

Казахский национальный исследовательский технический университет
им. К.И. Сатпаева

УДК 550.83:004.8

На правах рукописи

САДУОВ АЛИШЕР БЕРИКЖАНОВИЧ

8D07104 - Нефтегазовая и рудная геофизика

Технология создания и использования искусственного интеллекта при решении
геопрогнозных задач (на примере минералогических, скважинных и
региональных геолого-геофизических данных Казахстана и зарубежных стран)

Диссертация на соискание степени
доктора философии (PhD)

Научный консультант: Шарапатов А.,
кандидат геолого-минералогических наук,
ассоциированный профессор
кафедры Геофизики и сейсмологии
КазННТУ им. К. И. Сатпаева

Зарубежный научный консультант:
Приезжев И. И.,
доктор технических наук, профессор
кафедры разведочной геофизики и
компьютерных систем
РГУ нефти и газа им. И. М. Губкина;
генеральный директор ТОО “Лаборатория
Приезжева” (Сколково, Москва)

Алматы, 2026

СОДЕРЖАНИЕ

НОРМАТИВНЫЕ ССЫЛКИ	4
ОПРЕДЕЛЕНИЯ	5
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	7
ВВЕДЕНИЕ	8
1 ОБЗОР СОВРЕМЕННОГО СОСТОЯНИЯ ИЗУЧЕННОСТИ ПРОБЛЕМЫ И МЕТОДОЛОГИЯ ИССЛЕДОВАНИЙ	16
1.1 Состояние исследований в минералогической аналитике, геофизике скважин и региональной геоинформатике.....	16
1.1.1 Минералогическая аналитика (минералы и аномалии).....	16
1.1.2 Геофизика скважин (петрофизические основы и методы ML).....	17
1.1.3 Региональная геоинформатика и Mineral Prospectivity Mapping.....	20
1.2 Проблематика неполной разметки и пространственной автокорреляции в геопрогнозных задачах.....	22
1.2.1 Обучение с позитивными и неразмеченными данными (PU-learning).....	22
1.2.2 Пространственно независимая валидация (spatially independent validation).....	24
1.2.3 Калибровка вероятностей и надежности прогнозов.....	24
1.2.4 Контроль области применимости (АОА).....	25
1.2.5 Интерпретируемость моделей и учёт неопределённости.....	25
1.3 Петрофизическая база и геологические предпосылки (электрические методы, плотностно-магнитные свойства, структурные индикаторы).....	26
2 МИНЕРАЛОГИЧЕСКАЯ АНАЛИТИКА РЕДКИХ И АНОМАЛЬНЫХ МИНЕРАЛОВ	28
2.1 Данные, предобработка, понижение размерности.....	28
2.2 Детектирование аномалий (Isolation Forest, One-Class SVM), группировка и геологическая типизация.....	32
2.3 Проверка устойчивости и практическая интерпретация.....	34
3 КАРТИРОВАНИЕ УРАНОПЕРСПЕКТИВНЫХ ГОРИЗОНТОВ ПО ДАННЫМ ГИС: ИНТЕГРАЦИЯ ML И ГЕОСТАТИСТИКИ	41
3.1 Геолого-геофизическая характеристика Северного рудного узла (Шу-Сарысуская впадина), состав и качество каротажных данных.....	41
3.2 Модели литологии и продуктивных интервалов (Random Forest, инженерные γ -признаки).....	45
3.3 Калибровка вероятностей, групповая (по скважинам) spatial-CV, исключение слабых меток из валидации.....	50
3.4 Визуализация по профилям (cubic spline), сравнение с экспертной интерпретацией.....	55
3.5 Чувствительность, ограничители и переносимость на родственные ролл-фронт системы.....	60
4 РЕГИОНАЛЬНАЯ ПЕРСПЕКТИВНОСТЬ (MPM) НА Au-Ag МИНЕРАЛИЗАЦИЮ В ПРЕДЕЛАХ ВЕЛИКОГО БАССЕЙНА (НЕВАДА,	

США): АОА-ОГРАНИЧЕНИЕ, КАЛИБРОВКА И TOP-k НАВИГАЦИЯ РЕШЕНИЙ	63
4.1 Исходные слои: геология, структура, тепловой поток, поля и производные; конструирование признаков.....	63
4.2 PU-learning в MPM, строго пространственная GroupKFold-валидация, OOF-инференс.....	74
4.3 АОА-маскирование (индекс диссимиларности), калибровка вероятностей, метрики.....	79
4.4 Карты неопределенности, риск-ориентированное ранжирование и top-k цели.....	84
4.5 Интерпретируемость (SHAP, PDP/ICE) и согласование с минерал-системной моделью.....	95
4.6 Переносимость методики в условия Казахстана.....	98
4.6.1 Апробация ML-ориентированного MPM в мировой практике: от карт перспективности к выбору буровых целей.....	103
5 СИНТЕЗ И СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭЛЕМЕНТОВ МЕТОДОЛОГИИ	107
5.1 Сравнительный анализ по компонентам методологической схемы: данные, валидация, калибровка, интерпретируемость, неопределённость.....	107
5.2 Универсальность приёмов от минералогии к ГИС и региональной MPM-геологоразведке.....	112
5.3 Ограничения и области применимости; рекомендации для практики.....	114
ЗАКЛЮЧЕНИЕ	118
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	120

НОРМАТИВНЫЕ ССЫЛКИ

В настоящей диссертации использованы ссылки на следующие документы и стандарты:

- 1 Правила присуждения ученых степеней, утвержденных приказом МОН РК от 28.09.2018 года №512.
- 2 Государственный общеобязательный стандарт образования Республики Казахстан послевузовское образование. Докторантура. ГОСО РК 5.04.034-2011.
- 3 ГОСТ 7.1-2003. Библиографическая запись. Библиографическое описание. Общие требования и правила составления.
- 4 ГОСТ 7.32-2001 (изменения от 2006 г.). Отчет о научно-исследовательской работе. Структура и правила оформления.
- 5 ГОСТ 7.9-95 (ИСО 214-74) Система стандартов по информации, библиотечному и издательскому делу. Реферат и аннотация. Общие требования.
- 6 ГОСТ 8.417-81 Государственная система обеспечения единства измерений. Единицы физических величин.

ОПРЕДЕЛЕНИЯ

Минералогическая аномалия - участок пространства, в пределах которого содержание одного или нескольких минералов статистически значимо отличается от фоновых значений и может указывать на наличие рудной минерализации.

Минералогическая аналитика - совокупность методов количественного и качественного анализа минерального состава пород и руд с использованием лабораторных, геохимических и вычислительных технологий.

Геофизические исследования скважин - комплекс методов регистрации физических параметров горных пород в стволе скважины, применяемый для определения литологических, коллекторских и фильтрационно-емкостных характеристик разреза.

Фильтрационно-емкостные свойства - совокупность параметров горных пород, характеризующих способность пород накапливать и пропускать флюиды (пористость, проницаемость, насыщенность и сопутствующие петрофизические величины).

Mineral Prospectivity Mapping - совокупность методов построения пространственно-распределенных индексов перспективности, количественно оценивающих вероятность присутствия рудной минерализации на основе геологических, геофизических, геохимических и пространственных признаков.

Позитивно-неразмеченное обучение (Positive-Unlabeled learning) - класс методов машинного обучения, в которых для целевого класса доступны только позитивные примеры, а оставшиеся наблюдения рассматриваются как смесь неразмеченных позитивов и негативов.

Пространственно независимая валидация - подход к оценке качества моделей, при котором обучающая и тестовая выборки формируются с учетом пространственного разнесения, что снижает эффект пространственной автокорреляции и переоценки прогностической способности.

Калибровка вероятностных прогнозов - совокупность процедур и моделей, направленных на согласование предсказанных вероятностей с фактической частотой наступления событий, обеспечивающее интерпретируемость и надежность вероятностных оценок.

Область применимости модели - часть признакового или географического пространства, в пределах которой распределение данных близко к обучающему множеству и прогнозы модели могут считаться статистически обоснованными.

Индекс диссимilarityности - числовая характеристика удаленности конкретного объекта от множества обучающих данных в пространстве

признаков, используемая для выявления областей с повышенной неопределенностью прогноза.

Интерпретируемость модели - свойство модели машинного обучения, заключающееся в возможности количественно и качественно оценить вклад отдельных признаков и их комбинаций в формирование прогноза.

Индекс перспективности - скалярный показатель, получаемый в результате моделирования, который используется для ранжирования пространственных ячеек по степени минералого-геологической перспективности.

Зона высокой перспективности - область пространства, которая по значению индекса перспективности и результатам геологической интерпретации относится к наиболее благоприятным для локализации рудной минерализации.

Акт верификации - комплекс полевых, лабораторных или камеральных процедур, направленных на подтверждение или опровержение прогноза модели в выделенных целевых зонах.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

- AI - Artificial Intelligence, искусственный интеллект.
- ML - Machine Learning, методы машинного обучения.
- ГИС - геофизические исследования скважин.
- ГК - гамма-картаж.
- ПС - поляризация самопроизвольная.
- ДЗЗ - дистанционное зондирование Земли.
- ЦМР - цифровая модель рельефа.
- MPM (Mineral Prospectivity Mapping) - минералого-поисковое картирование перспективности (прогноз и ранжирование территорий по вероятности локализации заданного типа минерализации).
- SVM (Support Vector Machine) - метод опорных векторов.
- RF (Random Forest) - ансамблевый алгоритм машинного обучения «случайный лес».
- GBM (Gradient Boosting Machine) - градиентный бустинг деревьев решений.
- PU-learning (Positive-Unlabeled learning) - обучение с позитивными и неразмеченными данными.
- SCAR (Selected Completely At Random) - допущение полностью случайного отбора размеченных позитивов.
- SAR (Selected At Random) - допущение зависимо от признаков отбора размеченных позитивов.
- ROC AUC (Receiver Operating Characteristic Area Under Curve) - площадь под ROC-кривой.
- PR AUC (Precision-Recall Area Under Curve) - площадь под кривой «precision-recall».
- Brier score - интегральная метрика калибровки вероятностных прогнозов.
- RTP (Reduction to the Pole) - преобразование аномального магнитного поля к полюсу.
- THD (Total Horizontal Derivative) - модуль полного горизонтального градиента поля.
- AOA (Area of Applicability) - область применимости модели.
- DI (Dissimilarity Index) - индекс диссимилярности, мера удаленности точки от обучающего множества.
- SHAP (SHapley Additive exPlanations) - метод шеплиевских аддитивных разложений для интерпретации моделей.
- PDP (Partial Dependence Plot) - график частичной зависимости.
- ICE (Individual Conditional Expectation) - графики индивидуальных условных ожиданий.
- precision@K - доля объектов целевого класса в верхних K процентах ранжирования по прогнозному индексу.

ВВЕДЕНИЕ

Актуальность темы исследования. Современная геологоразведка развивается в условиях постоянного увеличения объема, разнородности и пространственной сложности геолого-геофизической информации. При поиске и оценке месторождений твердых полезных ископаемых одновременно используются минералогические описания, результаты лабораторных анализов, данные геофизических исследований скважин, материалы геологического картирования, сведения о разломно-структурном строении, гравитационные и магнитные поля, геохимические показатели, тепловые данные и другие пространственно распределенные признаки. Такая информационная база создает предпосылки для более полного анализа рудоконтролирующих факторов, однако одновременно усложняет задачу их совместной интерпретации.

В традиционной практике геологоразведки прогнозные решения часто принимаются на основе экспертного анализа карт, разрезов, отдельных геофизических аномалий или пороговых значений по отдельным признакам. Такие подходы остаются необходимыми и геологически обоснованными, однако они не всегда позволяют учитывать сложные многомерные связи между признаками, неоднородность исходных данных, пространственную зависимость наблюдений и неполноту фактического контроля. Особенно остро эта проблема проявляется в условиях, когда достоверно подтверждены только отдельные продуктивные интервалы, проявления или месторождения, тогда как остальная часть разреза или территории не может быть уверенно отнесена к непродуктивной. В геологическом смысле это означает, что отсутствие подтвержденной минерализации не является доказательством ее отсутствия, а отражает ограниченность бурения, опробования, кернового контроля или полевой изученности.

В минералогических задачах указанная проблема проявляется в том, что редкие и статистически нетипичные минералы представлены малочисленными группами, а их поисковое значение не всегда может быть установлено с помощью стандартной классификации. В скважинной геофизике подтвержденные продуктивные интервалы обычно ограничены по числу скважин и глубинных отметок, тогда как значительная часть разреза остается фоном неопределенного статуса. В региональных задачах картирования перспективности известные месторождения и проявления составляют лишь небольшую часть потенциально рудоносной территории, а пространственное распределение данных определяется историей изученности, доступностью участков, глубиной залегания объектов и неоднородностью геологических условий.

В этих условиях актуальной становится разработка такой технологии геопрогноза, которая позволяла бы не только применять методы искусственного интеллекта к геоданным, но и адаптировать их к реальной логике геологоразведки. Такая технология должна учитывать неполноту подтверждающих данных, пространственную связанность наблюдений,

необходимость независимой проверки качества прогноза, геологическую интерпретируемость полученных результатов, границы допустимого переноса прогноза на новые участки, а также потребность в количественном ранжировании объектов для последующих поисково-разведочных работ.

Следовательно, актуальность диссертационной работы определяется необходимостью разработки и научного обоснования воспроизводимой технологии применения интеллектуальных методов анализа геолого-геофизических данных, обеспечивающей получение геологически интерпретируемых, вероятностно согласованных и практически применимых прогнозных решений при дефиците достоверно подтвержденной информации.

Тема диссертационной работы формулируется следующим образом: «Технология создания и использования искусственного интеллекта при решении геопрогнозных задач (на примере минералогических, скважинных и региональных геолого-геофизических данных Казахстана и зарубежных стран)».

Объект и предмет исследования. Объектом исследования являются геопрогнозные процессы, связанные с поиском и оценкой месторождений твердых полезных ископаемых на основе комплексных минералогических, геофизических и геоинформационных данных. В диссертации рассматриваются три репрезентативных класса задач: выявление редких и статистически нетипичных минералов по многомерным минералогическим данным, картирование потенциально уранопродуктивных горизонтов по данным геофизических исследований скважин и региональное прогнозирование перспективности Au-Ag объектов по интегрированным геологическим, структурным, геофизическим и тепловым признакам.

Предметом исследования является совокупность методологических принципов и алгоритмических решений, обеспечивающих построение, проверку качества, согласование вероятностной шкалы и практическое применение прогнозных моделей в указанных геопрогнозных задачах. В качестве ключевых компонентов предмета выделяются методы работы в условиях, когда подтвержденность объектов задана выборочно и неравномерно (смещенность, обусловленная сетью и приоритетами работ), схемы проверки качества по пространственно разнесенным группам (пространственно независимая валидация), подходы к выделению области допустимой экстраполяции прогноза и оценке неопределенности, а также методы анализа вклада факторов и формирование приоритетного ранжирования перспективных зон с контролируемым уровнем риска.

Цель исследования. Целью диссертационной работы является разработка и обоснование комплексной технологии применения интеллектуальных методов анализа минералогических, геофизических и геоинформационных данных в условиях неполного фактического контроля и пространственной зависимости наблюдений. В рамках цели должно быть обеспечено получение вероятностных прогнозов с проверяемым смыслом (то есть с согласованием вероятностной шкалы с фактическими подтверждениями), выделение области

допустимой интерпретации прогноза (ограничение экстраполяции), а также геологически обоснованная трактовка влияния ключевых признаков для риск-ориентированного принятия решений в геологоразведке.

Задачи исследования. Для достижения поставленной цели в диссертационной работе были решены следующие задачи:

1. Выполнить анализ современного состояния исследований по применению методов искусственного интеллекта в минералогической аналитике, интерпретации данных геофизических исследований скважин и региональном картировании перспективности месторождений, с выделением основных методических ограничений, связанных с неполнотой фактического контроля, пространственной зависимостью данных и неопределенностью прогноза.

2. Разработать подход к выявлению редких и статистически нетипичных минералов по многомерным минералогическим данным на основе совместного применения методов снижения размерности, выявления нетипичных объектов и группирования результатов, а также оценить геологическую интерпретируемость выделенных минералого-геохимических ассоциаций.

3. Сформировать методику вероятностного выделения потенциально уранопродуктивных интервалов по данным геофизических исследований скважин в условиях, когда достоверно подтверждены только отдельные продуктивные интервалы, а остальная часть разреза рассматривается как фон неопределенного статуса.

4. Разработать методологию регионального картирования перспективности Au-Ag минерализации на основе интеграции геологических, структурных, геофизических и тепловых признаков, с учетом пространственно независимой проверки качества, ограничения области допустимой геологической экстраполяции и ранжирования перспективных участков.

5. Синтезировать результаты минералогического, скважинного и регионального уровней исследования в единую технологическую схему применения искусственного интеллекта при решении геопрогнозных задач, определить области ее применимости, ограничения и практические направления использования в геологоразведке.

Методологическая база исследования. Методологическую основу диссертационной работы составляют положения минералогии, генетической геологии, петрофизики, разведочной геофизики, теории потенциальных полей, математической статистики, геостатистики и пространственного анализа. Исследование основано на представлении о том, что прогноз рудной минерализации должен учитывать не только расчетные зависимости, но и геологическую природу исходных признаков, условия формирования месторождений, пространственное положение рудоконтролирующих факторов и степень фактической подтвержденности объектов.

Для анализа многомерных минералогических данных применялись методы снижения размерности, выявления нетипичных объектов и

группирования, что позволило переходить от отдельных аномальных проб к устойчивым минералого-геохимическим ассоциациям. Для решения скважинных и региональных задач использовались методы машинного обучения, рассматриваемые как инструменты выявления многомерных зависимостей между геологическими, геофизическими и пространственными признаками в условиях подтвержденных объектов и неопределенного фона.

Проверка качества прогнозных моделей выполнялась с использованием независимого контроля по пространственно или геологически разобщенным группам: по отдельным скважинам в скважинной задаче и по пространственно разнесенным блокам в региональной задаче. Такой подход позволил снизить риск завышения качества прогноза вследствие пространственной близости данных.

Для практической интерпретации результатов применялись калибровка прогнозных вероятностей, оценка области допустимой экстраполяции, анализ неопределенности и оценка вклада исходных факторов. Это позволило выделять участки обоснованного прогноза и формировать перечень приоритетных объектов с учетом как прогнозного индекса, так и надежности его интерпретации.

Основные научные результаты. В ходе выполнения диссертационной работы был получен комплекс научных результатов, направленных на развитие технологии применения искусственного интеллекта при решении геопрогнозных задач в условиях неполного фактического контроля, пространственной зависимости данных и неоднородной геологоразведочной изученности. Разработанная технология была апробирована на трех уровнях геологического прогноза: минералогическом, скважинном и региональном. Каждый из этих уровней характеризуется собственным типом исходных данных, масштабом исследования и способом подтверждения прогнозируемых объектов, однако во всех случаях решалась общая научная задача - повышение обоснованности прогноза при ограниченности достоверных наблюдений.

На минералогическом уровне был разработан и реализован подход к выявлению редких и статистически нетипичных минералов по многомерным минералогическим данным. В отличие от простого поиска отдельных отклоняющихся объектов, в работе была применена последовательная схема, включающая снижение размерности признакового пространства, выявление нетипичных минералогических объектов и последующее группирование полученных результатов. Это позволило перейти от набора разрозненных статистических аномалий к устойчивым минералого-геохимическим ассоциациям, которые могут быть сопоставлены с определенными минералообразующими условиями и использованы как дополнительные поисковые индикаторы. Было показано, что выделенные ассоциации отличаются от фоновой совокупности минералов по комплексу признаков и имеют геологически интерпретируемый характер.

На скважинном уровне был сформирован комплекс моделей для вероятностного выделения потенциально уранопродуктивных интервалов по

данным геофизических исследований скважин. При постановке задачи была учтена реальная разведочная ситуация, при которой достоверно подтверждены только отдельные продуктивные интервалы, тогда как остальная часть разреза не может автоматически рассматриваться как непродуктивная. В связи с этим неподтвержденная часть разреза была интерпретирована как фон неопределенного статуса. Проверка качества выполнялась по независимым скважинам, что позволило исключить завышение оценки прогноза за счет сходства соседних глубинных отсчетов внутри одной скважины. Полученные вероятностные разрезы после геостатистического согласования по профилям показали соответствие керновым данным и экспертной геологической интерпретации.

На региональном уровне была разработана методология картирования перспективности Au-Ag минерализации на основе совместного анализа геологических, структурных, геофизических и тепловых признаков. В рамках данной методологии прогноз перспективности был построен с учетом ограниченного числа подтвержденных проявлений и месторождений, а его проверка была выполнена по пространственно разнесенным блокам. Для повышения надежности интерпретации была выделена область геологической аналогии, то есть часть территории, в пределах которой сочетание исходных признаков сопоставимо с условиями, представленными в обучающих данных. За пределами такой области прогноз рассматривался как экстраполяционный и требующий осторожной геологической трактовки. На основе вероятностного ранжирования, оценки неопределенности и выделения верхней доли наиболее перспективных участков был сформирован перечень приоритетных целей для дальнейших поисково-разведочных работ.

В совокупности полученные результаты позволили обосновать технологическую схему геопрогноза, в которой методы искусственного интеллекта используются не как изолированные вычислительные приемы, а как часть полного геолого-прогнозного цикла. В данный цикл входят геологическая постановка задачи, подготовка и параметризация информативных признаков, построение прогнозной модели, независимая проверка качества, вероятностная интерпретация результата, контроль области допустимого применения прогноза и формирование приоритетов для последующей геологоразведки.

Научная новизна диссертационной работы заключается в разработке и научном обосновании технологии применения искусственного интеллекта при решении геопрогнозных задач, ориентированной на условия неполного фактического контроля, пространственной зависимости геолого-геофизических данных и необходимости геологически интерпретируемого прогноза для поисково-разведочных решений.

1. Разработан подход к автоматизированному выявлению редких и статистически нетипичных минералов в многомерных минералогических данных, основанный на переходе от отдельных аномальных объектов к устойчивым минералого-геохимическим ассоциациям. Научная значимость данного подхода состоит в том, что статистически нетипичные минералы

рассматриваются не только как отклонения от фоновой совокупности, но и как возможные индикаторы различных минералообразующих условий и поисковые критерии соответствующих типов оруденения.

2. Обоснована постановка вероятностного выделения потенциально уранопродуктивных интервалов по данным геофизических исследований скважин в условиях неполной подтвержденности разреза. В данной постановке достоверно установленные продуктивные интервалы используются как опорные данные, тогда как остальная часть разреза не отождествляется с непродуктивной толщей, а рассматривается как фон неопределенного статуса. Это позволяет учитывать реальную разведочную ситуацию, при которой керновый контроль, опробование и достоверная информация о продуктивности охватывают только часть скважинного разреза.

3. Сформирована методология регионального ранжирования перспективности Au-Ag оруденения по интегрированным геолого-структурным, геофизическим и тепловым признакам. Методология основана на пространственно независимой проверке качества, ограничении интерпретации прогноза областью геологической аналогии и выделении верхней доли наиболее перспективной площади. Это позволяет использовать прогнозную карту не только как результат моделирования, но и как инструмент количественного обоснования приоритетных участков для дальнейших геологоразведочных работ.

Практическая значимость диссертационной работы определяется возможностью использования разработанной технологии для повышения обоснованности геологоразведочных решений в условиях ограниченного фактического контроля и неоднородной изученности территории или разреза.

В минералогической аналитике предложенный подход может применяться для выявления редких и статистически нетипичных минералов, типизации минералого-геохимических ассоциаций и отбора объектов, требующих экспертного анализа или дополнительного лабораторного подтверждения. Это позволяет повысить информативность больших минералогических баз и использовать редкие минеральные ассоциации как дополнительные поисковые признаки при прогнозировании оруденения.

В задачах интерпретации данных геофизических исследований скважин разработанный подход позволяет строить вероятностные разрезы потенциальной урановой продуктивности, пригодные для выбора интервалов опробования, уточнения геологической интерпретации, планирования доразведки и сопоставления каротажных признаков с керновыми данными. Такой результат особенно важен для объектов, где подтвержденные продуктивные интервалы представлены ограниченно, а значительная часть разреза остается недостаточно проверенной.

В региональных задачах картирования перспективности разработанная методология позволяет формировать ранжированный перечень перспективных участков с учетом прогнозной оценки, неопределенности и области допустимой геологической экстраполяции. Это может быть использовано

геологоразведочными организациями при выборе участков для полевых работ, детализации геофизических съемок, постановке бурения и распределении ресурсов на ранних стадиях поисково-оценочного процесса.

