication in China using an electronic nose for the identification of Xihu Longjing tea. *IEEE Sensors Journal.* 2021;21(6):8065–8077. DOI: 10.1109/JSEN.2020.3048534.

47. *Chen L., Xu Z., Li Q., Peng J., Wang S., Li H.* An empirical study of adversarial examples on remote sensing image scene classification. *IEEE Transactions on Geoscience and Remote Sensing*. 2021;59(9):7419–7433. DOI: 10.1109/TGRS.2021.3051641.

48. *Xing J., Sieber R.* The challenges of integrating explainable artificial intelligence into GeoAI. *Transactions in GIS*. 2023;27(3):626–645. DOI: 10.1111/tgis.13045.

49. *Kim M., Kim D., Jin D., Kim G.* Application of Explainable Artificial Intelligence (XAI) in Urban Growth Modeling: A Case Study of Seoul Metropolitan Area, Korea. *Land*. 2023;12(2):420. DOI: 10.3390/land12020420.

50. *Choi C., Choi J., Kim C., Lee D.* The Smart City Evolution in South Korea: Findings from Big Data Analytics. *The Journal of Asian Finance, Economics and Business.* 2020;7(1):301–311. DOI: 10.13106/JAFEB.2020.VOL7.NO1.301.

51. *Cheremisina Ye.N., Finkelstein M.Ya., Lyubimova A.V.* GIS INTEGRO — import substitution software for geological and geophysical tasks. *Geoinformatika*. 2018;(3):8–17.

52. *Cheremisina E.N., Finkel'shtein M.Ya., Deev K.V., Bol'shakov E.M.* GIS INTEGRO. status and prospects for development in the context of import substitution. *Russian oil and gas geology*. 2021;(3):31–40. DOI: 10.31087/0016-7894-2021-3-31-40.

53. *Volkov E.N.* Application of spatial attention module in convolutional neural network for OCT images analysis. In: V Mezhdunarodnaya konferentsiya po neironnym setyam i neirotekhnologiyam (NeuroNT'2024): sbornik dokladov konferentsii (St. Petersburg, 20 June 2024). St. Petersburg: SPBGEHTU «LEHTI», 2024. pp. 99–102.

54. *Volkov E.N., Averkin A.N., Yarushev S.A.* Toward the development of an interpretable software solution for diagnosing retinoblastoma. In: Informatsionno-telekommunikatsionnye tekhnologii i matematicheskoe modelirovanie vysokotekhnologichnykh sistem: materialy Vserossiiskoi konferentsii s mezhdunarodnym uchastiem (Moscow, 8–12 April 2024). Moscow: RUDN; 2024. pp. 468–471.

55. *Averkin A.N., Volkov E.N., Yarushev S.A.* Explainable Artificial Intelligence in Deep Learning Neural Nets-Based Digital Images Analysis. *Journal of Computer and Systems Sciences International*. 2024;63:175–203. DOI: 10.1134/S1064230724700138.

56. *Averkin A.N., Volkov E.N., Yarushev S.A.* Vozmozhnosti primeneniya neiro-nechetkikh setei dlya zadachi klassifikatsii oftal'mologicheskikh izobrazhenii [The Potential of Neuro-Fuzzy Networks for the Classification of Ophthalmological Images]. In: Dvadtsat' pervaya Natsional'naya konferentsiya po iskusstvennomu intellektu s mezhdunarodnym uchastiem (KII-2023): trudy konferentsii (Smolensk, 16–20 October 2023). Vol. 2. Smolensk: Print-Ehkspress; 2023. pp. 6–18.

57. *Yarushev S.A., Averkin A.N., Anurov A.O.* Development of a modular solution for service applications in personalized medicine and healthcare 5.0. *Soft Measurements and Computing*. 2024;79(6):68–78. DOI: 10.36871/2618-9976.2024.06.007.

Статья поступила в редакцию 10.06.2024 г., одобрена после рецензирования 25.09.2024 г., принята к публикации 27.11.2024 г.
The article was submitted 10.06.2024; approved after reviewing 25.09.2024; accepted for publication 27.11.2024.

Информация об авторах

Трофимов Юрий Владиславович
Аспирант,
Кафедра экологии и наук о Земле
ФГБОУ ВО «Университет «Дубна»
141980 Дубна, ул. Университетская, д. 19
Программист
Научно-исследовательский центр искусственного интеллекта Государственного университета «Дубна»
e-mail: ura_trofim@bk.ru
ORCID: 0009-0005-6943-7432
Scopus Author ID: 57221465588
ResearcherID: LRT-9153-2024
SPIN-код: 3718-8990
AuthorID: 1083410

Information about authors

Yuriy V. Trofimov
PhD Student
Department of Ecology and Earth Sciences
Dubna State University
19, Universitetskaya str., Dubna, 141980, Russia
Programmer
Research Center for Artificial Intelligence, Dubna State University
e-mail: ura_trofim@bk.ru
ORCID: 0009-0005-6943-7432
Scopus Author ID: 57221465588
ResearcherID: LRT-9153-2024
SPIN-код: 3718-899
AuthorID: 1083410

Аверкин Алексей Николаевич
Кандидат физико-математических наук
Ведущий научный сотрудник
Федеральный исследовательский центр «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН)
119333 Москва, ул. Вавилова, д. 44, кор. 2
Доцент кафедры системного анализа и управления
ФГБОУ ВО «Университет «Дубна»
141980 Дубна, Московской обл., ул. Университетская, д. 19
e-mail: averkin2003@inbox.ru
ORCID: 0000-0003-1571-3583
Scopus Author ID: 57192557075
ResearcherID: L-6541-2013
SPIN-код: 5753-1636
AuthorId: 61393

Черемисина Евгения Наумовна
Доктор технических наук, профессор
Заведующий отделением Геоинформатики
ФГБУ «Всероссийский научно-исследовательский геологический нефтяной институт» (ФГБУ «ВНИГНИ»)
117105 Москва, Варшавское ш., д. 8
Заведующий кафедрой системного анализа и управления
ФГБОУ ВО «Университет «Дубна»
141980 Дубна, Московская обл., ул. Университетская, д. 19
e-mail: e.cheremisina@geosys.ru
ORCID: 0000-0002-6041-8359
SPIN-код: 1472-2283
AuthorID: 119359

Aleksey N. Averkin
Candidate of Physical and Mathematical Sciences
Leading Researcher
Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (FRC CSC RAS)
44, build. 2, Vavilova str., Moscow, 119333, Russia
Associate Professor at the Department of Systems Analysis and Management
Dubna State University
19, Universitetskaya str., Dubna, 141980, Russia
e-mail: averkin2003@inbox.ru
ORCID:0000-0003-1571-3583
Scopus Author ID: 57192557075
ResearcherID: L-6541-2013
SPIN-код: 5753-1636
AuthorID: 61393

Eugenia N. Cheremisina
Doctor of Technical Sciences, Professor
Head of the Geoinformatics Department
All-Russian Research Geological Petroleum Institute (VNIGNI)
8, Varshavskoe shosse, 117105, Moscow, Russia
Head of the Department of Systems Analysis and Management
Dubna State University
19, Universitetskaya str., Dubna, Moscow region, 141980, Russia
e-mail: e.cheremisina@geosys.ru
ORCID: 0000-0002-6041-8359
SPIN-код: 1472-2283
AuthorID: 119359