Положения, выносимые на защиту:

1. Автоматизированное выделение устойчивых ассоциаций редких и статистически нетипичных минералов в глобальной минералогической базе данных позволяет формировать воспроизводимые минералого-геохимические группы, интерпретируемые как отражение различных генетических обстановок; такие группы пространственно приурочены к зонам повышенной рудоносности и могут использоваться в качестве минералогических критериев прогнозирования соответствующих типов оруденения.

2. Вероятностное выделение потенциально уранопродуктивных интервалов по данным геофизических исследований скважин может быть корректно реализовано в условиях неполной подтвержденности разреза, когда достоверно подтверждены только продуктивные интервалы, а остальная толща рассматривается как фон неопределенного статуса; при независимой проверке по скважинам исключается перенос информации между разрезами, а получаемые вероятностные разрезы урановой продуктивности после геостатистического согласования вдоль профилей согласуются с керновыми данными и экспертной интерпретацией.

3. Воспроизводимое региональное ранжирование перспективности территории на Au-Ag оруденение по интегрированным геолого-геофизическим признакам обеспечивается при проверке по пространственно разнесенным блокам и ограничении интерпретации прогнозов областью геологической аналогии; при этом известные проявления и месторождения концентрируются в пределах верхней доли наиболее перспективной площади, а результатом является количественно обоснованный перечень приоритетных участков для дальнейших геологоразведочных работ.

Апробация работы. Основные результаты научных исследований, представленных в диссертации, неоднократно обсуждались и получали одобрение научного сообщества. Промежуточные и итоговые выводы рассматривались на научных семинарах и заседаниях кафедры, а также на международной научно-практической конференции «75 лет кафедры «Геофизика и сейсмология» – образование, наука и судьбы» КазНИТУ имени К. И. Сатпаева. Отдельные результаты по минералогической аналитике, картированию урановых горизонтов и региональному Mineral Prospectivity Mapping докладывались и обсуждались в Бакинском государственном университете (г. Баку, Азербайджан) в рамках научных семинаров и мероприятий, приуроченных к юбилею профильных кафедр. В ходе работы осуществлялись регулярные консультации с зарубежным научным консультантом и другими исследователями, специализирующимися на машинном обучении в геологии и геофизике, что способствовало уточнению методологических подходов и сопоставлению полученных результатов с мировым уровнем исследований в данной области.

Результаты исследований опубликованы в 11 научных работах в изданиях Казахстана и дальнего зарубежья, рекомендованных Комитетом по обеспечению качества в сфере науки и высшего образования МНВО РК. Из них: 1 статья в журнале, входящем в квартиль Q1, и 1 статья в журнале, входящем в квартиль Q3 базы данных Scopus (БД Scopus); 4 статьи в журналах, рекомендованных МНВО РК; 5 трудов в сборниках материалов международных и республиканских конференций.

Научным консультант диссертационной работы – А. Шарапатов, кандидат геолого-минералогических наук, ассоциированный профессор кафедры Геофизика и сейсмология. Научный консультант осуществлял научно-методическое руководство исследованием, участвовал в постановке задач, обсуждении промежуточных и итоговых результатов, а также в формировании концептуальной структуры диссертации.

Зарубежным научным консультантом является И. И. Приезжев, доктор технических наук, профессор кафедры разведочной геофизики и компьютерных систем, генеральный директор ТОО «Лаборатория Приезжева». Зарубежный консультант обеспечивал сопоставление методологических решений и полученных результатов с международной практикой применения методов искусственного интеллекта в геофизике и геологии, способствовал формулировке ключевых направлений развития исследований и их увязке с задачами цифровой трансформации геологоразведки.

Автор выражает особую признательность коллективу кафедры Геофизики и сейсмологии КазННТУ им. К. И. Сатпаева за поддержку при выполнении исследований и подготовке диссертации, конструктивные замечания и обсуждения, способствовавшие уточнению методологических подходов и повышению научного уровня работы.

Структура и объем диссертации. Диссертация изложена на 132 страницах компьютерного текста и состоит из введения, 5 разделов, заключения и списка использованных источников, состоящего из 161 наименования. Диссертация включает 47 рисунков и 7 таблиц.

1 ОБЗОР СОВРЕМЕННОГО СОСТОЯНИЯ ИЗУЧЕННОСТИ ПРОБЛЕМЫ И МЕТОДОЛОГИЯ ИССЛЕДОВАНИЙ

1.1 Состояние исследований в минералогической аналитике, геофизике скважин и региональной геоинформатике

1.1.1 Минералогическая аналитика (минералы и аномалии)

Современная минералогическая аналитика выполняется в условиях, когда для подавляющей части проб отсутствует однозначная экспертная атрибуция по редким минералам и редким минеральным ассоциациям. Редкие по встречаемости минералы и их разновидности представлены малочисленными группами, вследствие чего наблюдается выраженная неравномерность представления классов в исходных массивах данных, а применение стандартных схем распознавания, ориентированных на хорошо представленные типы, оказывается методически ограниченным. Дополнительной особенностью является то, что для большинства проб корректная отметка «достоверно отсутствует данный минерал» обычно не фиксируется в явном виде: во многих базах данных регистрируется факт обнаружения, тогда как отсутствие может отражать как истинное отсутствие, так и особенности аналитической методики, предела обнаружения, отбора, подготовки проб или неполноту описания. В результате, вместо прямой классификации «есть - нет» все чаще используются подходы, ориентированные на выявление статистически редких и нетипичных проб относительно общей совокупности (в терминологии data analysis это часто описывается как anomaly detection и one-class classification - выявление нетипичных объектов без задания достоверного класса “отсутствия”) [1, 2]. Такие методы методически оправданы в ситуациях, когда достоверная контрольная выборка не формируется, а интерес сосредоточен на поиске необычных сочетаний физико-химических признаков, которые могут быть связаны с редкими минералами, редкими условиями кристаллизации или специфическими минералообразующими обстановками.

В литературе показано, что применение одноклассовых процедур поиска статистически редких проб позволяет выделять группы минералов, отличающиеся аномальными сочетаниями составов и физических характеристик, то есть выходящие за пределы типичного “фона” распространенных минералов [3]. Далее, для перевода статистического результата в минералого-геологическую интерпретацию выделенные нетипичные пробы, как правило, систематизируются и типизируются. Для этого используется снижение размерности (например, PCA как метод приведения многомерного пространства признаков к компактному представлению) и последующее группирование проб по сходству (clustering, в частности k-средних) [1]. Практический смысл такого шага заключается в том, что единичные “нетипичные” точки переводятся в набор устойчивых ассоциаций, каждая из которых может быть соотнесена с отдельными

минералого-геохимическими признаками, вероятными парагенезисами и возможными генетическими обстановками. Подобная связка «выявление нетипичных проб - типизация ассоциаций» способствует более предметной постановке дальнейших проверок: уточняются диагностические признаки, формируются гипотезы о происхождении, выделяются кандидаты для экспертного просмотра и независимого подтверждения.

При этом в исследованиях последовательно подчеркивается, что для малочисленных групп редких минералов чувствительность алгоритмов к настройкам и составу входных данных остается существенным источником неопределенности. При небольшом числе нетипичных проб даже умеренные изменения в наборе признаков, методах нормировки или параметрах алгоритма способны менять состав выделяемых “аномалий” и их группирование. По этой причине в современной литературе рассматривается необходимость строгой проверки устойчивости результатов, а также независимого подтверждения наиболее значимых находок минералогами-экспертами или дополнительными аналитическими методами [4, 5]. Вместе с тем, одноклассовые схемы уже доказали прикладную ценность в задачах поисковой геологии, поскольку позволяют в автоматизированном режиме выявлять новые или редкие минеральные ассоциации в больших массивах спектральных, химических и минералогических данных и затем сопоставлять их с пространственной структурой изучаемой территории. В рамках выполненных автором исследований, представленных в публикации в журнале первого квартала, была продемонстрирована практическая интеграция одноклассовых методов с минералогической интерпретацией: автоматически выделенные статистически нетипичные минералы и ассоциации были пространственно приурочены к участкам повышенного поискового интереса, что обеспечивает методически корректный переход от “выбросов” в данных к минералогическим критериям прогнозирования перспективных типов оруденения.

1.1.2 Геофизика скважин (петрофизические основы и методы ML)

Геофизические исследования скважин (ГИС) опираются на петрофизические закономерности, связывающие измеряемые поля с фильтрационно-емкостными параметрами и литологическим строением разреза. Классическим примером является закон Арчи, который устанавливает связь между удельным сопротивлением полностью водонасыщенной породы, сопротивлением пластовой воды и пористостью [6]. В базовой форме данная зависимость может быть записана как

$$\rho_0 = \rho_f * \varphi^{-m} \quad (1.1)$$

где ρ_0 - сопротивление полностью водонасыщенного образца, ρ_f - сопротивление порового раствора, φ - пористость, m - цементационный коэффициент, отражающий геометрию порового пространства [6,7]. Уравнение Арчи (1942) остается фундаментальной основой интерпретации электрических

диаграмм и широко применяется при анализе сопротивления, индукционного каротажа и родственных методов. В дальнейшем зависимость была обобщена: вводился коэффициент a , учитывающий отклонения реальных пород от идеализированной модели, и рассматривались случаи многофазного насыщения [8]. При этом физический смысл сохранялся: электрические параметры каротажа чувствительны к пористости и насыщению, а значит, к фильтрационно-емкостным свойствам и флюидному режиму пород [7].

Необходимо также учитывать, что электрические методы существенно реагируют на литологию и минералогический состав: глинистая составляющая повышает эффективную проводимость за счет поверхностного ионного переноса, в результате чего в глинистых песчаниках кажущееся сопротивление может быть занижено относительно чистых коллекторов при сходной пористости и насыщении. Для корректной интерпретации требуются поправки на влияние диаметра скважины, глинистой корки, инвазии промывочной жидкости и других технологических факторов, способных исказить измерения. Современные обзоры и переоценки закона Арчи и его обобщений рассматриваются как опорные для построения методически корректной интерпретационной рамки электрических методов [7, 8].

Классические петрофизические зависимости легли в основу ранних методик выделения продуктивных интервалов по ГИС (каротаж), однако в сложных разрезах ручная интерпретация и применение жестких эмпирических критериев не всегда обеспечивают достаточную устойчивость и воспроизводимость. По этой причине в последние десятилетия в скважинной геофизике все шире внедряются методы интеллектуального анализа данных для автоматизированной классификации литотипов и оценки потенциально продуктивных интервалов. Показано, что алгоритмы типа SVM (Support Vector Machine) могут быть эффективны при распознавании литофаций по каротажным параметрам, включая варианты с “нечетким” отбором признаков, обеспечивающие преимущества относительно классического дискриминантного анализа и ряда нейросетевых реализаций [9]. Параллельно демонстрировались возможности нейросетевых подходов и решений переноса, направленных на повышение устойчивости распознавания при изменении условий разреза и распределений признаков [10]. Деревья решений и ансамблевые модели получили широкое распространение вследствие устойчивости к шуму и удобства интерпретации; в частности, Random Forest показал практическую эффективность при выделении потенциально продуктивных интервалов в условиях слабоконтрастных коллекторов и сложных геофизических сигналов [9, 10].

Отдельного внимания заслуживает методически корректная проверка качества в скважинных задачах, поскольку при неправильной организации контроля неизбежно возникает переоценка точности. Если данные одной и той же скважины попадают и в обучение, и в проверку (даже при разбиении по глубине), формируется зависимость наблюдений, связанная с устойчивыми индивидуальными особенностями разреза, аппаратуры и условий измерений,

что приводит к завышенным метрикам. Поэтому корректной практикой считается проверка по независимым скважинам, то есть обучение на наборе скважин и контроль на других скважинах, ранее не участвовавших в построении модели [4, 5]. На практике для этого применяются групповые схемы перекрестной проверки, где единицей группировки выступает скважина: на каждой итерации целиком исключается одна (или группа) скважина из обучения и используется для контроля. Подобная схема исключает “перетекание” информации между обучением и тестом и обеспечивает оценку переносимости решения на новые разрезы [5].

При переходе от интерпретации отдельных скважин к построению разрезов и профилей дополнительно применяется интеграция интеллектуальных моделей с геостатистическими процедурами. Прогнозные значения, полученные на основе каротажа, используются как опорные точки, после чего выполняется пространственное согласование по профилям или площади методами геостатистики (кригинг, сплайны и др.), позволяющими учитывать пространственную корреляцию свойств. При этом методически критичным остается исключение ситуации, когда пространственно близкие точки одновременно участвуют в обучении и контроле, поскольку автокорреляция способна завышать показатели качества аналогично региональному картированию [4]. В современной литературе обосновано, что надежная оценка обобщающей способности должна базироваться на “блокировке” данных по естественным геологическим единицам независимости: скважинам, пластам, участкам или районам [5].

В целом методы интеллектуального анализа уже стали значимым инструментом интерпретации ГИС (каротаж), обеспечивая автоматизированное выделение литотипов и оценку потенциально продуктивных интервалов, а также выявление нелинейных многомерных связей между каротажными параметрами и свойствами пород. В исследованиях автора по потенциально уранопродуктивным интервалам были объединены ансамблевые модели (в том числе Random Forest) с последующим геостатистическим согласованием прогнозов по профилям, а также была учтена реальная разведочная ситуация, когда подтвержденность продуктивности задана выборочно и неравномерно, а значительная часть разреза остается фоном неопределенного статуса (в терминологии ML это соответствует постановке “обучение по подтвержденным объектам при неопределенном фоне”, PU learning). Тем самым была продемонстрирована практическая возможность корректного учета неравномерной подтвержденности интервалов и повышения воспроизводимости прогнозных разрезов. Для условий Шу-Сарысуской урановой провинции были получены зависимости, связывающие электрические параметры каротажа с оценками фильтрационных характеристик песчаников, что позволило сузить диапазон неопределенности при оценке проницаемости рудовмещающих пород (подробнее в разд. 1.3).

1.1.3 Региональная геоинформатика и Mineral Prospectivity Mapping

Mineral Prospectivity Mapping (MPM) представляет собой подход к региональному прогнозированию и картированию перспективности, в рамках которого разнородные геопространственные данные интегрируются для получения индекса перспективности каждой точки территории относительно заданного типа минерализации [11, 12]. При этом могут использоваться как знание-ориентированные решения, где вклад факторов задается на основе геологической модели и экспертных представлений, так и данно-ориентированные решения, где вклад факторов оценивается по распределению известных проявлений и месторождений. Независимо от выбранной методологии, ключевым условием эффективности MPM остается качество признаков: должны использоваться геофакторы, геологически обоснованно связанные с рудоконтролем и минералообразующими процессами [11]. В прикладной интерпретации к таким геофакторам обычно относятся структурные элементы (разломы, зоны пересечений, контакты), индикаторы магмато-гидротермальной активности (тепловой поток, флюидопроницаемые зоны), а также производные потенциальных полей, выделяющие границы блоков и контрасты физических свойств.

Наборы признаков в MPM обычно включают геологические, структурные, геофизические и геохимические компоненты. Геологические признаки отражают литологию, стратиграфию и показатели покрова, включая степень обнаженности, мощность осадочного чехла и т.п. Структурные признаки характеризуют тектонический каркас: положение разломов, узлов пересечений, плотность линейных элементов и расстояния до них; зоны влияния глубинных разломов рассматриваются как возможные каналы миграции флюидов и локализации минерализации [11]. Геофизические признаки представлены картами аномальных полей и их преобразованиями. Для магнитных данных широко используются фильтры, облегчающие структурную интерпретацию: наклонный градиент (tilt derivative) и производные по горизонтальному градиенту позволяют выделять границы источников и линеаменты [13, 14]. Комбинация tilt derivative и карты тотального горизонтального градиента (THD) увеличивает контраст линейных структур и контактов [15, 16]. Аналитический сигнал (амплитуда векторного градиента) формирует экстремумы над центрами намагниченных тел независимо от направления намагниченности и используется для локализации интрузивных тел и потенциально рудоконтролирующих объектов [17]. Геохимические и спектральные признаки включают концентрации рудных и индикаторных элементов, а также мульти- и гиперспектральные показатели альтерации; радиометрические данные (например, канал К как индикатор калийсодержащих минералов) часто интегрируются с геохимией как самостоятельная группа признаков [11].

Ключевой методологической проблемой регионального прогноза является пространственная зависимость данных, обусловленная автокорреляцией признаков и кластеризацией месторождений и проявлений.

Близко расположенные точки характеризуются сходными значениями признаков, поэтому при случайном разбиении выборки обучающая и контрольная части оказываются географически “смешаны”, и показатели качества становятся завышенными [18, 19]. Показано, что при строгой пространственно раздельной проверке, когда обучение и контроль отделяются блоками или субрегионами, многие модели демонстрируют снижение показателей, что отражает реальную переносимость на новые районы [18]. Для получения честной оценки используются схемы пространственно независимой проверки: block cross-validation, leave-one-area-out и родственные варианты, где единицей разделения выступает рудный узел, лист карты или иной геологически сопоставимый субрегион; размер блоков выбирается не меньше характерного диапазона автокорреляции признаков, который может оцениваться по вариограммному анализу. В условиях выраженной редкости объектов и дисбаланса акцент переносится на метрики, чувствительные к ложноположительным срабатываниям: PR-кривую и Average Precision, а также на прикладные показатели “доля известных объектов в верхней части ранжирования” ($\text{precision}@K$ как доля известных объектов в верхних K процентах территории по индексу перспективности) [12, 20].

Кроме способности к ранжированию важна согласованность вероятностной шкалы прогноза с фактическими частотами подтверждений. Выходы многих алгоритмов часто оказываются “некалиброванными” и требуют согласования по независимым данным. Для этого применяются стандартные процедуры: сигмоидальная калибровка (Platt scaling) и изотоническая регрессия; для нейросетевых моделей используется температурная калибровка (temperature scaling) [21-23]. Качество согласования вероятностей оценивается по Brier score и ESE, а подстройка выполняется на независимой проверке или по сквозным предсказаниям, исключаящим утечку информации [21, 22].

Отдельной проблемой является корректность интерпретации прогноза при переносе на зоны, не представленные в обучающем материале. Для этого вводится контроль области применимости модели - область геологической аналогии (Area of Applicability, AOA), которая определяет, где условия по признакам сопоставимы с обучающими аналогами, а где начинается экстраполяция. На практике используется индекс различия (DI), оценивающий удаленность новой точки от обучающих данных в пространстве признаков; выше порога прогноз маркируется как относящийся к зоне экстраполяции и не рассматривается как надежный [24]. Маскирование карт вне области аналогии обеспечивает корректную интерпретацию результатов и снижает риск ложных “аномалий”, вызванных отсутствием аналогов в данных [24].

Интерпретируемость прогнозов рассматривается как обязательное условие их геологической применимости, поскольку она позволяет увязать результат ранжирования с исходными факторами и формировать проверяемые гипотезы. В современной литературе широко применяются методы анализа вклада факторов для нелинейных моделей (SHAP) и графики зависимости прогноза от факторов (PDP/ICE), которые используются как инструмент

объяснения и контроля согласованности прогнозных зависимостей с геологическими представлениями [25-27]. Дополнительно для управленческих решений требуется оценка неопределенности, поскольку ранжирование целей должно учитывать не только величину индекса перспективности, но и устойчивость прогноза. Ансамблевые схемы и анализ разброса предсказаний в сочетании с калиброванными вероятностями и маской АОА позволяют формировать приоритеты по принципу “высокая перспективность при приемлемой неопределенности”, что соответствует риск-ориентированной постановке задач MPM [12].

В геологоразведочном смысле перечисленные элементы методологии решают четыре практические проблемы интерпретации прогнозных карт. Во-первых, PU-learning отражает реальную асимметрию знаний: наличие выявленных проявлений и месторождений фиксируется надежно, тогда как “фон” не является доказанным отсутствием минерализации и может включать скрытые объекты, не обнаруженные из-за покрова, недостаточной изученности или слабой выраженности геосигналов. Во-вторых, контроль области применимости модели (АОА) вводится как количественный аналог границы допустимой геологической экстраполяции: прогноз считается обоснованным лишь там, где сочетание геофакторов сопоставимо с обучающими условиями, а в иных доменах модель ограничивается во избежание ложных “горячих зон”. В-третьих, калибровка переводит выход модели в интерпретируемую вероятностную шкалу, что позволяет трактовать значения как оценку шанса обнаружения целевого типа минерализации при заданных данных и использовать их в планировании работ. В-четвертых, интерпретируемость (SHAP, PDP/ICE) применяется как средство геологического контроля: проверяется, согласуются ли выявленные моделью зависимости с минерал-системной моделью и ожидаемыми рудоконтролирующими факторами, а также формируются проверяемые гипотезы для последующей верификации.

1.2 Алгоритмические основы: обучение с PU-данными, пространственная валидация, калибровка, АОА, интерпретируемость

1.2.1 Обучение с позитивными и неразмеченными данными (PU-learning)

В задачах геологоразведочного прогноза типичной является ситуация, когда достоверно подтверждены только “положительные” объекты - например, точки известных проявлений и месторождений или подтвержденные интервалы потенциальной продуктивности в скважинах, тогда как остальная территория или разрез представляют собой фон неопределенного статуса. Такая постановка соответствует задаче Positive-Unlabeled learning (PU learning), в которой множество P включает подтвержденные объекты, а множество U включает все остальные объекты, среди которых могут присутствовать как истинно “фоновые”, так и невыявленные объекты того же типа [28]. Формально вместо полного разбиения на “наличие - отсутствие” задаются P и $U = All \setminus P$, при этом предполагается, что ненулевая доля объектов положительного класса может

быть скрыта внутри U . В терминах геологии это означает, что отсутствие фиксации проявления или рудного тела не интерпретируется как доказанное отсутствие минерализации, а рассматривается как неопределенность наблюдения. Поэтому “фон” в PU-постановке трактуется как смесь истинно неблагоприятных участков и потенциально рудоносных зон, скрытых факторами покрова, глубины или недостаточной изученности.

Классическое допущение SCAR (Selected Completely At Random) означает, что вероятность попадания положительного объекта в подтвержденный набор P не зависит от его признаков и является постоянной β [28,29]. При выполнении SCAR задача может быть приведена к взвешенной бинарной классификации после оценки β или общей доли положительного класса $\pi = P(y=1)$ [29]. Однако в геологических данных чаще реализуется ситуация, когда подтвержденность зависит от условий изученности и геологических предпосылок: вероятность обнаружения определяется доступностью района, интенсивностью работ, глубиной покрова и выраженностью геосигналов. Это соответствует более реалистичному сценарию SAR (Selected At Random), при котором подтвержденные “плюсы” образуют смещенную подвыборку всех потенциальных объектов [28]. В таком случае требуется осторожная интерпретация вероятностных выводов, поскольку без явного учета механизма отбора меток прогноз может отражать не “истинную вероятность наличия”, а “сходство с подтвержденными аналогами” при заданной структуре выборки [28,29].

В литературе рассматриваются несколько семейств методов PU. Во-первых, используются двухшаговые схемы, где из U выделяются объекты, крайне маловероятные как “позитивы”, и формируется набор “надежного фона” для обучения. Во-вторых, применяются методы оценки π с последующим приведением исходных скорингов к вероятностной шкале с учетом π [29]. В-третьих, развиты методы на основе непредвзятой оценки риска, где функционал качества выражается через наблюдаемые величины без явных меток U ; распространенным решением является $npPU$ (неотрицательная оценка риска), предотвращающая вырождение при гибких моделях [30,31]. В современных модификациях дополнительно используются приемы расширения представительности фона и уменьшения смещения, что может быть особенно важно при неоднородной изученности территорий и выраженной неравномерности геоданных [32].

Практические рекомендации для геологоразведочных данных заключаются в необходимости учитывать механизм формирования подтвержденных объектов, особенно если допущение SCAR не выполняется. Предметные знания должны использоваться для корректного формирования фона, выбора признаков и проверки устойчивости результатов. Дополнительно рекомендуется оценивать чувствительность к дисбалансу и проводить проверки устойчивости на модельных или синтетических сценариях, приближенных по структуре к реальным данным, чтобы контролировать корректность восстановления π и стабильность вероятностной шкалы [28,32]. При этом

вероятностные выходы PU-моделей целесообразно трактовать как риск-ориентированный индекс перспективности или вероятностную оценку при заданных предположениях, а не как безусловное утверждение наличия минерализации [28,29].

1.2.2 Пространственно независимая валидация (spatially independent validation)

Пространственная зависимость данных предъявляет особые требования к процедурам проверки качества. При случайном разбиении близко расположенные точки попадают одновременно в обучение и контроль, что ведет к переоценке качества вследствие повторения сходных геосигналов и пространственной кластеризации подтвержденных объектов [33, 34]. Для получения честной оценки применяются схемы пространственной перекрестной проверки: блоковая k-fold, исключение целых субрегионов (leave-one-area-out), а также буферизация вокруг подтвержденных объектов, предотвращающая “подсказку” в их окрестности [33, 35]. Размер блоков выбирается с учетом диапазона автокорреляции признаков или остатков, который может быть оценен по вариограммному анализу, чтобы минимизировать зависимость между фолдами [33]. В условиях дисбаланса предпочтение отдается PR-кривой и Average Precision, а также метрикам ранжирования, отражающим долю подтвержденных объектов в верхней части списка приоритетов (precision@K), поскольку ROC AUC может оставаться высокой даже при значительном числе ложноположительных зон [34, 36]. Переход от случайной к пространственно раздельной проверке закономерно снижает метрики, однако именно это снижение отражает реальную переносимость модели на новые районы, что является ключевым для геологоразведочной интерпретации [34, 35]. В дальнейшем сквозные предсказания, полученные в режиме “вне обучения”, используются для согласования вероятностной шкалы и построения калибровочных кривых без утечки информации.

1.2.3 Калибровка вероятностей и надежности прогнозов

Большинство алгоритмов интеллектуального анализа формируют скоринговые индексы или “сырые вероятности”, которые не обязаны совпадать с эмпирическими долями подтверждений. Для приведения прогнозов к вероятностной интерпретации применяется посткалибровка: Platt scaling и изотоническая регрессия являются базовыми решениями, а для нейросетевых моделей эффективна температурная калибровка (temperature scaling), которая уменьшает переуверенность без изменения ранжирования [37-39]. Качество согласования оценивается по Brier score и Expected Calibration Error, отражающим расхождение между предсказанной вероятностью и фактической долей подтверждений в группах прогнозов [38, 39]. Методически принципиально исключается утечка: параметры калибровки подбираются на

независимой проверке или по сквозным предсказаниям, после чего фиксируются и применяются к тестовым данным [38]. Современные расширения включают Bayesian Binning into Quantiles и beta calibration, позволяющие улучшать согласование при сложной форме калибровочной зависимости [40, 41].

1.2.4 Контроль области применимости (АОА)

Контроль области применимости (Area of Applicability, АОА) позволяет выделять зоны экстраполяции, где в пространстве признаков отсутствуют аналоги обучающих данных. Используется индекс различия (DI), оценивающий удаленность новой точки от обучающего множества с учетом весов признаков, связанных с их вкладом в модель; порог DI выбирается по кросс-валидационным остаткам и задает границу “внутри - вне” области геологической аналогии [42]. Маскирование зон экстраполяции повышает корректность интерпретации карт перспективности и предотвращает переоценку прогнозов в геологически непредставленных условиях [42]. В сочетании с согласованными вероятностями и картами неопределенности ансамбля формируется риск-ориентированная схема приоритизации целей: при равных значениях индекса предпочтение отдается зонам с высокой перспективностью, меньшей неопределенностью и расположенным внутри области аналогии [36, 43].

1.2.5 Интерпретируемость моделей и учёт неопределённости

Интерпретация прогнозных моделей рассматривается как необходимое условие их применения в геонауках, поскольку позволяет увязать расчетный индекс перспективности с геологическими факторами и обеспечить экспертный контроль корректности зависимостей. Анализ вклада факторов может выполняться как встроенными оценками важности признаков (например, для ансамблей деревьев), так и модельно-агностическими методами. SHAP-разложения обеспечивают аддитивные вклады признаков на уровне отдельных объектов и устойчивые оценки глобальной важности для нелинейных моделей [44]. Графики частичной зависимости (PDP) и индивидуальные условные ожидания (ICE) визуализируют характер влияния признаков, выявляют нелинейности и возможные взаимодействия [45]. Показано, что применение интерпретируемого анализа способствует формированию проверяемых геологических гипотез и повышает научную ценность прогнозных решений [43].

Количественная оценка неопределенности дополняет интерпретацию и необходима для управленческих решений в задачах картирования перспективности. Используются ансамбли моделей и разброс предсказаний, энтропийные меры вероятностного выхода и карты стандартного отклонения по реализациям. Для задач MPM зоны повышенной неопределенности часто совпадают с областью экстраполяции (вне АОА) и должны учитываться при

формировании перечня приоритетных участков, поскольку высокий индекс при высокой неопределенности является менее надежным основанием для выбора цели [49]. Таким образом, интерпретация вклада факторов, оценка неопределенности, пространственно раздельная проверка качества и согласование вероятностной шкалы образуют единый методический каркас, соответствующий современным рекомендациям литературы 2018-2025 гг. [50].

1.3 Петрофизическая база и геологические предпосылки (электрические методы, плотностно-магнитные свойства, структурные индикаторы)

Электрические методы геофизического каротажа (сопротивления, индукционный каротаж, самопроизвольная поляризация - СП) относятся к наиболее информативным комплексам ГИС (каротаж) для выделения литологического разреза, оценки насыщения и косвенной характеристики фильтрационно-емкостных свойств. Физическая основа связана с переносом электрического тока через породу и регистрацией электрохимических потенциалов, формирующихся между пластами и скважинной жидкостью; на показания влияют пористость, минерализация поровой воды, глинистая проводимость, температура и геометрия порового пространства [51].

Ключевая количественная связь задается законом Арчи в форме

$$R_0 = a * R_w * \varphi^{-m} \quad (1.2)$$

где R_0 - сопротивление полностью водонасыщенной породы, R_w - сопротивление пластовой воды, φ - пористость, m - цементационная экспонента, a - эмпирический коэффициент [52,53]. Для практических разрезов рассматриваются обобщения, учитывающие многофазное насыщение и глинистые коллекторы; модель Waxman-Smits вводит вклад связанной воды и обменных катионов, что критично для глинистых толщ и глинистых песчаников [54, 55]. В прикладной интерпретации используются поправки, таблицы и эмпирические корректировки, необходимые для учета литологических эффектов и условий измерений [56, 51].

Метод ПС регистрирует естественные электрохимические потенциалы и применяется для выделения проницаемых пластов при контрасте минерализации; амплитуда ПС зависит от диффузионных и фильтрационных потенциалов и чувствительна к глинистости и минерализации. В практической обработке учитываются эффекты инвазии, диаметра ствола, температурные поправки и сглаживание логов; сопротивления могут приводиться к стандартной температуре и сопоставляться с лабораторными данными по керну [51].

Для задач прогнозирования потенциально уранопродуктивных интервалов комплекс ГИС (каротаж) включает гамма-каротаж, фиксирующий радиоактивный фон, и электрические методы, отражающие минерализацию поровых вод и особенности коллекторов. При построении прогнозных моделей по ГИС физически осмысленные производные показатели (градиенты по глубине, скользящие оценки, индексы, комбинирующие ГК, ПС и сопротивления) часто повышают устойчивость и геологическую

интерпретируемость решения, поскольку фактически расширяют классические петрофизические зависимости на многомерный режим анализа [51,56].

Плотностные и магнитные свойства пород формируют региональные аномалии гравитационного и магнитного полей: плотность изменяется от порядка 1.8 г/см³ для рыхлых осадков до 3.3-3.5 г/см³ для базальтов и ультрамафитов, а магнитная восприимчивость варьирует от десятков 10⁻⁵ СИ для осадков и кислых магматитов до 1000 * 10⁻⁵ СИ и выше для основных магматитов и руд с магнетитом, что обуславливает выраженные контрасты аномальных полей [51]. Для магнитных данных применяются RTP-преобразование, производные и модуль горизонтального градиента (THD) для выделения границ источников, а наклонный градиент (tilt derivative) используется для дешифровки контактов и линейных структур с удобной пороговой интерпретацией [57, 58]. Аналитический сигнал (амплитуда вектора градиента) формирует экстремумы над центрами намагниченных тел и удобен при локализации интрузий и потенциально рудоконтролирующих объектов, поскольку слабо зависит от направления намагниченности [59].

Структурные факторы рассматриваются как один из ведущих контролирующих элементов минерализации: разломы выступают каналами флюидопотока, а узлы пересечений часто связаны с зонами повышенной трещиноватости и проницаемости. Для выделения линейных элементов используются данные ДЗЗ, ЦМР и геофизические поля; выделенные линеаменты трактуются как проекции разломов и структурных зон [60]. В МРМ структурная информация переводится в количественные признаки - расстояния до разломов и узлов, плотность разломов в окрестности, индикаторы принадлежности к зонам разломов, что усиливает геофизические и геохимические сигналы и обычно входит в число наиболее информативных факторов [50].

В целом методическая рамка может быть сведена к следующему: для минералогических массивов при дефиците экспертной атрибуции оправдано выделение статистически нетипичных проб с последующей типизацией ассоциаций; для скважинных данных ключевую роль играют физически осмысленные производные по каротажным кривым и корректная проверка по независимым скважинам; для регионального картирования перспективности требуется интеграция геологии, структуры, геофизики и ДЗЗ с обязательной пространственно раздельной проверкой, согласованием вероятностной шкалы и контролем области геологической аналогии. Такой “геологически осмысленный” подход обеспечивает одновременно воспроизводимость, интерпретируемость и снижение риска методических переоценок при переносе решения на новые участки.

2. МИНЕРАЛОГИЧЕСКАЯ АНАЛИТИКА РЕДКИХ И АНОМАЛЬНЫХ МИНЕРАЛОВ

2.1 Данные, предобработка, понижение размерности

В данном подразделе излагаются состав исходной минералогической выборки, последовательность предобработки и обоснование понижения размерности, с акцентом на интерпретируемость результатов в терминах минералогии, петрофизики и ожидаемых геофизических контрастов. Принципиально важно, что выбранный дизайн признакового пространства был ориентирован не только на статистическое разделение образцов, но и на обеспечение физического смысла выделяемых ключевых факторов, что критично при последующем использовании аномально-ориентированных алгоритмов и геологической типизации редких объектов.

Исходные данные. Исходный набор признаков включал физические, оптические, химико-структурные и кристаллографические характеристики минералов, необходимые для корректной интерпретации аномалий как “нетипичных” сочетаний свойств, а не как случайных шумовых отклонений. К физическим признакам были отнесены твердость по Моосу, диафанитность, показатель преломления, удельный вес, молярная масса, молярный объем и вычисленная плотность. Химико-структурный блок включал кристаллографическую систему, количественные показатели содержания ключевых элементов в формульной единице (например, O, Si, Fe) и бинарные индикаторы наличия анионных групп (например, CO_3^{2-} , PO_4^{3-}). Подобная композиция признаков была выбрана исходя из того, что именно эти параметры обеспечивают связку между лабораторной минералогией и полевыми геофизическими наблюдениями. Так, удельный вес, молярная масса и вычисленная плотность непосредственно связаны с плотностными контрастами горных пород и минералов, которые определяют вклад в гравитационные аномалии и, в более общем виде, в вариации аномалий Буге. Одновременно содержание железа, особенности кристаллохимии и тип кристаллографической системы потенциально связаны с вариациями магнитной восприимчивости и, следовательно, с выраженностью магнитных аномалий и их производных, особенно в случаях присутствия ферромагнитных фаз и их вторичных преобразований.

Для обеспечения воспроизводимости и прозрачности дальнейших вычислительных процедур использовалась сводная спецификация признаков, типов и шкал кодирования (Таблица 2.1). Следует подчеркнуть, что включение категориально закодированных переменных (кристаллографическая система, оптические свойства) было методически оправдано тем, что эти признаки несут самостоятельную минералогическую информацию и позволяют различать семейства минералов, которые при близких массовых характеристиках могут существенно отличаться по структуре и оптическому поведению, что отражает различия в кристаллохимии и генетических условиях.

Таблица 2.1 - Ключевые признаки минералогической выборки и используемые шкалы кодирования.

Характеристика	Описание	Код / Шкала
Кристаллическая структура	Классификация по кристаллографической системе	1 = Триклинная, 2 = Моноклинная, 3 = Орторомбическая, 4 = Тетрагональная, 5 = Гексагональная, 6 = Тригональная, 7 = Кубическая, 8 = Аморфная
Твердость по Моосу	Твердость по шкале Мооса	От 1 до 10
Удельный вес	Отношение плотности минерала к плотности воды	Числовое значение (например, 2.65 для кварца)
Оптические свойства	Поведение света при прохождении через минерал	1 = Анизотропный, 2 = Изотропный, 3 = Униаксиальный, 4 = Биаксиальный
Показатель преломления	Показатель преломления минерала	Числовое значение
Кислород (O)	Число атомов кислорода в формульной единице	Числовое значение (например, 2 для кварца)
Кремний (Si)	Число атомов кремния в формульной единице	Числовое значение
Железо (Fe)	Число атомов железа в формульной единице	Числовое значение
Молярная масса	Молекулярная масса минерала	Числовое значение (например, 60.08 для кварца)
Молярный объем	Объем, занимаемый одним молем минерала	Числовое значение
Карбонат (CO ₃ ²⁻)	Наличие карбонат-ионов	1 если присутствует, 0 если отсутствует
Фосфат (PO ₄ ³⁻)	Наличие фосфат-ионов	1 если присутствует, 0 если отсутствует

Предобработка и масштабирование. Разнородность шкал исходных признаков (непрерывные физические величины, дискретные структурные индикаторы, бинарные признаки анионных групп) устранялась стандартизацией, при которой каждое значение центрировалось относительно среднего и нормировалось на стандартное отклонение. Этот шаг был необходим для исключения доминирования признаков с большим численным диапазоном и для обеспечения того, чтобы метрики близости, используемые далее при понижении размерности и в алгоритмах детектирования аномалий, отражали согласованный вклад всех классов признаков. Примененная стандартизация записывается стандартным выражением:

$$z = \frac{x - \mu}{\sigma} \quad (2.1)$$

где μ и σ соответствуют среднему и стандартному отклонению признака, рассчитанным по исходной выборке. Такая нормировка повышает устойчивость к масштабирующим эффектам и обеспечивает сопоставимость плотностных, оптических и категориально кодированных структурных признаков, что особенно важно при последующем использовании PCA и моделей, чувствительных к метрике расстояния [61].

Понижение размерности методом главных компонент. Для уменьшения мультиколлинеарности и перехода к ортогональным направлениям максимальной изменчивости был использован метод главных компонент. В минералогических данных наличие коррелированных блоков признаков является типичным: например, удельный вес, вычисленная плотность и молярная масса нередко демонстрируют согласованные изменения, отражая вариации состава и упаковки структуры, а некоторые структурные индикаторы могут быть связаны с определенными химическими группами. В таких условиях PCA позволяет перейти от исходного пространства признаков к компактному набору композитов, которые легче интерпретируются как латентные “факторы” минералогической изменчивости. Важно, что PCA в данной работе рассматривался не как исключительно технический шаг для ускорения вычислений, а как способ выделить взаимосогласованные комбинации признаков, которые затем могли быть трактованы в терминах петрофизических и кристаллохимических контрастов, релевантных геологической интерпретации.

Выбор числа компонент обосновывался графически по ломаной Кэттелла (Рисунок 2.1), где представлены индивидуальные и накопленные доли объясненной дисперсии. Для двумерной визуализации аномалий использовались первые две компоненты как наиболее информативная проекция, тогда как для последующего моделирования сохранялся ограниченный набор компонент, обеспечивающий достижение порога накопленной дисперсии 90-95 % [62]. Такая стратегия обеспечивала баланс между сохранением геологически значимой информации и снижением размерности, что уменьшает риск переобучения и повышает стабильность алгоритмов выявления аномалий.

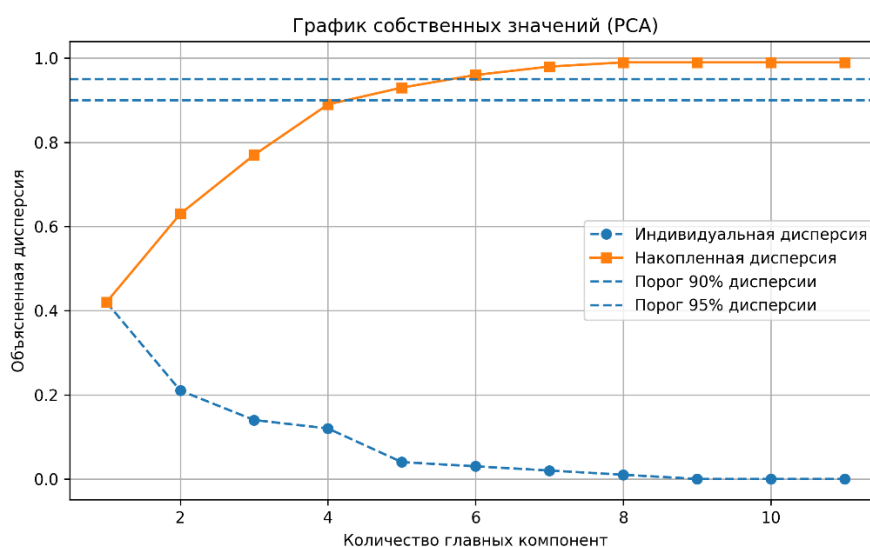


Рисунок 2.1 - Ломаная Кэттелла для главных компонент: индивидуальная и накопленная доли дисперсии; порог выделения компонент по накопленной дисперсии 90-95 %.

Математические соотношения. Пусть X - матрица стандартизированных признаков размера $n \times p$, где n - число образцов, p - число признаков. Тогда ковариационная матрица определяется как:

$$\Sigma = \frac{1}{n-1} X^T X, \quad (2.2)$$

собственное уравнение, задающее ранжирование компонент по собственным значениям (см. (2.3)):

$$\Sigma \vartheta = \lambda \vartheta, \quad (2.3)$$

и линейная проекция исходных признаков в пространство главных компонент (см. (2.4)):

$$Z = XW \quad (2.4)$$

где W - матрица, составленная из выбранных собственных векторов (нагрузок), а Z - координаты образцов в пространстве главных компонент. Приведение этих выражений является необходимым, поскольку именно через нагрузки W обеспечивается интерпретация компонент как сочетаний исходных минералогических и петрофизически значимых признаков, а также обеспечивается корректное сопоставление выделенных латентных факторов с физическими ожиданиями (плотностные контрасты, потенциальная связь с магнитными свойствами, структурно-оптические различия).

Геолого-петрофизическая интерпретация пониженного пространства. Полученные главные компоненты интерпретировались как согласованные факторы признаков, отражающие совместную вариабельность физических и химико-структурных характеристик. В прикладном смысле это означало, что “плотностно-химический” и “оптико-структурный” композиты позволяли перевести статистически выделенные направления изменчивости в интерпретационные оси, релевантные геологическим процессам и потенциальным геофизическим откликам. Повышенные значения компоненты, связанной с плотностными прокси, концептуально согласуются с минералогическими ассоциациями, способными создавать локальные положительные плотностные контрасты, что потенциально коррелирует с аномалиями Буге на соответствующих масштабах. Одновременно вариации, связанные с кристаллохимией, кристаллографической системой и присутствием железа, рассматривались как потенциальные индикаторы различий в магнитной восприимчивости и, следовательно, различий в структурной организации магнитного поля и его производных. Важно, что такая трактовка не подменяет прямое измерение петрофизических параметров, но обеспечивает объяснимый мост между минералогическими данными и ожидаемыми геофизическими сигнатурами, повышая интерпретируемость последующего выделения аномалий [61, 62].

2.2 Детектирование аномалий (Isolation Forest, One-Class SVM), группировка и геологическая типизация

Постановка задачи. рамках выявления редких и аномальных минералов задача рассматривалась как одно-классовая: подавляющая часть минералов формирует распределение “нормы”, тогда как редкие минералы или геологически необычные сочетания свойств проявляются как статистически нетипичные объекты в многомерном пространстве признаков. Данная постановка методически оправдана, поскольку отрицательный класс в минералогических данных, как правило, не формируется в явном виде, а интерес направлен на обнаружение “необычных” объектов, потенциально связанных с редкими генетическими обстановками или редкими ассоциациями. Для повышения надежности применялись два комплементарных алгоритма, различающиеся по механизму выявления нетипичности: Isolation Forest (IF) и One-Class Support Vector Machine (OC-SVM). Совместное применение было обосновано тем, что IF эффективно выделяет “глобальные” выбросы, располагающиеся на периферии плотных облаков наблюдений, тогда как OC-SVM способен выявлять более тонкие отклонения, включая локальные девиации внутри общей структуры данных [63-65].

Isolation Forest: принцип и мера аномальности. Isolation Forest строит ансамбль случайных деревьев изоляции, в которых объект считается более аномальным, если он изолируется меньшим числом случайных разбиений. Иными словами, ожидаемая длина пути $h(x)$ до объекта обратно связана с его “нормальностью”: редкие и нетипичные объекты отделяются быстро, а типичные требуют большего числа разбиений [63]. Для интерпретации использовалась стандартная мера аномальности $S(x, n)$, зависящая от ожидаемой длины пути и нормирующего коэффициента $c(n)$, соответствующего среднему пути в случайном дереве при размере выборки n :

$$S(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (2.5)$$

где $E[h(x)]$ – ожидаемая длина пути в ансамбле, $c(n)$ - нормировка. Значения $S(x, n)$, близкие к 1, интерпретировались как высокая аномальность; значения около 0.5 трактовались как пограничные; значения ниже 0.5 соответствовали области “нормы” [63]. Подчеркнем, что данная шкала удобна именно для ранжирования кандидатов и последующей геологической проверки, поскольку позволяет перейти от бинарного решения к управляемому списку приоритетов.

One-Class SVM: постановка оптимизационной задачи. OC-SVM аппроксимирует границу области, в которой сосредоточено большинство “нормальных” объектов, и относит точки, выходящие за пределы этой области, к аномалиям. Метод реализуется как задача построения разделяющей поверхности в (возможно) высокомерном ядровом пространстве, что позволяет моделировать нелинейные формы распределения нормы [64]. В праймальной форме оптимизационная задача записывается как:

$$\min_{\omega, \xi, \rho} \frac{1}{2} \|\omega\|^2 + \frac{1}{\nu n} \sum_{i=1}^n \xi_i \quad (2.6)$$

где $\varphi(\cdot)$ - отображение в ядровое пространство, $\nu \in (0,1]$ - параметр, задающий верхнюю оценку доли выбросов и одновременно контролирующей жесткость границы, ξ_i - слабые переменные [64]. В контексте минералогических данных данный подход был важен тем, что способен фиксировать “локальные редкости”, то есть нетипичные комбинации признаков, которые могут располагаться не только на периферии, но и внутри общей структуры облака, что дополняет изоляционную логику IF [65].

Визуализация распределения аномалий. Для интерпретации результаты IF и OC-SVM визуализировались в пространстве первых двух главных компонент. Данная проекция использовалась исключительно для наглядной проверки геометрии распределений и взаимного расположения аномалий относительно основной массы минералов. Ожидается, аномалии, выделенные IF, преимущественно локализовались по периферии облака в PCA-пространстве, тогда как аномалии OC-SVM чаще проявлялись также вблизи внутренних областей, отражая чувствительность к локальным отклонениям. Эта разница рассматривалась как эмпирическое подтверждение комплементарности методов и оправдание их совместного применения для формирования расширенного набора кандидатов [65].

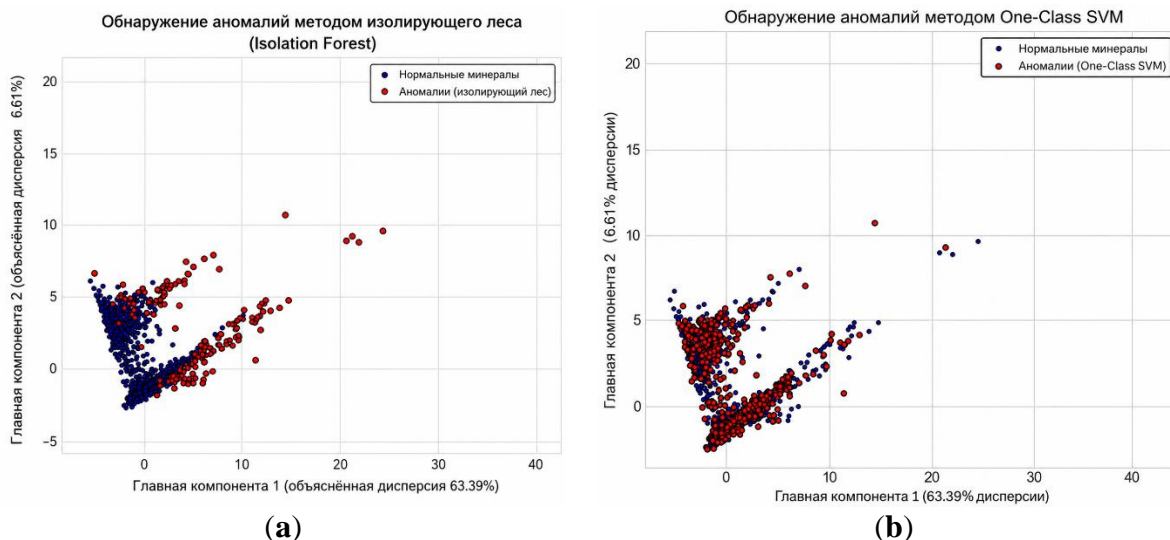


Рисунок 2.2 - Проекция аномалий в пространстве первых двух главных компонент: (a) Isolation Forest, (b) One-Class SVM. Нормальные образцы показаны синим, аномалии - контрастным цветом. Первая компонента объясняет 63.39 %, вторая - 36.61 % суммарной дисперсии. Для обучения моделей использовались пять компонент.

Сравнительная результативность алгоритмов и практическая ценность. Оценивание качества на разметке редких минералов демонстрировало ожидаемый компромисс между “строгостью” и полнотой. IF характеризовался высокой точностью при выделении “жестких” выбросов,

однако полнота была ограниченной, что типично для алгоритмов, ориентированных на крайние значения и периферийные объекты. OC-SVM демонстрировал более высокую полноту при сохранении высокой интегральной дискриминации, что отражает способность выявлять широкий спектр редких и геологически интересных комбинаций признаков, включая неочевидные отклонения. С практической точки зрения это означало, что IF может быть предпочтителен на этапе первичной фильтрации “самых редких” объектов, тогда как OC-SVM целесообразно использовать для расширенного скрининга, когда требуется максимально полный пул кандидатов для экспертной проверки и дальнейшей типизации [63-65]. Численные значения метрик приводятся в Таблице 2.2, где представлены Precision, Recall, F1-score и AUROC.

Таблица 2.2. Оценка качества моделей детектирования аномалий: точность (Precision), полнота (Recall), F1-мера и AUROC для Isolation Forest и One-Class SVM.

Модель	Precision	Recall	F1-score	AUROC
Isolation Forest	1.000	0.277	0.433	0.638
One-Class SVM	0.851	1.000	0.919	0.981

Группировка выявленных аномалий (KMeans) и геологическая типизация. Дальнейший анализ был сосредоточен на подмножестве объектов, идентифицированных как аномальные хотя бы одним из методов, то есть IF и OC-SVM. Такой выбор обеспечивал, что кластеризация выполнялась именно в пространстве редких и нетипичных объектов, а не размывалась доминирующей массой типичных минералов. Для изучения внутренней структуры множества аномалий применялась кластеризация KMeans после понижения размерности PCA. Число кластеров выбиралось методом “локтя”, где излом кривой инерции указывал на $k = 3$ как на значение, обеспечивающее баланс между интерпретируемостью и внутрикластерной компактностью. Соответствующие результаты представлены на Рисунке 2.3 (локоть) и Рисунке 2.4 (карта кластеров) [66]. Подчеркнем, что выбор k трактовался не как сугубо статистическая оптимизация, а как выбор гранулярности геологически интерпретируемых типов: слишком малое k приводило бы к слиянию генетически различных групп, тогда как слишком большое k могло бы создавать статистически неустойчивые подкластеры, затрудняющие интерпретацию и практическое использование.

2.3 Проверка устойчивости и практическая интерпретация

В данном подразделе оценивается устойчивость полученных результатов к изменению гиперпараметров, параметров предобработки и условиям выборки, а также демонстрируется практическая интерпретация выделенных групп редких и аномальных минералов с позиций минералогии, геологии и ожидаемых геофизических проявлений. Целью является показать, что

выделенные закономерности обладают не только статистической воспроизводимостью, но и прикладной ценностью для построения осмысленных гипотез и последующего таргетинга лабораторных либо полевых проверок [60, 68].

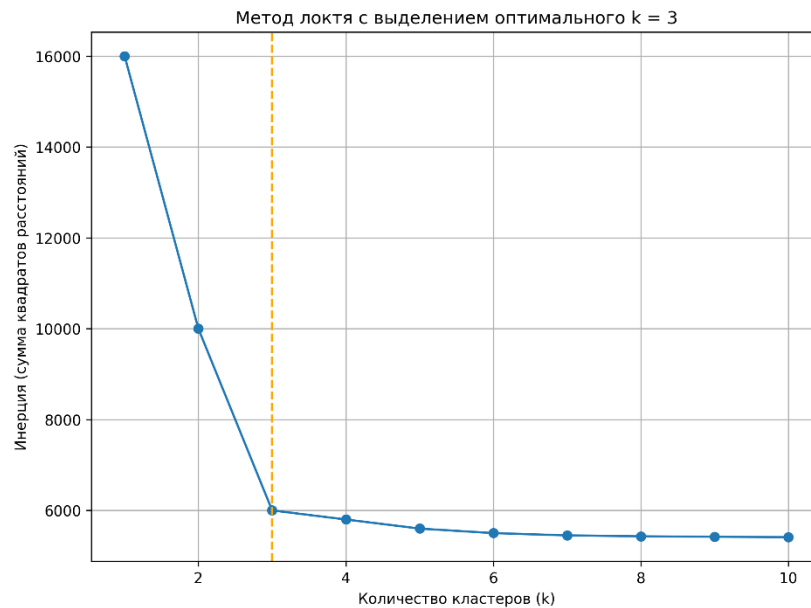


Рисунок 2.3 - Метод «локтя» для выбора числа кластеров в KMeans: излом при $k = 3$ соответствует оптимальному балансу внутрикластерной компактности и числа групп.

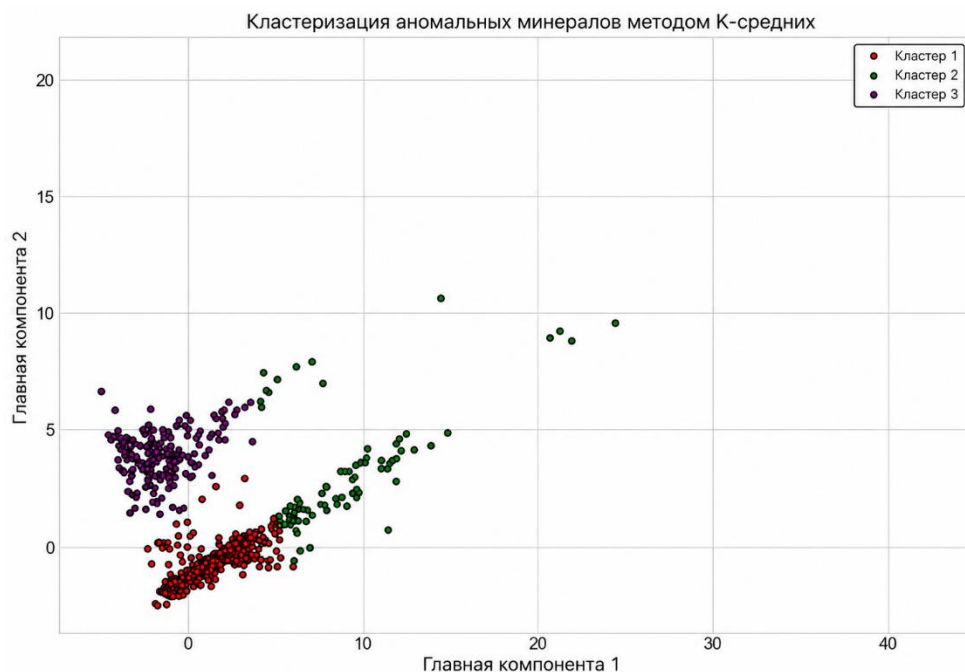


Рисунок 2.4. Кластеризация аномальных минералов методом KMeans после понижения размерности PCA (первые две компоненты суммарно информативны для визуализации). Каждая группа отражает различающиеся геологические процессы формирования минералов.

Геологическая интерпретация кластеров аномальных минералов. Выделенные три группы рассматривались как типологические классы редких объектов, различающиеся по вероятным условиям образования и по ожидаемым физико-химическим контрастам.

Первый кластер интерпретировался как группа, связанная преимущественно с эвпоритовыми и щелочными условиями минералообразования, поскольку в нем доминировали галогениды и сульфаты. Подобные ассоциации типичны для высокосолёных бассейнов испарительного типа и щелочных рассольных систем, где происходит концентрация ионов и последующая кристаллизация соответствующих фаз. В прикладном плане данная группа рассматривалась как индикаторная для оценки перспектив солеродных и рассольных систем и может иметь опосредованную связь с рассольными типами полезных компонентов при наличии соответствующего геологического контекста.

Второй кластер был отнесен к метаморфическим и метасоматическим обстановкам, поскольку он характеризовался преобладанием силикатных фаз и сочетаниями признаков, согласующимися с условиями повышенных давлений, умеренных температур и/или гидротермальной переработки. В практической постановке именно такие группы наиболее релевантны для диагностики зон метасоматической минерализации, часто контролируемых разломными и разломно-сдвиговыми системами, а также для постановки гипотез о возможных гидротермальных путях миграции флюидов.

Третий кластер интерпретировался как отражающий магматические и последующие супергенные процессы, поскольку в нем сочетались признаки первичных высокотемпературных минералов (алюмосиликаты, оксиды) и вторичных фаз выветривания/окисления. Такая смешанная природа рассматривалась как индикатор ступенчатой эволюции минералообразующей системы, в которой первичная магматическая генерация могла сопровождаться последующей переработкой и перераспределением компонентов в зоне гипергенеза.

Петрофизический и геофизический контекст. Кластерная типизация была увязана с физическими свойствами минералов, прежде всего с удельным весом, вычисленной плотностью и оптическими прокси. Повышенная плотность и высокие значения оптических показателей обычно соответствуют присутствию тяжелых катионов и/или плотной упаковке структуры, что согласуется с экстремальными условиями минералообразования (высокие Р-Т, специализированная химическая среда) и объясняет их статистическую редкость в общей массе минералов. С позиции геофизики такие группы рассматриваются как потенциальные индикаторы зон контрастов плотностных и магнитных свойств, а также возможного аномального радиометрического отклика при определенных ассоциациях. Это обеспечивает основу для последующей интеграции лабораторных результатов с данными потенциальных полей и гамма-спектрометрии на этапе пространственного анализа и таргетинга.

Верификация результатов и устойчивость к параметрам. Верификация выполнялась в двух логически связанных контурах. Во-первых, оценивалась согласованность машинного выделения аномалий с консервативным статистическим эталоном на уровне 2σ для ключевых признаков, что позволяло трактовать Precision, Recall, F1-score и AUROC как индикаторы соответствия алгоритмического решения строгому критерию нетипичности [67]. Во-вторых, проводилась геологическая интерпретация аномалий и их групп, что обеспечивало проверку “физического смысла” выделенных объектов и снижало риск того, что модель извлекает артефакты кодирования или масштаба.

Устойчивость структуры данных к вариациям параметров понижения размерности оценивалась по воспроизводимости геометрии облака точек и взаимному положению кластеров при изменении числа главных компонент и при вариациях шага стандартизации. При сохранении доли объясненной дисперсии не ниже 80 % локализация основных групп и относительное разделение кластеров сохранялись, что свидетельствовало о наличии устойчивой латентной структуры данных, а не о доминировании случайных выборочных эффектов [60]. В прикладной интерпретации это означает, что типы аномальных минералов отражают устойчивые вариации состава и структуры, а не результат “хрупкой” настройки.

Чувствительность кластерной структуры проверялась по зависимости от числа кластеров и начальной инициализации центроидов. Оптимальность выбранного k дополнительно интерпретировалась через согласование нескольких критериев: характер излома на кривой внутрикластерной дисперсии, динамика межкластерной делимости и поведение средних значений силуэт-коэффициента. При уменьшении k происходило слияние геологически разных групп и утрата интерпретационной дифференциации, тогда как увеличение k приводило к дроблению однородных наборов на статистически неустойчивые подкластеры, что снижало практическую ценность типизации [64, 72]. Тем самым выбранное значение k соответствовало естественной гранулярности минералого-геохимических типов в множестве аномалий.

Качество детектирования аномалий при дисбалансе и ранжирование кандидатов. В условиях редкости целевых объектов особое значение имеет качество ранжирования в верхней части списка кандидатов. В этой связи, наряду с AUROC, концептуально более информативными считались метрики, чувствительные к ложноположительным в области высоких скорингов, включая PR-ориентированные показатели, поскольку именно верхние ранги определяют эффективность подтверждающих анализов и экономику лабораторной либо полевой верификации [66, 67]. Практический смысл результатов Таблицы 2.2 заключался в том, что IF обеспечивал “консервативный” режим, отбирая небольшое число наиболее контрастных объектов с высокой точностью, тогда как OC-SVM формировал расширенный пул кандидатов с высокой полнотой, что удобно для сценариев, где важнее не

пропустить редкий объект, чем минимизировать число ложноположительных на первом шаге.

Корреляционная структура и интерпретационные связи. Для дополнительной интерпретации использовался анализ корреляций между стандартизированными признаками, позволяющий выявить согласованные блоки свойств, которые могут трактоваться как признаки минеральных ассоциаций или единых петролого-геохимических факторов. Тепловая карта попарных корреляций (Рисунок 2.5) применялась как инструмент контроля: наличие ожидаемых связей между плотностными прокси и химико-структурными параметрами рассматривалось как поддержка физической состоятельности латентных факторов PCA и последующей кластеризации [60, 68].

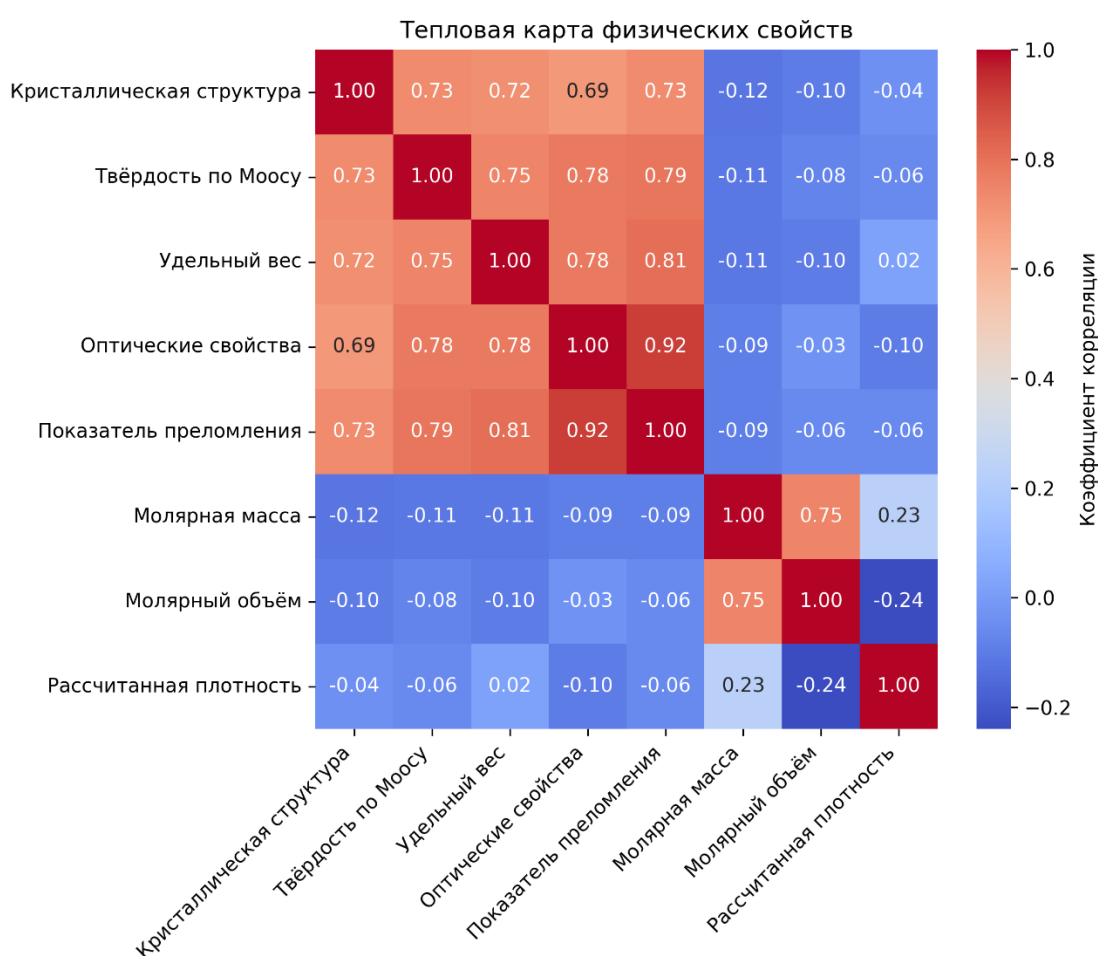


Рисунок 2.5 - Тепловая карта попарных корреляций между стандартизированными минералогическими и геохимическими признаками (коэффициент Пирсона r). Более темные ячейки отражают более сильные положительные корреляции, а отрицательные корреляции показаны контрастными оттенками.

Сходимость результатов двух моделей и двухконтурный пул приоритетов. Согласованность результатов IF и OC-SVM рассматривалась как

дополнительный критерий устойчивости. Пересечение верхней части рейтинга (например, совпадение значимой доли top-N кандидатов) трактовалось как индикатор того, что объекты действительно выделяются как нетипичные независимо от механизма детектирования. Расхождения рассматривались как область повышенной неопределенности, требующая адресной экспертной проверки. В прикладной постановке рекомендуется формировать двухуровневый список кандидатов: приоритет 1 - объекты, подтвержденные обоими алгоритмами и одновременно согласованные с геологической типологией кластера; приоритет 2 - объекты, выделенные только одним методом или демонстрирующие противоречивые сигнатуры, что предполагает необходимость дополнительных проверок.

Анализ стабильности при варьировании гиперпараметров. Дополнительная проверка устойчивости выполнялась путем варьирования доли ожидаемых выбросов ν в OC-SVM и ключевых параметров ансамбля в IF (в частности, числа деревьев). При изменении ν в диапазоне 0.05-0.20 и числа деревьев в диапазоне 200-700 сохранялась высокая сходимость верхней части списка кандидатов и стабильность относительных позиций объектов, что указывает на отсутствие “хрупкого” решения, зависящего от одной настройки, и подтверждает использование устойчивой многомерной структуры признакового пространства [65]. Наблюдаемая стабильность согласуется с тем, что первичные факторы вариабельности (плотностные, химико-структурные и оптические блоки) были выделены корректно и не искажались масштабирующими эффектами благодаря стандартизации.

Перенос результата в практический геологический контекст. Ключевым итогом раздела являлась трансляция статистического ранжирования в осмысленную геологическую логику. Для каждой типологической группы аномальных минералов диагностические наборы свойств сопоставлялись с вероятными сценариями минералообразования и ожидаемыми геофизическими проявлениями. Группы, демонстрирующие повышенные плотностные прокси и признаки, согласующиеся с присутствием Fe-содержащих или потенциально ферромагнитных компонентов, рассматривались как кандидаты, которые в соответствующем геологическом окружении могут быть связаны с источниками локальных гравитационных и магнитных аномалий и потому рационально проверяются через интеграцию с картами производных потенциальных полей. Напротив, группы, ассоциируемые с окислительно-гидротермальной переработкой и демагнетизацией, интерпретировались как потенциальные маркеры гидротермальных систем, для которых более характерны пониженные магнитные свойства и приуроченность к структурным узлам. Такая интерпретация задает мост между лабораторной минералогией и пространственными моделями перспективности, повышая эффективность последующих этапов интеграции с геофизическими наблюдениями и МРМ-уровнем анализа [68].

Вывод по главе. В главе было показано, что сочетание одно-классовых методов детектирования аномалий (Isolation Forest и One-Class SVM) в PCA-

пространстве стандартизированных минералогических признаков с последующей кластерной типизацией обеспечивает устойчивое и геологически интерпретируемое выявление редких и аномальных минералов. Было установлено, что верхняя часть ранжирования кандидатов сохраняется при разумном варьировании ключевых гиперпараметров, выделенные кластеры согласуются с физико-химическими сценариями минералообразования, а типологические группы могут быть напрямую увязаны с ожидаемыми плотностными и магнитными контрастами. Тем самым создается воспроизводимый контур “редкие минералы - интерпретируемая типизация - приоритетные кандидаты для верификации”, который методически пригоден для дальнейшей интеграции с геофизическими данными и для прикладной приоритизации подтверждающих анализов и полевых работ.

3 КАРТИРОВАНИЕ УРАНОПЕРСПЕКТИВНЫХ ГОРИЗОНТОВ ПО ДАННЫМ ГИС: ИНТЕГРАЦИЯ ML И ГЕОСТАТИСТИКИ

3.1 Геолого-геофизическая характеристика Северного рудного узла (Шу-Сарысуская впадина), состав и качество каротажных данных

Северный рудный узел расположен в пределах Шу-Сарысуская впадина, где ураноносные горизонты, как правило, приурочены к песчано-гравелистым коллекторам верхнемелового возраста, сформированным в пределах палеодолинных систем и представленным палеорусловыми (канальными) телами, развитие которых контролируется редокс-границами. В подобных пластово-инфильтрационных системах положение рудных фронтов обычно определяется сочетанием проницаемых коллекторов, по которым мигрируют окисленные ураносодержащие растворы, и низкопроницаемых глинисто-алевролитовых экранов, создающих условия для фиксации урана в восстановительной среде. Такая конфигурация гетерогенных коллекторно-экранирующих пакетов рассматривается как типичная для осадочных урановых систем Центральной Азии и задает вероятностно выраженную природу прогнозных признаков: устойчивые границы минерализации встречаются реже, тогда как чаще формируются зоны переходного характера, в которых продуктивность может проявляться локально и сегментировано [73]. В настоящем исследовании дальнейшая характеристика геологической обстановки, положения палеодолинных систем (контуры врезов) и размещения наблюдательной сети скважин опиралась на детальную схему района работ, позволяющую связать стратиграфию с ожидаемыми особенностями каротажных откликов. Как показано на рисунке 3.1, рудное поле прослеживается в пределах контуров палеодолин, а сеть скважин обеспечивает покрытие ключевых фрагментов палеорусловых коллекторских тел и потенциально благоприятных зон редукции, что является принципиальным для постановки задач прогнозирования ураноносности в вероятностной форме [73].

Комплекс геофизических методов скважинных исследований, применённый в районе работ, в первую очередь отражает физические основы редокс-контроля ураноносности и литолого-гидродинамических контрастов. Гамма-каротаж рассматривается как основной индикатор ураноносных интервалов при условии корректной калибровки и контроля стабильности регистрации, поскольку его повышенные значения могут отражать присутствие урана и продуктов его фиксации. Электрические параметры, включая кажущееся сопротивление и самопроизвольную поляризацию, используются как вспомогательные признаки, поскольку ими фиксируются вариации пористости, насыщения, минерализации пластовых вод и доли глинистого материала, то есть факторов, определяющих коллекторские свойства и положение экранов [74]. В методических руководствах по урановым скважинам подчёркивается необходимость сопоставления геофизических кривых с литологическими данными и соблюдения процедур контроля дрейфа, фильтрации шумов и обеспечения межскважинной сопоставимости [74]. При

интерпретации электрических откликов существенным является петрофизический контекст: параметры, чувствительные к электропроводности и электрохимическим эффектам на контактах фаций, могут косвенно указывать на изменения гидрогеохимического режима и на положение вероятных зон фильтрационного обмена, что потенциально повышает устойчивость разграничения коллекторов и экранов в ураноносных комплексах [75].

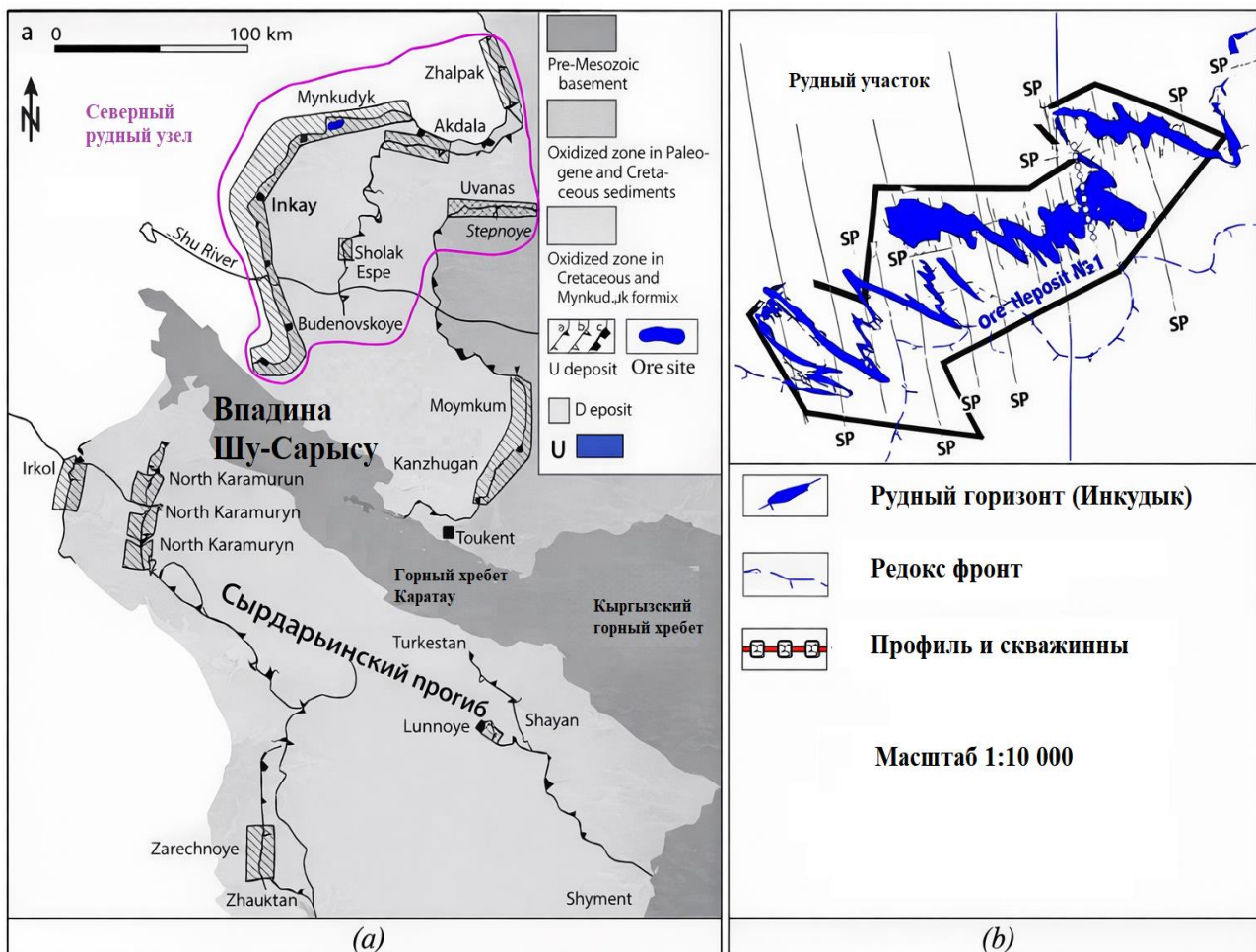


Рисунок 3.1 - Геологическое положение и размещение скважин в пределах Северного рудного узла; показаны контуры палеодолинных систем (древних врезов), положение ураноносных горизонтов и элементы редокс-контроля, а также основные литологические комплексы

Исходный массив данных включает гамма-каротаж, кажущееся электрическое сопротивление и самопроизвольную поляризацию, зарегистрированные с шагом 0.1 м по стволу, а также литологические и технологические атрибуты для стратиграфической привязки. Такая дискретизация обеспечивает высокую вертикальную детальность, однако интерпретироваться она должна с учётом эффективного разрешения методов, определяемого длиной зондирования и сглаживанием отклика в тонких пачках. Перед аналитическими и модельными этапами были выполнены процедуры обеспечения качества данных, включающие проверку диапазонов и единиц

измерения, выявление и корректировку нехарактерных значений, визуальный контроль согласованности кривых в разрезе, а также калибровку гамма-канала по контрольным участкам и межскважинную нормализацию при необходимости. Такая последовательность соответствует рекомендациям по обеспечению воспроизводимости каротажных данных в урановой тематике и направлена на снижение риска внесения систематических смещений на этапе статистического обучения [74]. Дополнительно выполнялась диагностическая проверка согласованности кривых с литологической рамкой и разметкой продуктивности. Как показано на рисунке 3.2, репрезентативный пример каротажных кривых демонстрирует согласованное поведение γ -активности и электрических параметров на границах фаций и в пределах интервалов, где наличие ураноносности может быть предположено по совокупности признаков, что критически важно для корректной постановки задач классификации.

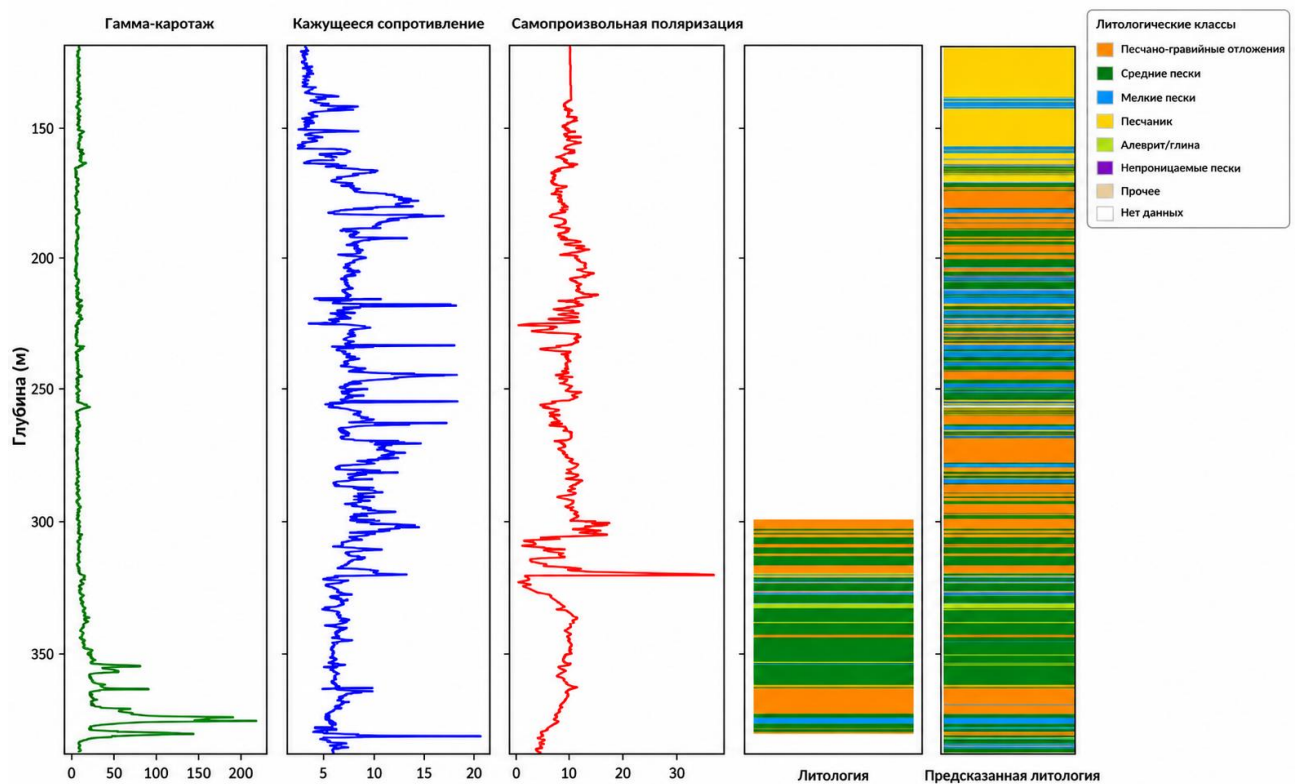


Рисунок 3.2 - Пример каротажных кривых по скважине Well_XX: гамма-каротаж, кажущееся электрическое сопротивление и самопроизвольная поляризация, сопоставленные с фактической и прогнозной литологией. Пустые сегменты «Lithology» соответствуют отсутствию исходной литологической метки.

С физико-геологической точки зрения каротажные отклики в пределах палеорусловых (канальных) песчаников палеодолинных систем могут быть объяснены сочетанием гранулометрии, минерального состава и гидрогеохимического режима. Повышенные значения γ -канала интерпретируются как индикаторы концентрации урана и/или сопутствующих

радионуклидов, однако в переходных условиях они также могут отражать повышенную глинистость, что требует осторожного различения потенциально продуктивных интервалов и литологически обусловленного радиоактивного фона. Пониженные значения кажущегося сопротивления могут быть связаны с ростом минерализации поровых вод, увеличением доли глинистой фазы либо сочетанием обоих факторов, что приводит к неоднозначности интерпретации без привлечения дополнительных признаков. Тренды ПС чувствительны к электрохимическим эффектам на контактах фаций и градиентам минерализации, поэтому с их помощью могут уточняться вероятные границы фациальных переходов и зоны активного фильтрационного обмена. Учитывая известную связь электрических и фильтрационно-емкостных свойств песчаников, совместное использование сопротивления и ПС рассматривается как потенциально полезное для стратиграфической корреляции палеорусловых коллекторских тел и для косвенной оценки проницаемости в рамках гидрогеологической модели ураноносной системы [75]. В совокупности эти соображения определили выбор минимально достаточного набора признаков и обоснование процедур предобработки и нормализации.

Критическим этапом подготовки данных являлась унификация шкал и проверка статистической согласованности измерений перед входом в моделирование. В соответствии с рекомендациями по контролю качества для урановой геологоразведки, выполнялась проверка внутрискважинной согласованности кривых, анализ межскважинных сдвигов и выявление аномальных значений, связанных с инструментальными и технологическими факторами [76]. На рисунке 3.3 представлены диагностические распределения и перекрёстные диаграммы ключевых каротажных признаков, позволяющие выявлять асимметрию, возможные систематические смещения и межскважинные эффекты. Такая визуальная аналитика использовалась как основание для решений о сглаживании, нормализации и локальной фильтрации выбросов, при этом принципиально сохранялись геологически осмысленные сигналы, не сводимые к шуму.

Особое внимание было уделено калибровке, воспроизводимости и сопоставимости каротажных измерений. Международные руководства по урановой тематике фиксируют необходимость регулярных процедур калибровки γ -детекторов и учёта инструментальной стабильности; сопоставимость между партиями приборов и бригадами обеспечивается стандартами и эталонными участками, а также строгой документацией режимов измерений [74]. При отсутствии единого эталона по всей сети скважин ключевую роль приобретает статистический и визуальный контроль распределений и межскважинных сдвигов, поскольку именно систематические смещения способны искусственно улучшать или ухудшать метрики и исказить прогнозируемые вероятности. Для электрических методов практическая интерпретация также выигрывает от учёта связи электрических и гидравлических параметров коллекторов, что повышает обоснованность

разграничения коллекторских и экранирующих толщ и, следовательно, вероятностной связи с ураноносностью [75].

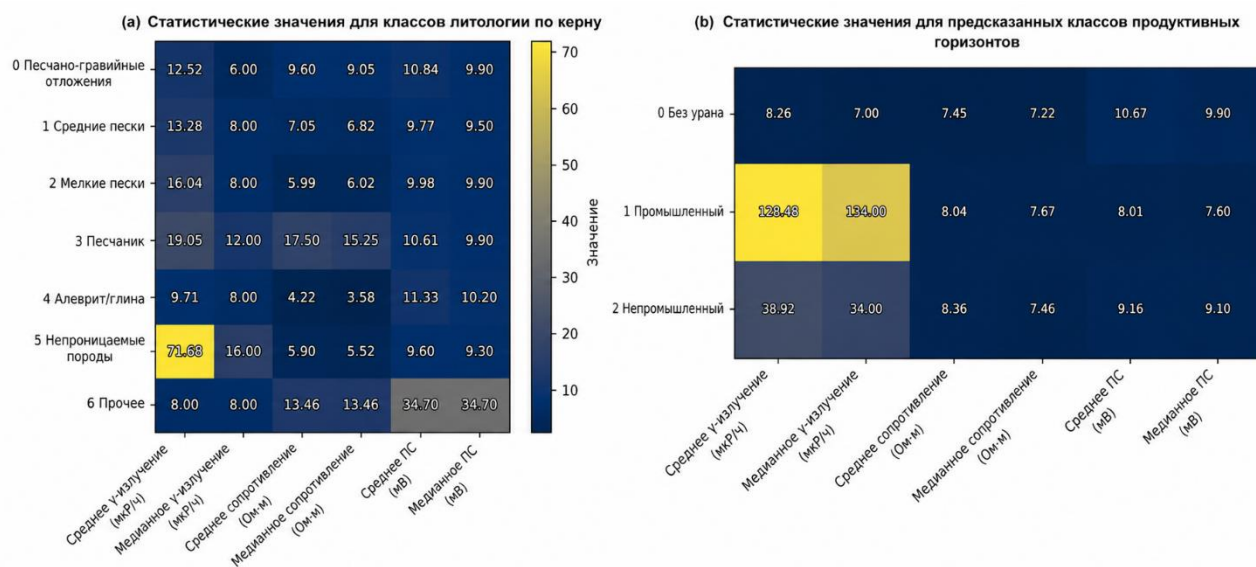


Рисунок 3.3 - Распределения и перекрёстные диаграммы ключевых каротажных признаков (γ в мкР/ч, кажущееся сопротивление в Ом·м, ПС в мВ) для классов литологии и продуктивности; приведены средние и медианные значения для учёта асимметрии и выбросов.

В итоге геологическая постановка задачи, физика регистрируемых сигналов и принципы обеспечения качества данных сформировали базис для последующих шагов моделирования: региональная схема и положение скважин задали стратиграфическую рамку (рисунок 3.1), каротажные разрезы подтвердили диагностический характер γ -активности и электрических параметров (рисунок 3.2), а диагностические распределения показали пригодность массива для статистического обучения при сохранении петрофизической интерпретируемости (рисунок 3.3) [74].

3.2 Модели литологии и продуктивных интервалов (Random Forest, инженерные γ -признаки)

В данном подразделе рассматривается построение предсказательных моделей литологии и продуктивных ураноперспективных интервалов по данным каротажа с целью восстановления литологической структуры разреза и вероятностного выделения ураноносных горизонтов в интервалах, где отсутствует прямое керновое подтверждение. Для пластово-инфильтрационных месторождений урана, где геометрия рудных линз контролируется сочетанием гранулометрии, проницаемости и локальных редокс-границ, такая реконструкция имеет прикладное значение, поскольку качество восстановления разреза влияет на обоснованность проектирования и приоритизации скважин, а также на корректность последующего пространственного моделирования.

В рамках работы решались две взаимосвязанные задачи, различающиеся по целевой переменной. Первая задача была сформулирована как многоклассовая классификация литологии с целью выделения основных типов осадочных толщ, интерпретируемых как коллекторские песчано-гравийные тела, песчаные разности различной granulometрии и низкопроницаемые глинисто-алевролитовые пачки, выполняющие роль экранов. Вторая задача была сформулирована как бинарная классификация продуктивности, где положительным классом задавались интервалы, подтверждённые по керну и геолого-технологической документации как промышленно ураноносные, а отрицательным классом - пустые интервалы без экономически значимых концентраций урана. Такое совместное решение позволило не только восстановить литологический разрез, но и получить вероятностную оценку ураноносности для каждого глубинного шага, что является критичным для дальнейшей геостатистической интерполяции и построения профильных/объёмных представлений.

Устойчивость статистического обучения была обеспечена литологической реклассификацией керновых описаний. Исходная детализация литологий, включавшая ряд редких и неоднородных типов, была агрегирована в ограниченное число укрупнённых классов, отражающих доминирующие петрофизические группы. В схему перекодировки были включены коды 0-6, однако для целей машинного обучения в окончательную обучающую выборку были включены только хорошо представленные и геологически относительно однородные классы 0, 1, 2 и 4. Коды 3, 5 и 6, характеризующиеся меньшей представленностью и/или петрофизической неоднозначностью, были исключены из обучения, но сохранены в общей схеме для последующего геологического анализа. Такая стратегия соответствует практике укрупнения литологических классов в задачах распознавания по данным каротажа, где избыточная детализация приводит к статистически нестабильным оценкам и снижает надёжность предсказаний по редким классам [77, 78]. Схема соответствия между исходными литологическими описаниями, укрупнёнными классами и их геологическим толкованием приведена в таблице 3.1.

Таблица 3.1 – Литологические коды после рекатегоризации.

Категория	Описание	Характеристики
0	Песчано-гравийные отложения и пески различной крупности	Высокопроницаемые, крупнозернистые осадки, способствующие миграции флюидов
1	Среднезернистые пески	Средняя проницаемость, хорошо отсортированные кварцсодержащие породы
2	Мелкозернистые пески	Более низкая проницаемость, мелкие детритовые осадки с повышенной степенью уплотнения
3	Песчаник	Индурированные породы с умеренной проницаемостью, потенциальные

		коллекторы
4	Супеси, алевролиты и глины	Низкопроницаемые толщи, обычно формирующие экранирующие слои над ураноносными образованиями
5	Непроницаемые пески	Плохо отсортированные, глинистые пески со значительно сниженной пористостью
6	Прочие классы (разнородные/неопределенные литологии)	Включают сложные литологии, зоны изменения пород и вторичные отложения

Анализ распределения интервалов по укрупненным литологическим классам показал умеренный, но геологически ожидаемый дисбаланс, обусловленный строением палеодолинных флювиальных разрезов. Наиболее широко представлены песчано-гравийные и средне-крупнозернистые песчаные разности, соответствующие главным палеорусловым коллекторским телам и их переработанным русловым пакетам. Мелкозернистые пески, алевроито-глинистые и глинистые разности встречаются в меньшем объеме и преимущественно приурочены к пойменно-аллювиальным, боковым и кровельным экранирующим пачкам, а также к зонам фациальных переходов на границе "коллектор - экран". Данное соотношение классов является методически значимым, поскольку оно заранее определяет неоднородность уверенности модели: хорошо представленные коллекторские фации, как правило, распознаются устойчивее, тогда как редкие и пограничные разности формируют зоны повышенной неопределенности и потенциально более высокой частоты ошибок. Распределение интервалов по укрупненным литологическим классам приведено на рисунке 3.4.

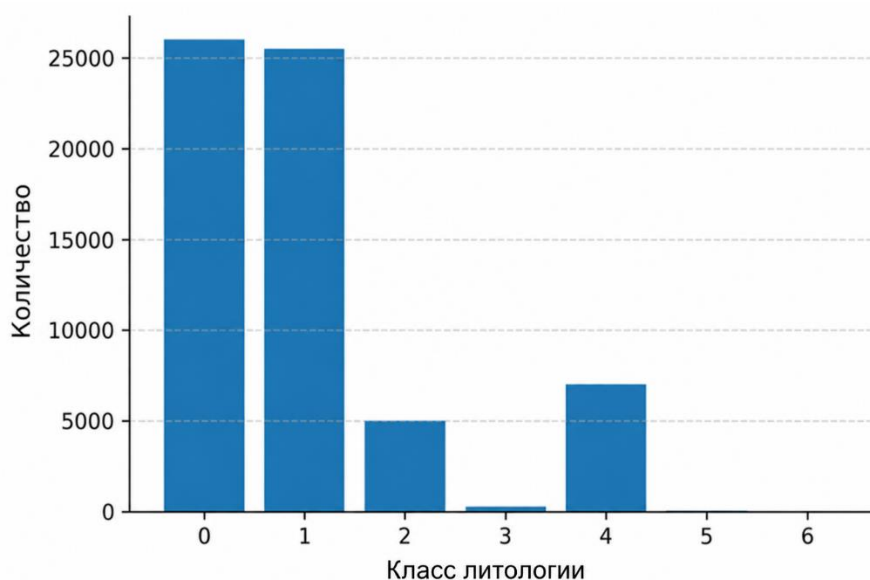


Рисунок 3.4 - Распределение классов литологии после повторной категоризации.

Признаковое пространство для обеих задач формировалось как сочетание базовых каротажных кривых и инженерных γ -признаков. Базовый набор включал гамма-каротаж, кажущееся электрическое сопротивление и самопроизвольную поляризацию, измеренные с шагом 0.1 м по стволу. Для усиления чувствительности к границам ураноносности и редокс-фронтам по γ -кривой были сконструированы два инженерных признака: производная по глубине $\Delta\gamma/\Delta z$ и квадрат γ^2 . Производная акцентирует локальные изменения радиоактивности и позволяет подчеркнуть потенциальные переходы от фоновых значений к аномально высоким уровням, характерным для ураноносных линз и зон вторичного накопления. Преобразование γ^2 усиливает вклад высоких γ -значений в дисперсии признака и тем самым повышает статистический «вес» интервалов с выраженной радиоактивной аномалией по сравнению с фоном. Важно, что эти признаки не подменяют геологическую интерпретацию порогами, а обеспечивают модели возможность учитывать форму сигнала и резкость переходов как потенциально информативные характеристики.

Алгоритмическая реализация предсказательных моделей была основана на методе случайного леса как ансамбля решающих деревьев, обучаемых на бутстрэп-подвыборках данных [77]. Разбиения в деревьях формировались по порогам признаков таким образом, чтобы максимизировать однородность классов в дочерних узлах (обычно по индексу Джини), при этом на каждом шаге использовалось случайное подмножество признаков, что снижало коррелированность деревьев и уменьшало риск переобучения. В результате нелинейные зависимости между глубиной, γ -откликом, его локальными изменениями и электрическими параметрами могли быть аппроксимированы без введения жёстких эвристических правил. Структура workflow обобщена на рисунке 3.5: каротажные кривые после контроля качества объединялись с ядерными литологическими индексами, выполнялась перекодировка в укрупнённые классы, рассчитывались инженерные γ -признаки, после чего обучались две модели случайного леса - для литологии и для продуктивности. На выходе формировались вероятностные оценки по глубине, которые далее использовались для валидации, профильной визуализации и подготовки входов для геостатистических процедур.

Несмотря на единое признаковое пространство, физическая природа решаемых задач различалась. Для литологической модели глубина рассматривалась как стратиграфический прокси-индикатор, отражающий повторяемость и вертикальную последовательность литологических пакетов в пределах палеодолинного разреза, включая чередование палеорусловых коллекторов и пойменно-аллювиальных (экранирующих) разностей, тогда как γ и его производные обеспечивали дифференциацию песчаных и глинистых литотипов, включая переходные фации. Для модели продуктивности на первый план выходили локальные γ -аномалии и их контраст относительно фонового уровня, а также резкость изменения γ , фиксируемая $\Delta\gamma/\Delta z$, что соответствует редокс-контролируемой природе ураноносности в пластово-инфильтрационных

системах. Электрические признаки рассматривались как поддерживающие, поскольку они повышали устойчивость различения интервалов с повышенной минерализацией поровых вод и/или глинистостью от промытых песчаных коллекторов, особенно в зонах пограничных γ -значений, где ураноносность проявляется вероятно и может быть сегментированной.

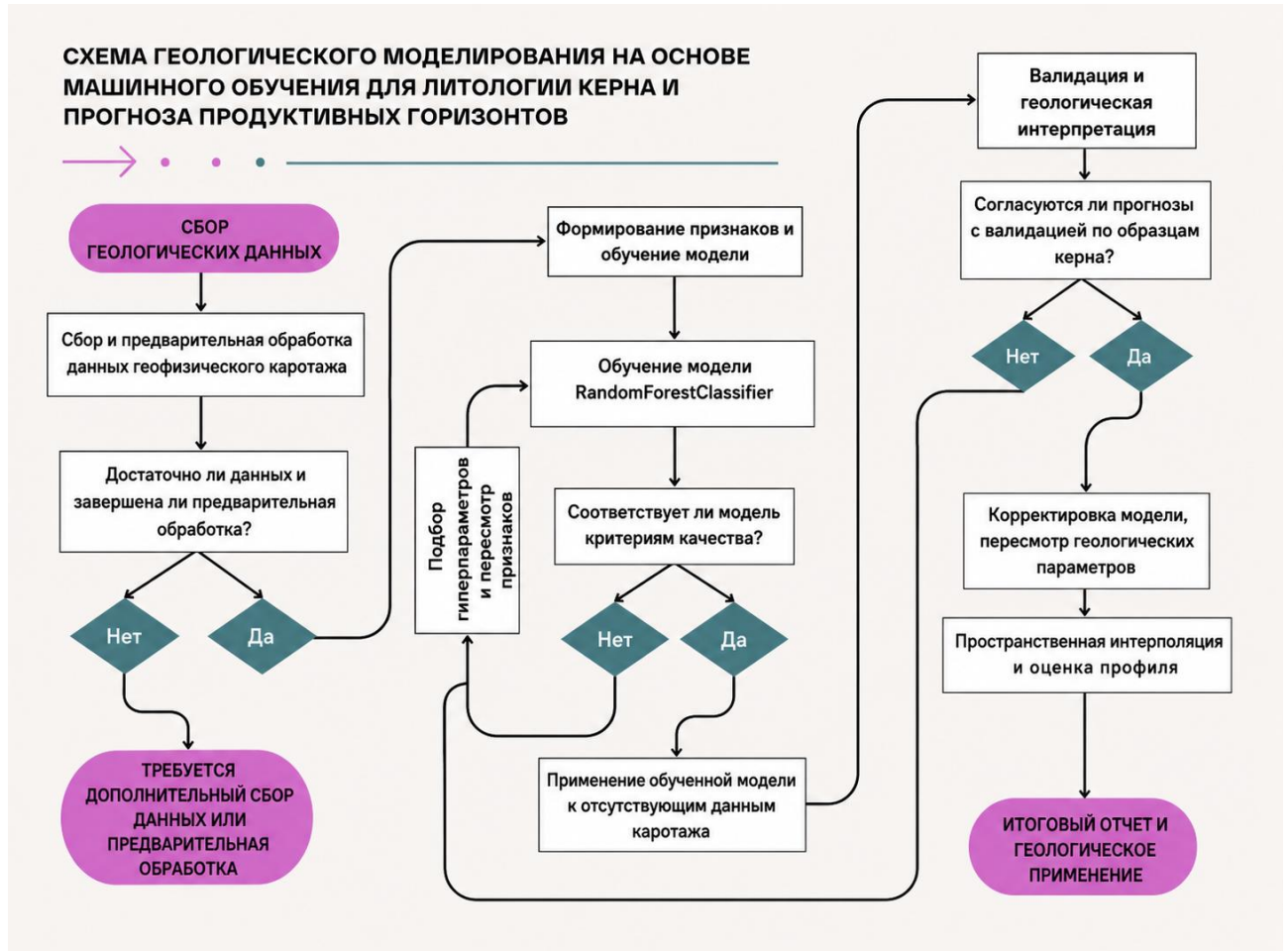


Рисунок 3.5 - Рабочий процесс геологического моделирования на основе машинного обучения, включающий предобработку данных, обучение модели, этапы валидации и геологическую интерпретацию с использованием логики принятия решений.

Для количественной оценки вклада признаков использовалась метрика важности случайного леса по среднему уменьшению неоднородности (индекса Джини) [77]. Как показано на рисунке 3.6, для литологической задачи глубина ожидаемо занимала доминирующее положение, γ и инженерные γ -признаки формировали следующую по значимости группу, а электрические параметры давали вклад второго порядка. Для задачи продуктивности относительная важность γ -признаков возрастала, а роль глубины несколько снижалась, что интерпретируется как отражение более прямой связи ураноносности с радиоактивными аномалиями по сравнению с общей стратиграфической архитектурой. Такое ранжирование согласуется с практикой применения

случайных лесов в задачах прогнозирования минерализации, где наиболее информативными оказываются параметры, непосредственно связанные с минерализацией, а признаки контекста обеспечивают стратиграфическую и гидродинамическую рамку [78, 79].

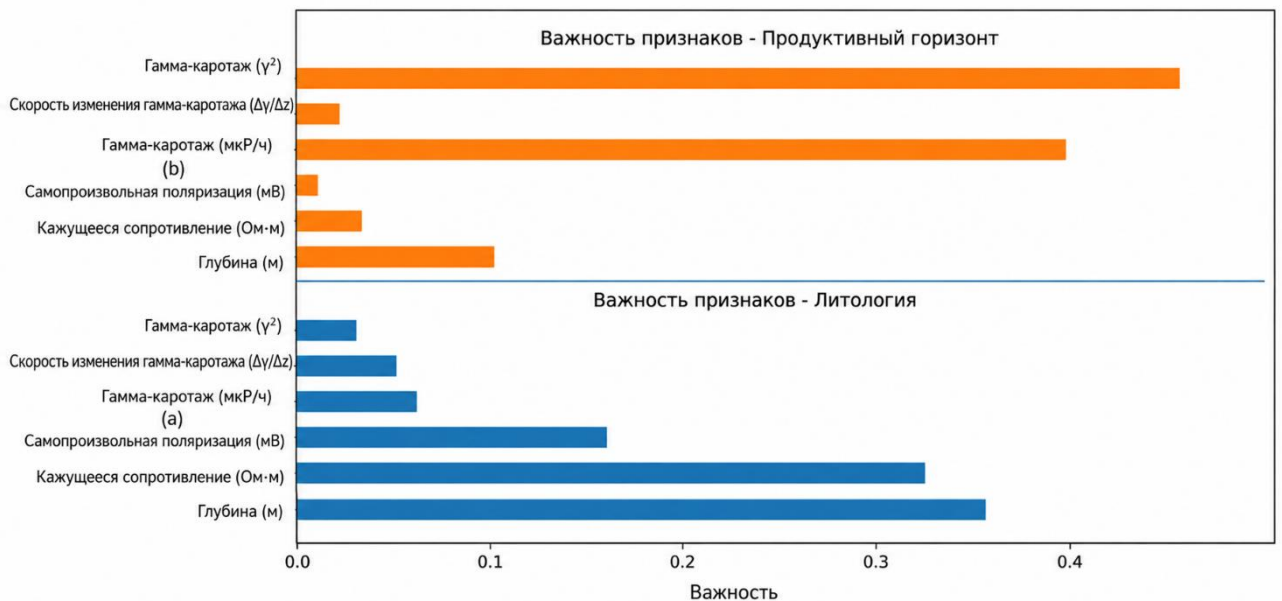


Рисунок 3.6 - Диаграммы важности признаков для: (а) литологии; (б) моделей продуктивных горизонтов.

В целом сочетание геологически осмысленной перекодировки литологии, инженерных γ -признаков и ансамблевых моделей случайного леса позволило сформировать интерпретируемый инструмент прогнозирования литологической структуры и вероятностного выделения ураноносных горизонтов. Существенным является то, что на выходе формировалось не бинарное решение, а поле вероятностей по глубине, пригодное для дальнейшей интеграции с геостатистическими процедурами. Протокол групповой кросс-валидации по скважинам, калибровка вероятностей и обращение со слабыми метками рассматриваются далее, поскольку именно они определяют, насколько полученные вероятности могут трактоваться как практические оценки риска/перспективности при планировании бурения.

3.3 Калибровка вероятностей, групповая (по скважинам) spatial-CV, исключение слабых меток из валидации

В условиях пространственно организованных данных критичным является способ оценки качества предсказаний. Скважины Северного рудного узла образуют сеть, приуроченную к палеодолиненным системам и связанным с ними палеорусловым коллекторским телам, поэтому случайное перемешивание интервалов между обучающей и проверочной выборками (в том числе при разбиении только по глубине) приводило бы к завышению метрик за счет пространственной утечки информации и повторяемости литолого-фациальных

пакетов. Чтобы воспроизвести реалистичный сценарий применения модели, была использована групповая кросс-валидация по скважинам, при которой интервалы каждой скважины целиком относились либо к обучающей, либо к валидационной выборке. Такой протокол имитирует ситуацию, когда разрез по части скважин подтвержден по керну и каротажу, а по другим скважинам модель используется для прогноза литологии и ураноносности, и соответствует рекомендациям по пространственно независимой оценке предсказательных моделей в геонаучных задачах [77, 78]. В результате полученные метрики интерпретируются как более консервативные и близкие к ожидаемому качеству при переносе на новые скважины и при оценке межскважинного пространства.

Дополнительной особенностью массива являлось наличие слабых меток продуктивности, сформированных по порогам γ -каротажа и вспомогательным критериям в интервалах без прямого керна подтверждения. Эти метки потенциально увеличивают объём обучающих данных и позволяют лучше охватить вариативность откликов вблизи редокс-границ, однако их неопределённость делает рискованным использование их при расчёте метрик качества наравне с надёжными метками. Поэтому была принята компромиссная стратегия: слабые метки допускались в обучающую выборку для стабилизации параметров модели, но полностью исключались из валидационных подвыборок. Таким образом, метрики качества рассчитывались только по интервалам с надёжной геологической разметкой, что снижало вероятность переоценки модели и делало выводы по кросс-валидации более релевантными для практики, где цена ложной уверенности в продуктивности является высокой.

Эффект сгруппированной кросс-валидации и фильтрации слабых меток наглядно проявлялся в структуре матриц ошибок для обеих задач. Как показано на рисунке 3.7, для литологической модели основная масса верных классификаций сосредоточена на диагонали, а ошибки имеют выраженный «локальный» характер и преимущественно приходятся на соседние по геологическому смыслу классы. Наиболее заметные взаимные подстановки наблюдаются между песчано-гравийными и среднезернистыми песками, а также между песчаными и слабо глинистыми разностями, что отражает постепенность смены фаций и неизбежное перекрытие диапазонов петрофизических откликов. Для модели продуктивности характер ошибок иной: истинно продуктивные интервалы в большинстве случаев выделяются корректно, а ложные срабатывания концентрируются в субпромышленных и переходных зонах с повышенным фоном γ и на границах линз, где ураноносность может проявляться неоднородно и не всегда формирует «чистый» сигнал. Такое поведение интерпретируется как закономерное следствие редокс-контроля, при котором максимальная неопределённость приходится на переходные области между явно пустыми и явно ураноперспективными продуктивными пакетами.

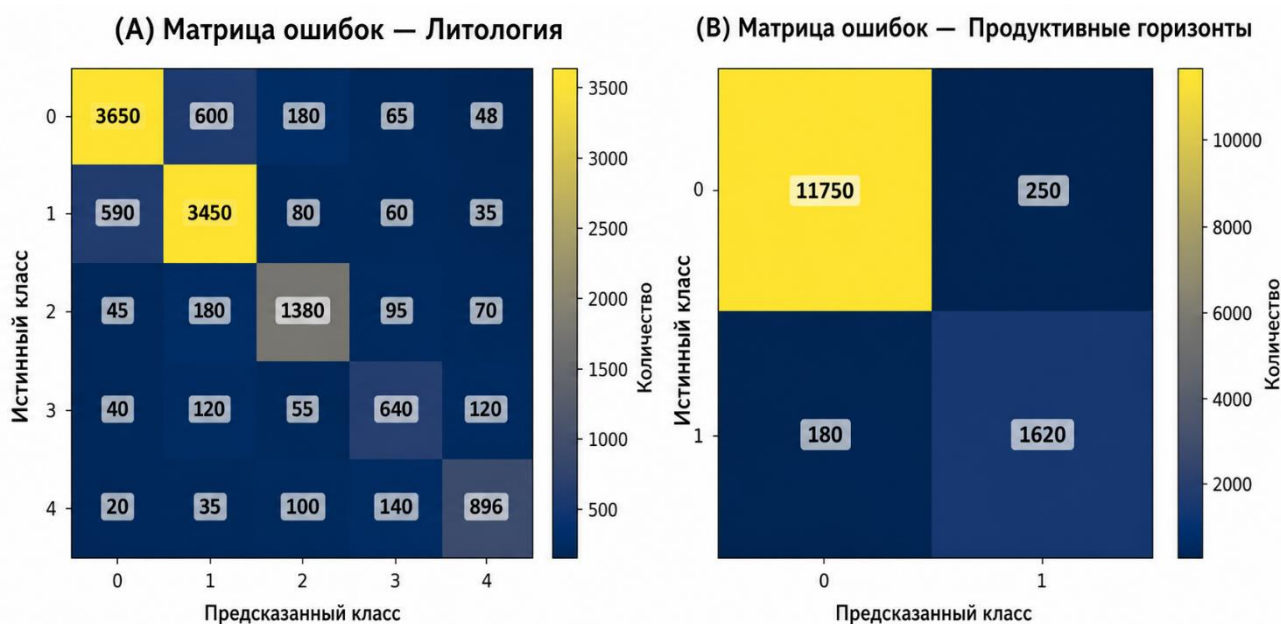


Рисунок 3.7 - Матрицы ошибок моделей Random Forest: а - классификация литологии по укрупненным классам 0-4, б - классификация продуктивных интервалов в бинарной схеме «пустой - продуктивный».

Количественная оценка многоклассовой литологической модели на сгруппированной кросс-валидации показала общую точность порядка 0.79 при макроусреднённых значениях точности и полноты около 0.76. Эти значения свидетельствуют о сбалансированной работе модели по основным классам при сохранении дисбаланса данных. Несколько более низкая полнота для мелкозернистых песков и глинистых разностей интерпретируется как следствие их меньшей представленности и петрофизической неоднозначности: диапазоны γ , сопротивления и ПС частично перекрываются с фоновыми значениями коллекторских песков, что затрудняет однозначное разделение. В таких условиях применение ROC-метрик может приводить к оптимистичным оценкам из-за большого числа истинно отрицательных примеров, поэтому более информативным инструментом анализа считаются кривые точность-полнота, чувствительные к качеству распознавания именно тех классов, которые практическая интерпретация рассматривает как критичные [80].

Для литологической модели кривые точность-полнота были построены по каждому классу с расчётом средней точности (average precision). Как показано на рисунке 3.8, значения AP варьируют примерно от 0.69 для глинистых толщ до 0.88 для грубообломочных коллекторов, что согласуется с геологическими ожиданиями: наиболее контрастные петрофизические различия характерны для грубообломочных пакетов, тогда как тонкозернистые и глинистые фации чаще формируют переходные зоны. Дополнительно была выполнена калибровка вероятностей, в которой предсказанные моделью вероятности сопоставлялись с фактическими долями наблюдений в бинах вероятности. Значение Brier-метрики порядка 0.139 указывает на умеренную консервативность

вероятностной шкалы: вероятности высоко уверенных предсказаний слегка занижаются, что проявляется в отклонении калибровочной кривой от диагонали идеальной калибровки [81]. Практически это означает, что вероятности модели следует рассматривать как несколько осторожные оценки, а пороги отбора классов целесообразно согласовывать с задачей (например, приоритетность коллекторов vs. экраны) и с допущениями последующей интерполяции.

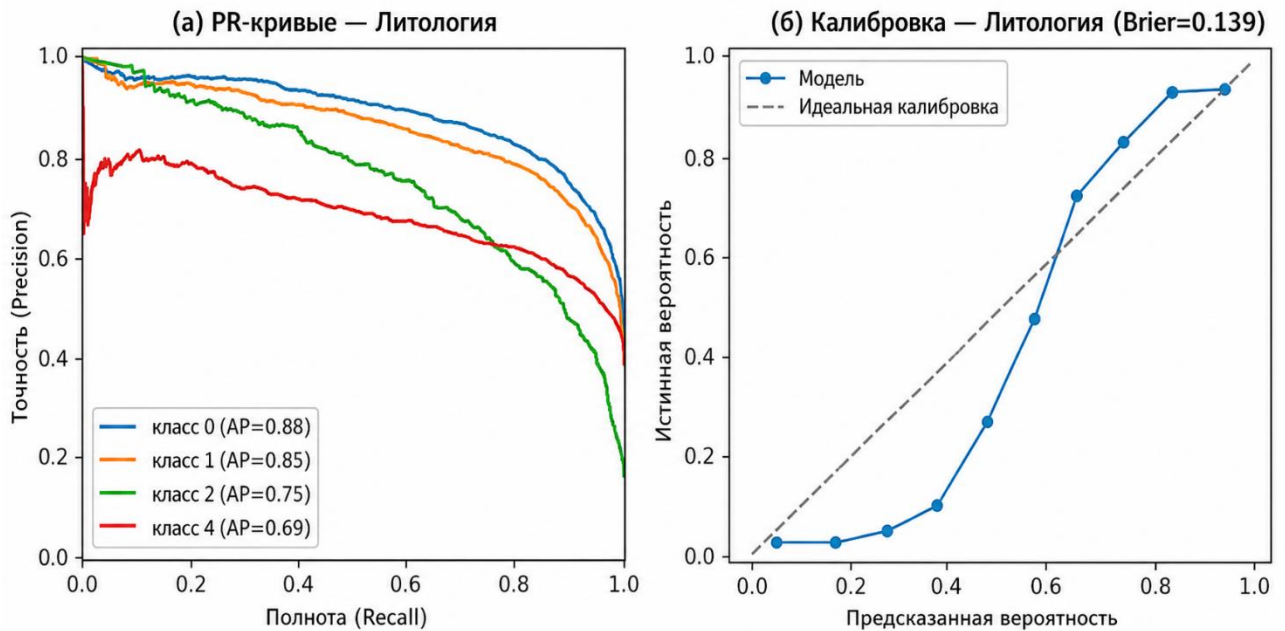


Рисунок 3.8 - Оценка многоклассовой модели литологии: а - кривые точность-полнота по укрупненным литологическим классам с приведенными значениями средней точности (AP); б - калибровочная кривая с указанием значения Brier-метрики.

Для модели продуктивности акцент при оценке качества ещё сильнее был смещён в сторону PR-кривой, поскольку положительный класс ураноперспективных продуктивных интервалов является редким, а практическая цена ошибок асимметрична. На сгруппированной кросс-валидации модель демонстрировала высокую общую точность порядка 0.97, при этом точность и полнота положительного класса составляли около 0.87 и 0.90. Средняя точность по PR-кривой достигала 0.94, что указывает на хорошую разделяемость ураноперспективных продуктивных и пустых интервалов по совокупности γ -признаков и электрических параметров. Калибровочный анализ показал, что Brier-метрика порядка 0.067 существенно ниже, чем в литологической задаче, и отражает более согласованную вероятностную шкалу для продуктивности. Незначительные отклонения в области высоких вероятностей интерпретируются как проявление естественной геологической неоднородности переходных зон, где часть факторов, влияющих на фиксацию урана, не представлена в признаковом пространстве.

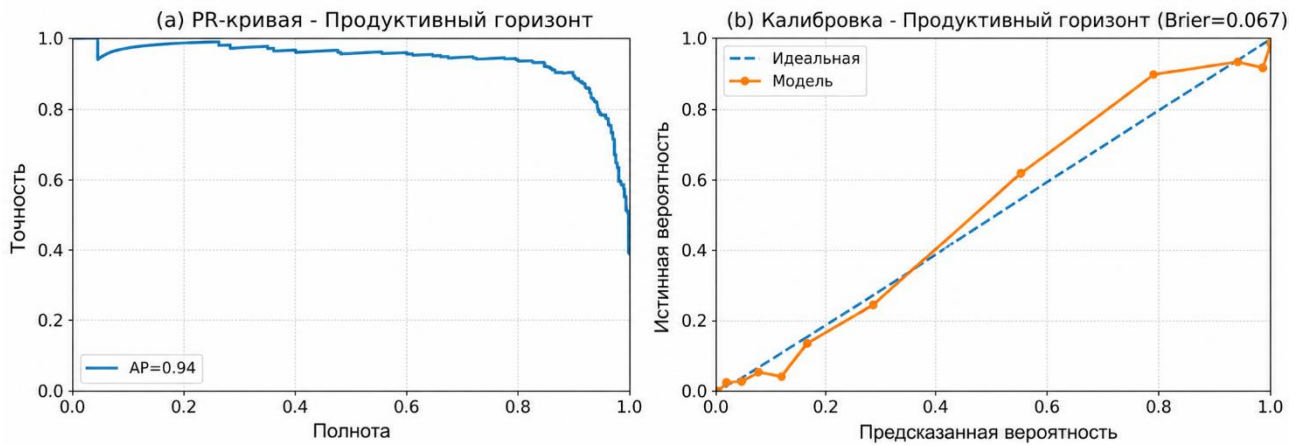


Рисунок 3.9 - Оценка модели продуктивного горизонта: а - кривая точность-полнота с указанной средней точностью (AP); б - калибровочная кривая с Brier-метрикой для вероятностных предсказаний продуктивных интервалов.

С практической точки зрения ключевым является выбор порога вероятности, отделяющего продуктивные интервалы от пустых. Стандартный порог 0.5 обеспечивает симметричный подход, однако не отражает реальную асимметрию потерь: пропуск продуктивного интервала в поисково-разведочном контексте обычно более критичен, чем ложная приписка продуктивности. Вместе с тем чрезмерное снижение порога неизбежно ведёт к росту ложных срабатываний и увеличению объёма потенциально неоправданного бурения. Поэтому порог подбирался с учётом формы PR-кривой, вида калибровочной зависимости и геолого-экономической логики балансировки полноты и точности. Наличие калиброванной вероятностной шкалы позволяет трактовать вероятности не как абстрактные оценки модели, а как приближённые оценки частоты продуктивных интервалов в группах точек с близкими значениями вероятности, что облегчает коммуникацию результатов с интерпретаторами и планировщиками бурения [81].

Для иллюстрации работы протокола по реальному разрезу была проанализирована скважина, не участвовавшая в обучении на соответствующей итерации CV. Как показано на рисунке 3.10, совместное рассмотрение каротажных кривых, надёжной разметки продуктивности и прогнозов модели позволяет проследить, как вероятностная модель воспроизводит ураноносные линзы и межлинзовые пустые зоны. В интервалах с надёжными метками модель, как правило, формирует высокие вероятности, тогда как в областях со слабой или отсутствующей разметкой вероятности принимают промежуточные значения, отражающие потенциальную продуктивность при наличии неопределённости. Принципиально важно, что такие интервалы не участвовали в расчёте метрик качества, но их визуальная диагностика позволяет оценить, где модель может быть полезна как инструмент предварительного ранжирования.

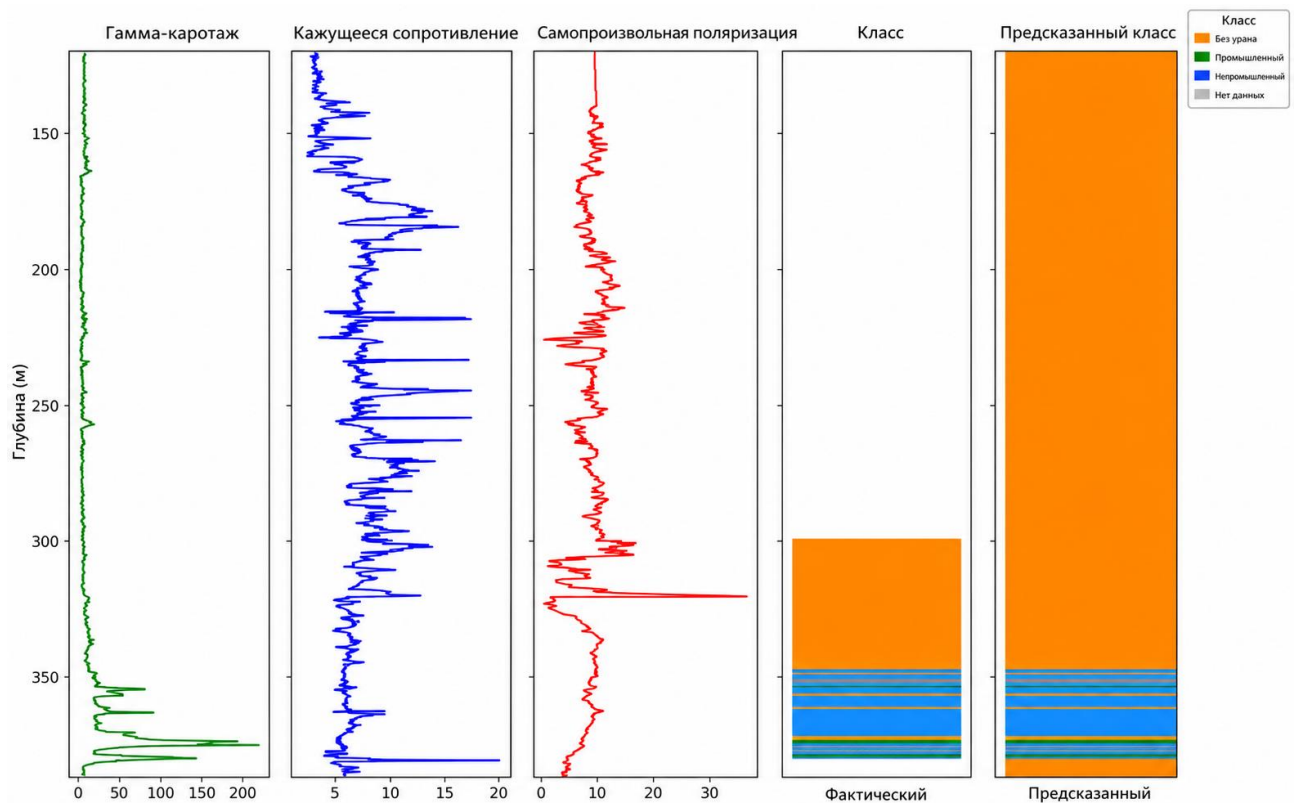


Рисунок 3.10 - Пример по скважине Well_XX: каротажные кривые, фактическая разметка продуктивных интервалов и прогноз модели Random Forest с указанием вероятности продуктивного горизонта и маскировкой интервалов со слабыми метками в валидации.

В целом выбранный протокол оценки качества объединяет групповую кросс-валидацию по скважинам, исключение слабых меток из валидации и анализ как дискретных метрик (точность, полнота, F1), так и вероятностных характеристик (PR-кривые, калибровка, Brier-метрика). Такой подход обеспечивает консервативные, но более реалистичные оценки качества, снижает риск пространственной утечки информации и позволяет трактовать вероятностные прогнозы как практически интерпретируемую шкалу, пригодную для дальнейшей интеграции с геостатистическим моделированием и задачами приоритизации бурения.

3.4 Визуализация по профилям (cubic spline), сравнение с экспертной интерпретацией

Профильная визуализация результатов моделирования использовалась как связующее звено между дискретными вероятностными оценками по скважинам и непрерывным представлением ураноносного горизонта в пределах Северного рудного узла. В подразделах 3.2 и 3.3 было показано, что модели случайного леса, обученные на петрофизически осмысленном наборе γ -признаков и электрических параметров, формируют калиброванные вероятности литологических классов и продуктивности при пространственно

независимой оценке. В настоящем подразделе эти вероятности были перенесены в плоскость профильных разрезов для группы скважин Профиля 1, где требовалось получить геологически правдоподобное, но при этом количественно управляемое описание формы горизонта и его связи с литологическими пакетами. Подчёркивается, что речь идёт не о восстановлении «истинной» геометрии горизонта, а о построении вероятностно согласованной интерпретации, отражающей наиболее поддерживаемую данными тенденцию при наличии неизбежных зон неопределённости.

Переход от одномерных предсказаний по стволам к двумерной профильной картине особенно важен для пластово-инфильтрационных урановых систем, где редокс-фронт и ураноносные линзы обладают линзовидной, субгоризонтальной или слабонаклонной геометрией и осложняются локальными вариациями гранулометрии и проницаемости. В таких условиях интерпретация требует согласования вероятностных оценок между скважинами, вскрывающими палеодолинную систему и палеорусловые коллекторские тела. Для построения гладкой профильной аппроксимации между узловыми точками, заданными по скважинам, был использован кубический сплайн как метод, обеспечивающий непрерывность кривой и ее первых двух производных, что соответствует требованию геологической правдоподобности при реконструкции стратиформных горизонтов и плавных фациальных переходов вдоль профиля [82].

Математически кубический сплайн задаётся кусочно-кубическими полиномами на интервалах между узлами профиля с условиями гладкости в точках стыка [82-83]. Пусть вдоль профиля заданы координаты узлов x_i , соответствующие положениям скважин, а функция $S(x)$ описывает глубину кровли (или подошвы) продуктивного горизонта. На каждом интервале $[x_i, x_{i+1}]$ сплайн представляется полиномом третьей степени (3.1):

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3, \quad (3.1)$$

коэффициенты которого подбираются так, чтобы выполнялись условия прохождения через узлы и гладкость по первой и второй производным, а также заданные граничные условия на концах профиля [82-83]. Тем самым формируется непрерывная линия, проходящая через узловые точки и избегающая резких изломов, которые в большинстве случаев не ожидаются для стратиформной геометрии ураноносных тел.

В рассматриваемой постановке узловыми точками для ураноперспективного продуктивного горизонта служили глубины, где по итоговой вероятностной модели вероятность промышленной ураноносности превышала выбранный порог. Для каждого ствола в пределах Профиля 1 определялись интервалы, интерпретируемые как продуктивные с высокой вероятностью, после чего из них извлекались характерные глубины (например, кровля/подошва или центр тяжести по мощности), использованные как узлы сплайна. Аналогично по литологической модели формировались узлы для границ основных литологических пакетов, что позволило визуализировать коллекторно-экранирующее строение разреза совместно с вероятной геометрией ураноперспективного горизонта. Подчёркивается, что

сглаживающий характер сплайна уменьшает влияние локального шума и единичных аномальных значений вероятности, однако одновременно может приводить к частичному «усреднению» мелкомасштабной сегментации, что должно учитываться при интерпретации.

На рисунке 3.11 представлены разрезы по Профилю 1: в панели (а) показана стратиграфическая схема, восстановленная по результатам литологической модели, а в панели (б) - положение продуктивных горизонтов, построенное по узловым точкам, полученным из вероятностной классификации, с последующей профильной аппроксимацией кубическим сплайном. В совокупности разрезы демонстрируют вероятную приуроченность ураноносного горизонта к коллекторским песчано-гравийным и песчаным пакетам, ограниченным экранирующими глинистыми толщами, и показывают, что модельно-сплайновое решение воспроизводит ожидаемую стратиформную тенденцию в пределах палеодолинной системы и связанных палеорусловых коллекторов, сохраняя зоны повышенной неопределенности в межскважинных интервалах.

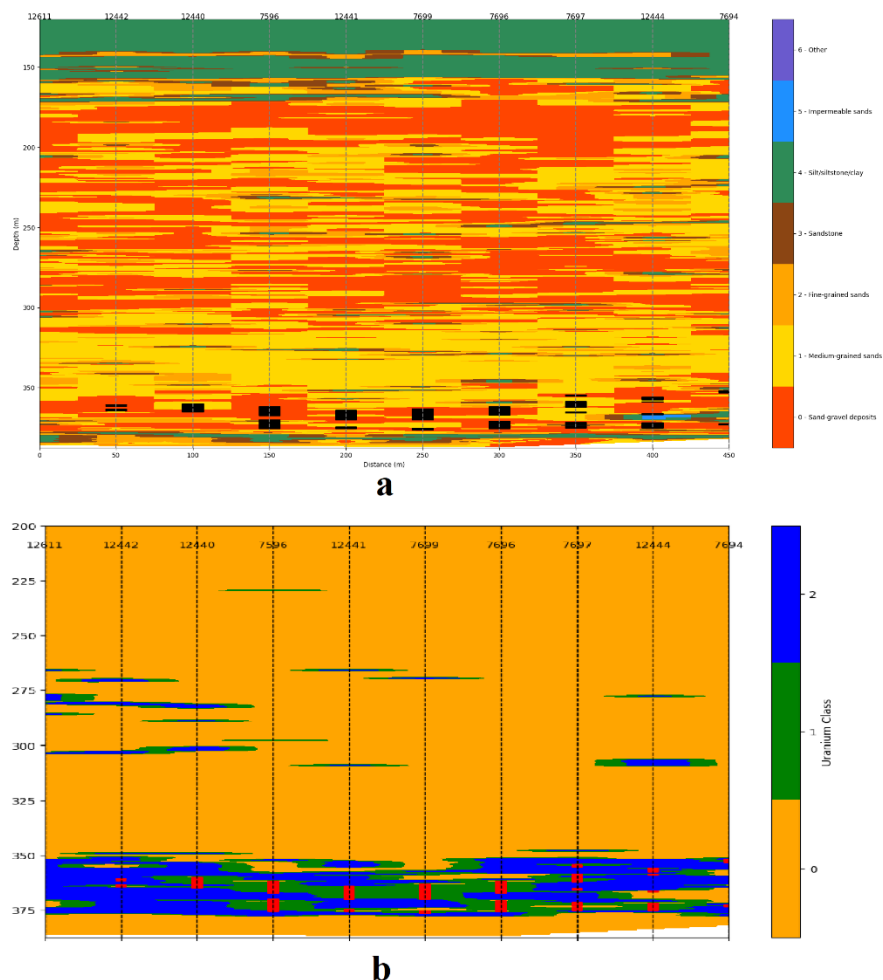


Рисунок 3.11 - Поперечные разрезы по Профилю 1: а - стратиграфические слои, восстановленные на основе литологической модели; б - ураноносные продуктивные горизонты, интерполированные кубическим сплайном по вероятностным прогнозам».

Геолого-геофизическая интерпретация профиля на рисунке 3.11 показывает, что ураноносный горизонт преимущественно приурочен к интервалам, интерпретируемым как песчано-гравийные и среднезернистые песчаные разности, формирующие относительно непрерывные палеорусловые коллекторские тела в пределах палеодолинной системы. Профильная аппроксимация кубическим сплайном подчеркивает согласованную тенденцию изменения глубины горизонта вдоль профиля, которая может отражать влияние регионального гидродинамического режима, контролирующего миграцию окисленных ураносодержащих растворов и положение редокс-фронта. Локальные участки увеличения мощности и изменения положения горизонта интерпретируются как потенциальные зоны усиленного рудонакопления, обусловленные сочетанием повышенной проницаемости коллектора, неоднородности фациального строения и благоприятных редокс-условий. При этом в межскважинных участках, удаленных от точек наблюдения, рост неопределенности является ожидаемым и должен учитываться при последующем пространственном моделировании и при выборе участков для уточняющего (опережающего) бурения.

Сопоставление модельно-сплайновой интерпретации с экспертной ручной интерпретацией было выполнено на рисунке 3.12. На сравнительном профиле представлены два варианта очертания горизонта: первый построен по вероятностным прогнозам модели с последующей сплайновой интерполяцией, второй отражает традиционную интерпретацию по γ -кривым, керновым данным и геологическому опыту. Такое сопоставление позволяет оценить согласованность модели с экспертным знанием и локализовать сегменты, где расхождения могут быть связаны либо с ограничениями признакового пространства, либо с различиями в трактовке переходных интервалов.

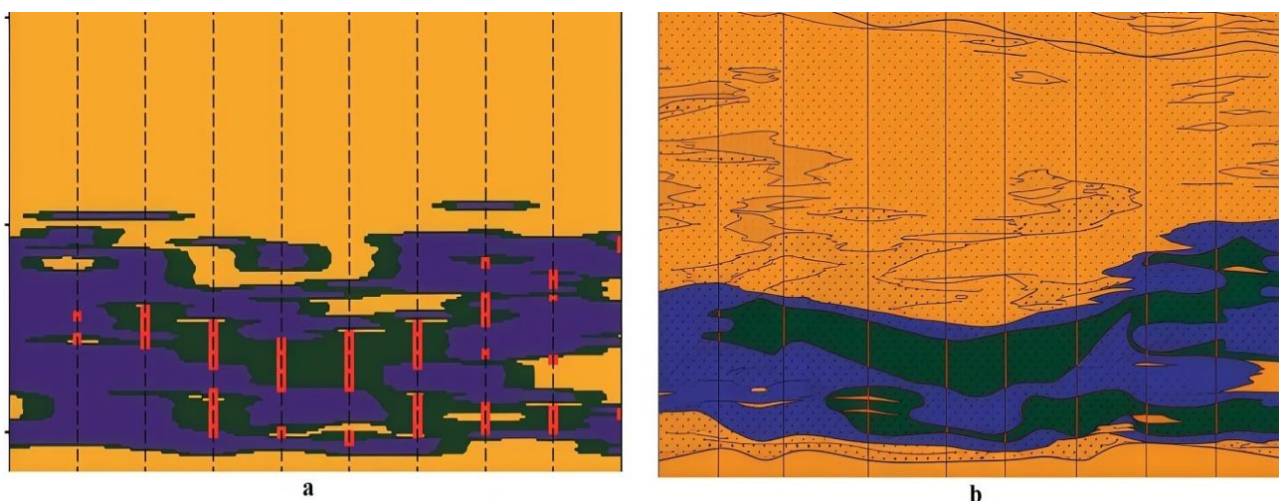


Рисунок 3.12 - Сравнение профилей ураноносного горизонта по Профилю 1: модель Random Forest с кубическим сплайном и ручная геологическая интерпретация. Условные обозначения по рисунку 3.11b

Анализ рисунка 3.12 показывает, что на большей части профиля положение горизонта по модели и по ручной интерпретации совпадает в пределах нескольких метров, что рассматривается как приемлемый уровень расхождения для песчано-гравийных тел данного масштаба и для постановки, ориентированной на вероятностное картирование. В ряде сегментов фиксируются систематические отличия: модель местами «сглаживает» кровлю горизонта, тогда как экспертная интерпретация выделяет локальные выступы или ступени, потенциально связанные с мелкомасштабной литологической неоднородностью, изменениями гранулометрии или вторичными процессами перераспределения урана. В других зонах модель, напротив, расширяет область вероятной продуктивности по сравнению с более консервативным очертанием интерпретатора, что может отражать повышенную чувствительность к сочетаниям признаков и указывать на участки, где продуктивность была потенциально недооценена при пороговом подходе.

Такое поведение объясняется различиями в обращении с переходными интервалами. Ручная интерпретация, как правило, опирается на более жёсткие пороги и стремится минимизировать риск завышения продуктивности, тогда как модель предоставляет непрерывную шкалу вероятностей, дополнительно подтверждённую калибровкой, и может присваивать умеренно высокие вероятности там, где признаки находятся в пограничной зоне. Сглаживающая природа сплайна усиливает тенденцию к более регулярной геометрии, что повышает удобство последующей интерполяции, но снижает детализацию мелкомасштабной сегментации. В итоге сопоставление на рисунке 3.12 следует рассматривать как согласование двух типов знания: вероятностного статистического и контекстного экспертного, где расхождения являются не ошибкой как таковой, а индикатором зон повышенной неопределённости и приоритетных участков для доразведки.

Практические преимущества профильной визуализации заключаются в том, что формируется непрерывное представление горизонта, пригодное в качестве входа для трёхмерного геостатистического моделирования и построения объёмных карт вероятности перехвата горизонта в малых ячейках. Дополнительно профили позволяют заранее выявлять участки, где расхождения между моделью и экспертом максимальны, что делает их естественными кандидатами для уточняющего бурения и для уменьшения неопределённости. Наконец, профильная форма представления облегчает коммуникацию с практиками, для которых интерпретация разрезов является базовым инструментом принятия решений. Таким образом, сплайновая профильная визуализация завершает связку между моделями случайного леса, калиброванными вероятностями и геологической интерпретацией, формируя вероятностно согласованное решение, пригодное для последующей интеграции с геостатистическими процедурами.

3.5 Чувствительность, ограничители и переносимость на родственные ролл-фронт системы

Результаты подразделов 3.2-3.4 показывают, что поведение моделей случайного леса для литологии и продуктивности определяется ограниченным, но физически интерпретируемым набором признаков и строгим протоколом оценки по скважинам. При этом практическая ценность подхода проявляется не в «абсолютной» реконструкции горизонтов, а в вероятностной приоритизации интервалов и зон, где сочетание признаков делает ураноносность наиболее правдоподобной. В данном подразделе рассмотрены чувствительность к признакам, ограничения при генерализации тонких и сложных интервалов и переносимость workflow на родственные ролл-фронт системы.

Диаграммы важности признаков (рисунок 3.6) показывают доминирование глубины и γ -признаков (γ , γ^2 , $\Delta\gamma/\Delta z$), тогда как кажущееся сопротивление и ПС дают вклад второго порядка. Такое распределение важности интерпретируется как геологически закономерное: глубина выступает прокси стратиграфического положения и косвенно отражает расположение относительно региональных гидродинамических и редокс-границ [73], а γ -канал напрямую связан с радиоактивностью, обусловленной урановой минерализацией и/или глинистостью, что делает его основным источником диагностической информации [74, 81]. Инженерные γ -признаки усиливают именно те аспекты сигнала, которые наиболее информативны для ролл-фронт систем: резкость переходов и выраженность аномалий, потенциально соответствующих границам редокс-фронтов и зонам фиксации урана. Умеренная, но устойчиво ненулевая роль электрических параметров указывает на их значимость в пограничных интервалах, где γ -сигнал является неоднозначным и где изменения минерализации и глинистости могут формировать альтернативные объяснения γ -аномалии [75, 77]. Следовательно, при расширении набора данных или переносе workflow на другие участки приоритет должен отдаваться стабильной калибровке γ -канала и физически осмысленным γ -преобразованиям, тогда как электрический блок может рассматриваться как модуль, чувствительный к качеству измерений и к локальным гидрогеохимическим условиям.

Ограничения моделей наиболее наглядно проявляются в структуре ошибок и особенностях калибровки вероятностей (рисунки 3.7-3.10). Для литологической задачи основные ошибки приходятся на соседние по смыслу классы, что отражает постепенность фациальных переходов и перекрытие петрофизических диапазонов. При шаге 0.1 м создаётся высокая дискретизация, однако эффективное вертикальное разрешение определяется длиной зондирования и мощностью пластов, поэтому тонкие переходы внутри коллекторских пакетов неизбежно оказываются сглаженными в регистрируемом отклике. Для задачи продуктивности зона максимальной неопределённости локализуется на границах линз и в субпромышленных интервалах, где значения признаков близки к пороговым и где продуктивность может проявляться мозаично. PR-кривые и калибровочные зависимости

(рисунки 3.8 и 3.9) показывают, что модель уверенно отделяет «явно пустое» от «явно продуктивного», однако сохраняет промежуточный интервал вероятностей, который следует трактовать как область потенциальной продуктивности, требующую доразведки, а не как основание для жёстких решений [80, 81]. В практическом смысле именно эта область является наиболее значимой, поскольку она отвечает реальному геологическому сценарию переходных редокс-зон.

Профильные сопоставления (рисунки 3.11 и 3.12) демонстрируют, как указанные ограничения проявляются в геометрии горизонта. Кубический сплайн, являясь сглаживающей интерполяцией, способствует формированию более регулярных и непрерывных поверхностей, удобных для последующего объёмного моделирования, однако локальная сегментация и мелкомасштабные линзы могут быть частично усреднены. Ручная интерпретация, напротив, способна фиксировать такие детали, но ценой повышенной субъективности и ограниченной возможности количественно оценивать неопределённость. Поэтому модельно-сплайновые горизонты целесообразно рассматривать как вероятностную «опорную» интерпретацию, пригодную для ранжирования и планирования, а не как замену комплексной интерпретации по керну и сопутствующим данным. В прикладной схеме целесообразно использовать совпадение модельной и экспертной интерпретации как зону высокого доверия, а зоны расхождений - как приоритетные участки для дополнительного опробования и уточняющего бурения.

Переносимость подхода на родственные ролл-фронт системы должна трактоваться как переносимость workflow, а не переносимость конкретно обученной модели. Общие предпосылки - редокс-контролируемое оруденение в песчано-гравийных коллекторах, приуроченность продуктивности к определённым стратиграфическим уровням и доминирующая роль γ -каротажа - характерны для многих объектов Центральной Азии и Китая, что делает структуру workflow потенциально универсальной [73, 81]. Однако диапазоны γ , сопротивления и ПС, а также форма переходов через редокс-границы способны существенно различаться из-за вариаций минералогии цемента, диагенеза, состава пластовых вод и гидрогеохимической эволюции бассейна. Поэтому при переносе методики должны сохраняться принципы реклассификации литологии, конструирования инженерных признаков, групповой кросс-валидации по скважинам, исключения слабых меток из валидации и калибровки вероятностей, но параметры моделей, пороги вероятностей и, при необходимости, сам состав признаков должны переоцениваться на данных нового объекта.

Дополнительным ограничителем переносимости является качество исходных данных. Устойчивость результатов критически зависит от калибровки γ -канала, межскважинной сопоставимости кривых и отсутствия систематических смещений, которые могут быть внесены различиями приборного парка и полевой методики. В условиях недостаточного контроля качества воспроизведение workflow может привести к формально высоким

метрикам на «внутренней» проверке, но к деградации качества при переносе на новые скважины. Следовательно, при применении методики на других участках первоочередным должно быть обеспечение минимальных стандартов контроля качества каротажных измерений, рекомендуемых для урановой геологоразведки [74].

В обобщённом виде можно заключить, что модели случайного леса, построенные на петрофизически обоснованном наборе γ -признаков, электрических параметров и глубины, в сочетании с групповой кросс-валидацией по скважинам, калибровкой вероятностей и профильной сплайновой интерполяцией, обеспечивают количественно проверяемое и геологически интерпретируемое вероятностное выделение ураноносных горизонтов в пределах Северного рудного узла. Полученные результаты являются сопоставимыми с экспертной интерпретацией на уровне общей геометрии горизонта и пригодны для переноса на родственные ролл-фронт системы в виде унифицированного workflow при условии локальной адаптации и строгого контроля качества данных.

4 РЕГИОНАЛЬНАЯ ПЕРСПЕКТИВНОСТЬ (МРМ) НА Au-Ag МИНЕРАЛИЗАЦИЮ В ПРЕДЕЛАХ ВЕЛИКОГО БАССЕЙНА (НЕВАДА, США): АОА-ОГРАНИЧЕНИЕ, КАЛИБРОВКА И ТОР-к НАВИГАЦИЯ РЕШЕНИЙ

4.1 Исходные слои: геология, структура, тепловой поток, поля и производные; конструирование признаков

Региональная модель минерагенической перспективности Au-Ag для центрально-западного сегмента Великого Бассейна была сформирована на основе интеграции нескольких комплементарных семейств исходных данных, включающих обобщенную геологию, структурный каркас, поля потенциальных типов, производные гравитационных и псевдогравитационных аномалий, а также современные оценки теплового потока. Применение именно мультимодальной основы является методически оправданным, поскольку пространственное размещение эпитермальных Au-Ag систем и близких к ним по генетической связи магматически контролируемых рудных систем определяется совместным действием (i) литолого-стратиграфического строения и физико-химических свойств вмещающих толщ, (ii) архитектуры разломно-блоковой проницаемости и зон длительно поддерживаемой трещиноватости, и (iii) долговременного термального режима литосферы, включая роль тепловых аномалий и магматических источников тепла [84-85]. В настоящем подразделе последовательно описывается состав используемых карт и растровых полей, принципы их приведения к единой расчетной сетке, а также то, каким образом исходные слои были преобразованы в воспроизводимый набор признаков, пригодный для последующего применения методов машинного обучения в рамках региональной МРМ.

Геологическая основа включает обобщенную карту литостратиграфических комплексов Великого Бассейна, сгруппированных по их функциональной роли в минерал-системной модели Au-Ag. В частности, в качестве ключевых доменов учитывались докайнозойские карбонатные и кремнистые толщи, мезозойские магматические дуги, кайнозойские вулканогенные и субвулканические комплексы, а также покровы четвертичных отложений, которые способны частично маскировать или модифицировать выраженность геофизических сигналов в потенциальных полях, особенно при региональной гармонизации данных [86]. В пределах этой геологической основы была выделена площадь моделирования, ориентированная на крупнейшие и наиболее изученные Au-Ag районы, после чего были нанесены точки известных объектов минерализованности, включающие проявления, рудопроявления и месторождения. Такое представление позволяет сопоставить пространственное распределение Au-Ag объектов с доминирующими литологическими и стратиграфическими пакетами, а также оценить степень приуроченности минерализации к определенным вмещающим средам, которые обеспечивают сочетание (а) проницаемости, (б) благоприятной реакционной способности пород, и (в) устойчивых путей флюидопереноса. Как показано на

рисунке 4.1, наиболее плотные скопления Au-Ag объектов приурочены к зонам сочленения кайнозойских вулканогенных комплексов с более древними карбонатными и кремнистыми толщами, что согласуется с типовой минерал-системной интерпретацией, в которой высокопроницаемые структурные коридоры сочетаются с благоприятными литологическими доменами и контрастами физических свойств.

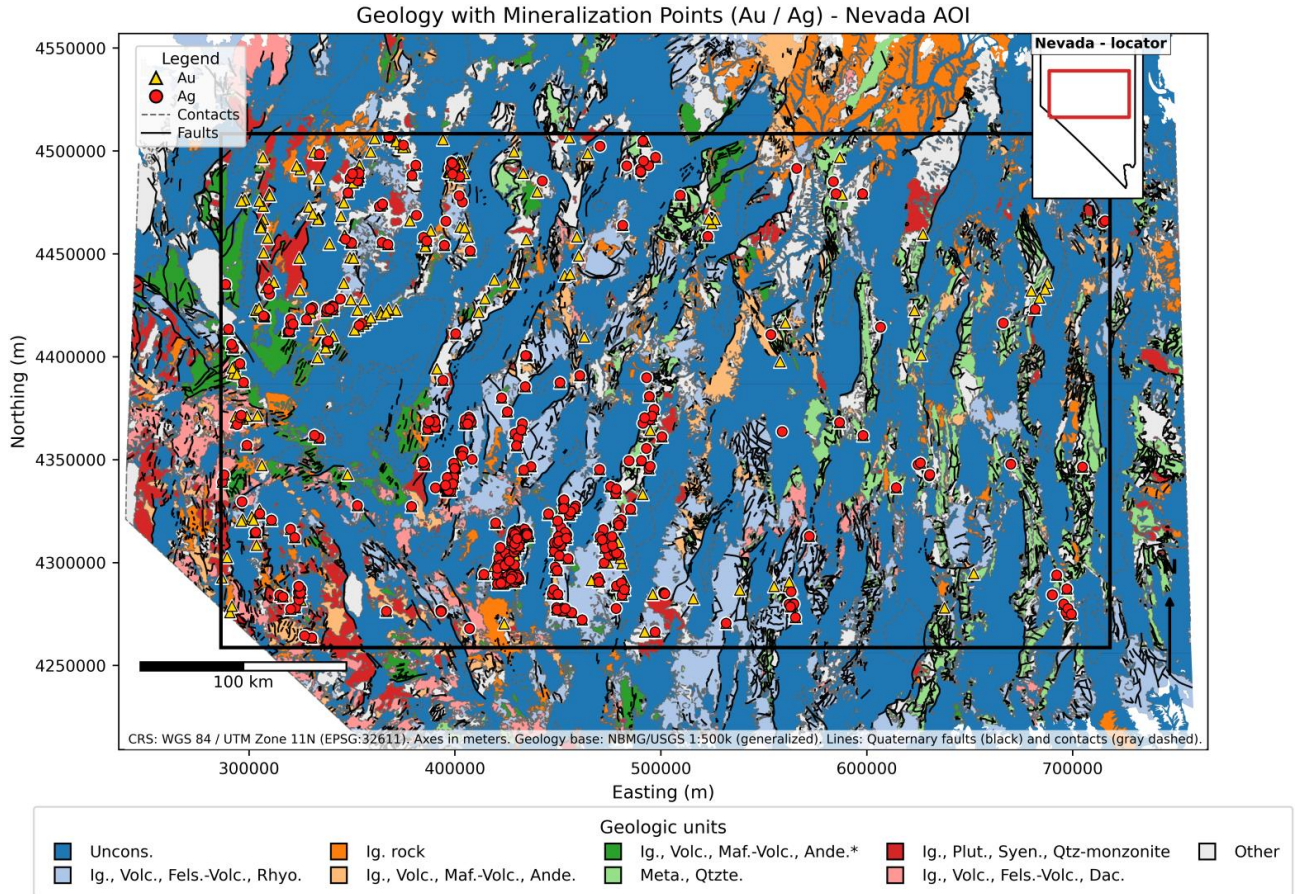


Рисунок 4.1 - Региональная геологическая основа площади исследования в пределах изучаемой части Великого Бассейна и расположение Au-Ag объектов в пределах области моделирования минерагенической перспективности

Для калибровки региональной модели перспективности в контексте реальных рудных систем использовался набор опорных месторождений и рудных районов, репрезентирующих различные типы Au-Ag оруденения в пределах изучаемой провинции [85-87]. В данный перечень были включены крупные и хорошо изученные объекты, характеризующиеся различной структурно-тектонической позицией и стилями минерализации, включая Comstock, Goldfield, Tonopah, Round Mountain, Paradise Peak, Aurora и Rochester. По каждому объекту фиксировались координаты, доминирующий генетический тип (например, эпитермальные системы низкой, промежуточной или высокой сульфидации, включая кварц-алунитовые ассоциации для высокосульфидационных систем), а также ключевые литературные

источники, служащие базой для геологического контекстирования. Как видно из таблицы 4.1, опорные месторождения пространственно охватывают основные сегменты провинции и обеспечивают репрезентативный спектр геолого-структурных обстановок формирования Au-Ag систем. Это позволяет использовать их не только как бинарные точки Р в постановке PU-learning, но и как геологический эталон при последующей интерпретации карт перспективности и процедур топ-k таргетинга, где требуется согласование статистического ранжирования с геологической правдоподобностью.

Таблица 4.1 - Опорные Au-Ag месторождения и рудные районы Великого Бассейна, используемые для геологического контекстирования региональной модели перспективности

ID	Месторождение (район)	Тип месторождения	Широта (°N, WGS84)	Долгота (°W, WGS84)	Восточная координата (м, EPSG:32611)	Северная координата (м, EPSG:32611)	Источник
1	Комсток (Вирджиния Сити)	Эпитермальные Ag-Au жилы низкой - промежуточной сульфидации	39.31	119.65	271520	4354530	(Vikre, 1989)
2	Голдфилд	Высокосульфидационные эпитермальные Au (кварц - алунит)	37.71	117.23	479727	4173660	(John, 2001)
3	Тонопа	Эпитермальные Ag-Au жилы промежуточной сульфидации	38.08	117.23	479829	4214720	(Bonham & Garside, 1979)
4	Раунд Маунтин	Эпитермальное Au низкой сульфидации (массового типа)	38.70	117.07	493913	4283490	(John, 2001)
5	Парадайз Пик	Высокосульфидационные эпитермальные Au-Ag	38.75	117.97	415710	4289480	(Castor et al., 1990)
6	Аурора	Эпитермальные Ag-Au жилы низкой сульфидации	38.29	118.89	334712	4239680	(John, 2001)
7	Рочестер	Эпитермальные Ag-Au месторождения низкой сульфидации (массового типа)	40.45	118.27	392307	4478480	(John, 2001)

Структурный каркас региона задавался картой четвертичных и более древних разломов, оцифрованных по данным геолого-съёмочных организаций и результатам предшествующих региональных структурных обобщений [86-

88]. В рамках минерал-системного подхода разломы и их пространственная организация рассматриваются как ключевые элементы, контролирующие фокусировку восходящих флюидов, развитие зон повышенной трещиноватости и формирование структурных ловушек для гидротермальных потоков. В этой связи особое внимание уделялось крупным нормальным и сбросо-сдвиговым нарушениям, линейным зонам растяжения, сегментированным участкам раздвигая, а также узлам пересечений и сопряжений разломов, где часто формируются зоны повышенной проницаемости. На картографической основе структурный каркас представлен совокупностью линий разломов различных рангов, контактов блоков различного возраста и состава, а также контуров основных вулканических центров и грабенообразных депрессий. Как показано на рисунке 4.2, крупные Au-Ag районы тяготеют к сочетанию высокоамплитудных разломов с сегментированными участками раздвигая и изгибов, где создаются благоприятные условия для формирования проницаемых коридоров и поддержания гидротермальной циркуляции. В дальнейшем структурные элементы были количественно параметризованы через расстояния до разломов, показатели плотности их пересечений и связанные геомеханические атрибуты, позволяющие переходить от описательной тектоники к воспроизводимым численным признакам.

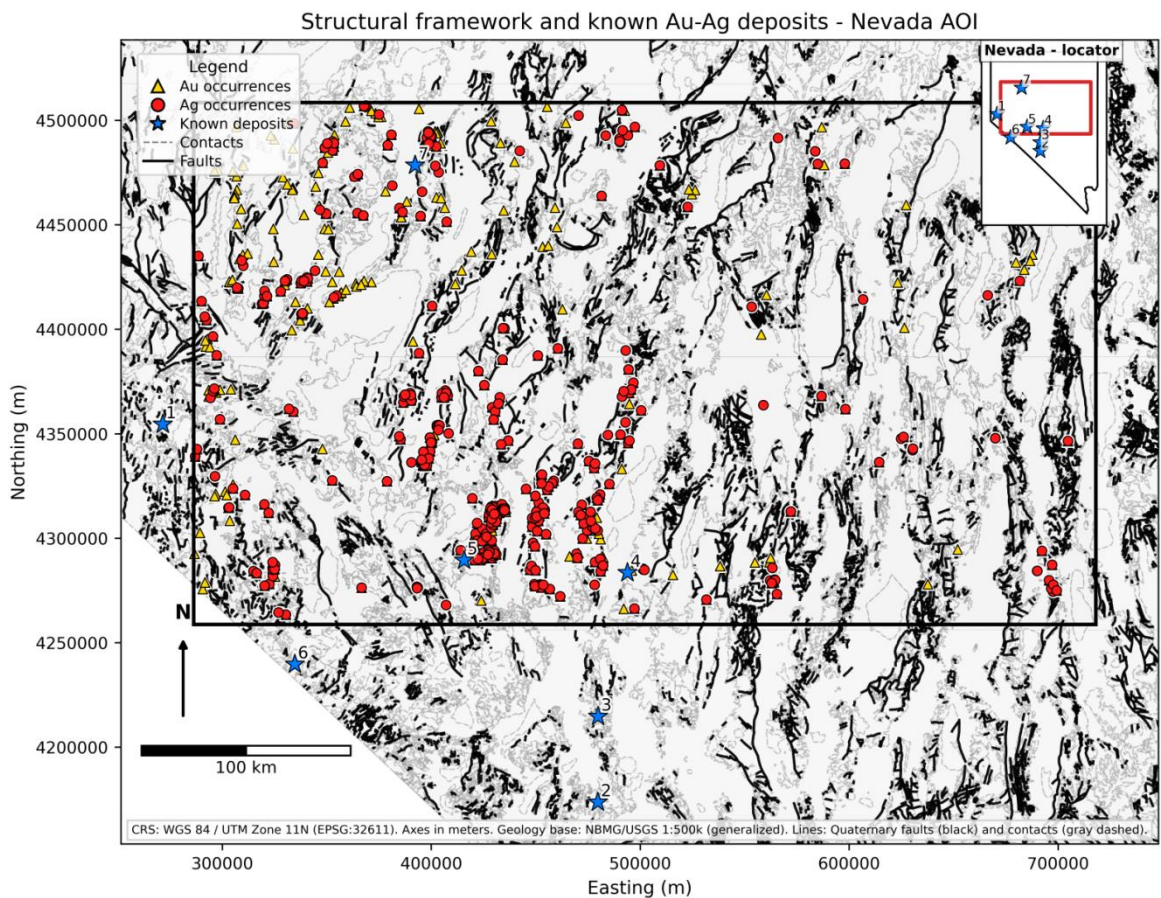


Рисунок 4.2 - Структурный каркас исследуемой территории и опорные Au-Ag месторождения, использованные для калибровки региональной модели перспективности

Наряду с геологическими и структурными данными ключевую роль в построении признакового пространства играли поля потенциальных типов и их производные, чувствительные к глубинному строению, латеральным контрастам плотности и намагниченности и, как следствие, к архитектуре блоков и границ доменов, релевантных магматизму и рудоконтролирующим структурам [89, 90]. В качестве базовых использовались гравиметрические и псевдогравитационные сетки, полученные путем гармонизации и фильтрации существующих карт Bouguer и производных по псевдогравитации. Псевдогравитационное преобразование магнитных данных применялось как стандартная процедура приведения магнитного сигнала к “гравитационно-подобному” поведению для унификации последующей интерпретации и получения сопоставимых производных, в частности горизонтальных градиентных характеристик. Для выделения региональных и субрегиональных аномалий применялись операции удаления длинноволнового тренда и построения остаточных полей, что позволяло сфокусироваться на компонентах, связанных с глубинными уровнями и структурами, сопоставимыми с масштабами источников и контролирующих элементов для эпитермальных и субэпитермальных систем. Высокочастотные компоненты полей анализировались через горизонтальные градиенты и производные по модулю, включая карты горизонтального градиентного модуля (HGM), которые подчеркивают резкие латеральные контрасты и часто используются как индикаторы границ блоков, разломных зон и контактов доменов, способных контролировать локализацию рудных систем. Дополнительным элементом признакового стека выступала карта комбинированного геомеханического индикатора, характеризующего распределение ориентированных напряжений и тенденции к раскрытию или сдвиговой активации разломов в зависимости от их ориентации в поле напряжений, что позволяет учитывать не только геометрию структур, но и их потенциальную гидравлическую “работоспособность” как путей флюидопереноса.

Отдельный блок исходных данных формировался полем теплового потока, полученным на основе региональных оценок по данным глубоких скважин, геотермических исследований и теплового моделирования [91-92]. Тепловой поток рассматривался как физически обоснованный прокси-параметр, отражающий интегральный эффект глубинных тепловых источников и долгоживущего термального режима коры, который определяет возможности поддержания гидротермальных систем и интенсивность теплопереноса, необходимого для формирования эпитермальных Au-Ag систем. Повышенные значения теплового потока в сочетании с благоприятной структурной обстановкой интерпретируются как индикатор повышенной вероятности магматически или мантийно-корово контролируемых тепловых аномалий, тогда как участки с относительно пониженным тепловым потоком при наличии выраженной разломной проницаемости могут соответствовать системам, где доминирует циркуляция метеорных вод и роль магматического тепла является менее выраженной. В практическом смысле данный блок предикторов вводит в

модель компоненту “энергетики” минерал-системы, дополняя геолого-структурные и геофизические признаки. Как показано на рисунке 4.3, пространственное распределение остаточных гравитационных и псевдогравитационных аномалий, их производных и теплового потока демонстрирует согласованные зоны, в которых формируются потенциально благоприятные комбинации глубинной структуры, доменной архитектуры и термального режима, релевантные Au-Ag минерализации.

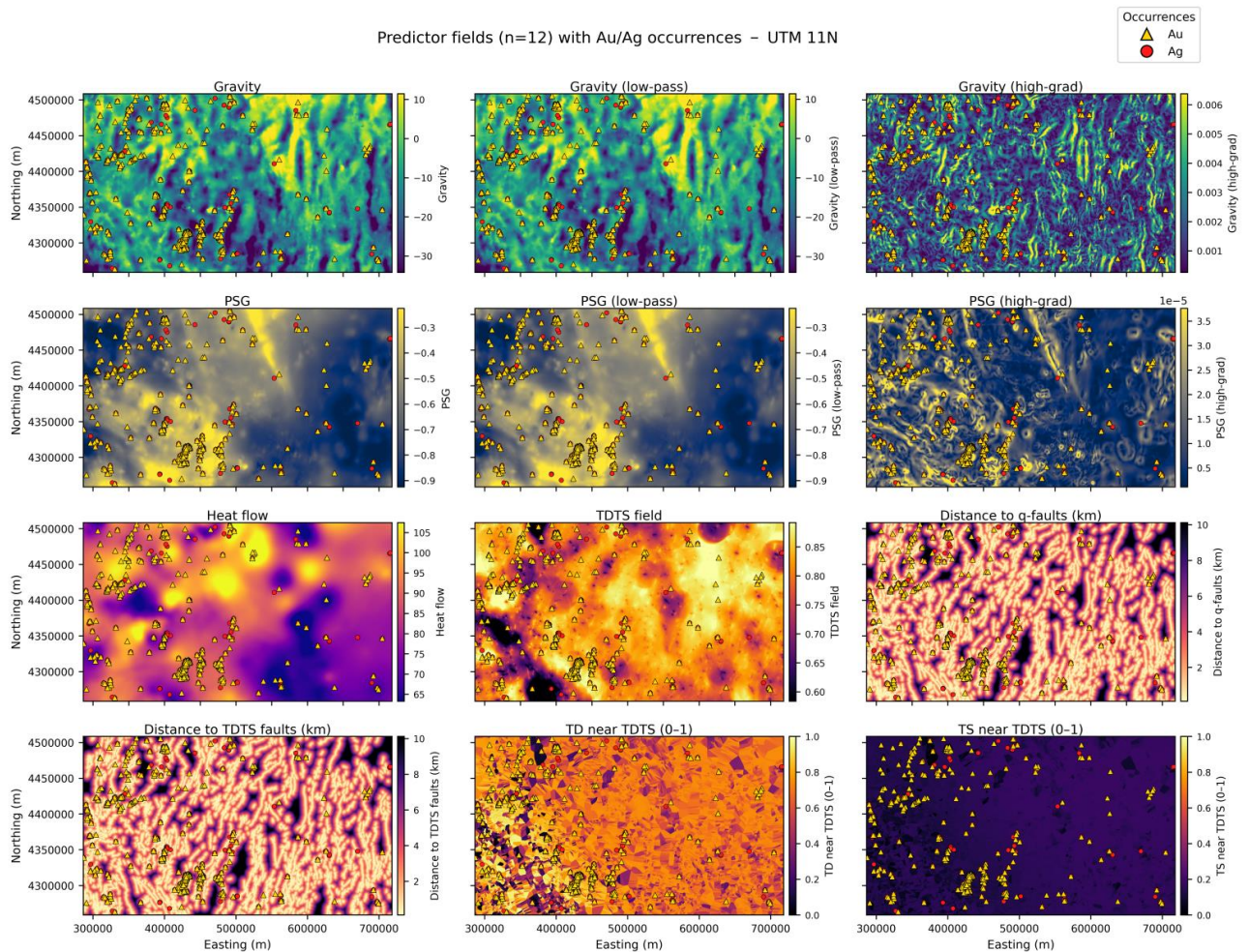


Рисунок 4.3 - Базовые растровые предикторы: остаточные поля плотностных и псевдогравитационных аномалий, их производные и тепловой поток в пределах области моделирования

Все пространственные слои были приведены к единой расчетной сетке с шагом 250 м и унифицированной системе координат, что обеспечило корректное совместное использование гетерогенных предикторов в рамках региональной задачи МРМ. Применение такого шага сетки является оправданным именно для регионального минерагенического моделирования, поскольку 250 м обеспечивает баланс между вычислительной эффективностью, совместимостью разнородных источников и сохранением информативности сигналов, чье фактическое пространственное разрешение определяется нативной дискретизацией исходных наборов данных. При приведении данных

учитывались различия в исходных пространственных разрешениях: гравиметрические и магнитные компиляции характеризуются относительно регулярной дискретизацией на масштабе сотен метров - первых километров, тогда как тепловой поток и отдельные структурно-геомеханические атрибуты имеют существенно более редкую и неоднородную поддержку, требующую аккуратной интерполяции и контроля сглаживания [89-92]. По этой причине было принципиально зафиксировано, что использование сетки 250 м служит прежде всего для гармонизации совместного анализа и формирования единого признакового стека, но не трактуется как искусственное “повышение” исходного разрешения там, где оно физически ограничено поставщиком данных и методами построения первичных поверхностей. Соответственно, применялись гармонизированные процедуры ресэмплинга и сглаживания, согласованные с общими рекомендациями по региональной обработке потенциальных полей и картированию структурных параметров в задачах минерагенического районирования [90], при обязательном учете ограничений валидности каждого набора.

На базе унифицированных растровых слоев были сформированы несколько семейств признаков, каждое из которых отражает отдельный аспект минерал-системной модели Au-Ag и, одновременно, обеспечивает статистическую различимость перспективных доменов в рамках дальнейшего машинного обучения. К семейству полей потенциальных типов относятся остаточные гравиметрические и псевдогравитационные аномалии, их горизонтальные градиенты, а также фильтрованные компоненты, чувствительные к блочно-структурной архитектуре коры, глубинным контактам и протяженным границам доменов. Структурные предикторы включают расстояния до разломов различных рангов, плотность пересечений разломов и геомеханические показатели, характеризующие относительные сдвиговые и растягивающие компоненты напряжений вблизи активных структур [88-90]. Тепловой блок представлен абсолютными значениями теплового потока и его отклонениями от регионального уровня, что имеет ключевое значение для учета пространственной изменчивости глубинного теплового режима и потенциала поддержания гидротермальных систем.

Каждый предиктор сопровождался сведениями об исходном наборе данных, нативном масштабе и разрешении, а также о методе перепроецирования или ресемплинга, примененном при приведении к единой сетке 250 м. Сводная характеристика указанных семейств представлена в таблице 4.2, которая служит структурной опорой для последующих подразделов главы 4: именно эти группы формируют конечный признаковый стек, используемый в PU-learning, АОА-ограничении и построении карт перспективности.

Таблица 4.2. Семейства предикторов, их геолого-геофизическое обоснование и исходные наборы данных, использованные в региональной модели МРМ для Au-Ag в пределах АОI (все слои приведены к базовой сетке 250 м; непрерывные поля ресэмплированы билинейно, дистанционные и категориальные - методом ближайшего соседа).

Группа предикторов	Переменная	Геолого-геофизический смысл	Источник данных	Нативное разрешение / тип данных	Приведение к аналитической сетке (250 м)	Ограничения и валидность
Потенциальные поля	Gravity_res	Остаточное гравитационное поле, отражающее крупные контрасты плотности и глубинную архитектуру	Regional isostatic gravity compilation (Glen et al., 2021)	Растр; эффективная поддержка ~0.5-1 км	Билинейный ресэмплинг; медианное центрирование и масштабирование по IQR	Сетка 250 м используется для гармонизации слоев и не повышает фактическое разрешение; спектральный диапазон сигнала задается исходной сеткой
Потенциальные поля	Gravity_HGM	Горизонтальная величина градиента гравитации, выделяющая структурные границы и контакты	Производная от гравитационной аномалии	Растр; эффективная поддержка ~0.5-1 км	HGM рассчитывается на нативной сетке, затем билинейный ресэмплинг; IQR-нормализация	Детализация градиента ограничена исходным шагом данных; структуры меньшего масштаба, чем нативная поддержка, не могут быть надежно восстановлены
Потенциальные поля	Gravity_match	Гравитация после match-filter, акцентирующая структуры заданного глубинного/масштабного диапазона	Match-filtered gravity grids (Glen et al., 2021)	Растр; масштаб сигнала задан параметрами фильтра	Билинейный ресэмплинг; параметры фильтрации сохраняются	Фильтрация фиксирует пространственную полосу сигнала; “уплотнение” до 250 м не добавляет информации и служит только для совместного анализа

						слоев
Потенциальные поля	PseudoG_res	Остаточная псевдогравитация для совместной интерпретации магнитных и плотностных контрастов	Regional magnetic compilation → pseudogravity	Растр; эффективная поддержка ~0.5-1 км	Билинейный ресэмплинг; медианное центрирование и IQR-нормализация	Ограничения аналогичны гравитации: гармонизация без прироста фактического разрешения; качество зависит от исходной магнитной съемки и преобразования
Потенциальные поля	PseudoG_HGM	HGM псевдогравитации, выделяющий границы магнитных источников	Производная от псевдогравитации и	Растр; эффективная поддержка ~0.5-1 км	HGM рассчитывается на нативной сетке, затем ресэмплинг; используется как непрерывный признак	Контроль структуры определяется нативной поддержкой магнитных данных, а не аналитической сеткой 250 м
Потенциальные поля	PseudoG_match	Псевдогравитация после match-filter, акцентирующая магнитные источники на целевых глубинах	Match-filtered pseudogravity grids	Растр; масштаб сигнала задан фильтром	Билинейный ресэмплинг; параметры фильтрации неизменны	250 м служит только для гармонизации; детализация ограничена исходными данными и параметрами фильтра
Тепловой режим	HeatFlow	Интерполированная поверхность теплового потока, отражающая региональные тепловые аномалии	NVML heat-flow surface (DeAngelo & Coolbaugh, 2021)	Растр; эффективная поддержка ≥5-10 км	Билинейный ресэмплинг; медианное центрирование и IQR-масштабирование	Сильное “переуплотнение” при 250 м: физически значимые вариации проявляются на км-масштабах; слой ожидаемо гладкий
Структурная	Dist_Qfault	Расстояние до	Quaternary faults	Вектор;	Евклидова дистанция в	Точность

близость		четвертичных разломов как мера структурного контроля и близости к активным/молодым разломным зонам	(Siler & DeAngelo, 2023)	картирование ~1:250 000-1:500 000	проекция (EPSG:32611) до ближайшей геометрии разлома; значение присваивается центрам ячеек	определяется масштабом картирования и позиционной неопределенностью разломов, а не размером ячейки 250 м
Структурная близость	Dist_TDTS_fault	Расстояние до разломов, параметризованных TD/TS, как мера близости к механически благоприятным структурам	NV TD/TS faults (Siler & DeAngelo, 2023)	Вектор; масштаб ~1:250 000-1:500 000	Евклидова дистанция до ближайшего сегмента; сохраняется индекс ближайшего сегмента для привязки атрибутов	Влияние ошибки положения разлома напрямую переносится на признак; повышение плотности сетки не улучшает исходную геометрию
Структурная близость	TD_near_tdts_fault	Дилатационная тенденция ближайшего разлома (TD) как индикатор потенциальной раскрытости/проницаемости	TD attributes (Siler & DeAngelo, 2023)	Векторные атрибуты разломов	Перенос значения TD от ближайшего сегмента (nearest neighbor) к центру ячейки	Эффективная поддержка задается точностью картирования разлома; локальные изменения TD внутри сегмента могут быть неразрешимы
Структурная близость	TS_near_tdts_fault	Сдвиговая тенденция ближайшего разлома (TS) как индикатор близости к критическому сдвигу	TS attributes (Siler & DeAngelo, 2023)	Векторные атрибуты разломов	Перенос значения TS от ближайшего сегмента (nearest neighbor) к центру ячейки	Ограничения аналогичны TD: точность определяется качеством/масштабом разломной базы и сегментацией линий
Структурная близость	TDTS_raster	Растровый индикатор TD×TS как прокси механической	TDTS raster (Siler & DeAngelo, 2023)	Растр; нативная поддержка	Билинейный ресэмплинг; стандартизация/нормализация	Фактическое разрешение ограничено

		благоприятности структуры		~0.5-1 км (по поставщику)		исходным растром и неопределенностью разломной геометрии; 250 м используется для совместимости слоев
Региональная геология	Geology500k (при использовании)	Литологические домены и крупные структуры как контекст для металлогении и пространственной сегментации	Vector geology map 1:500 000	Вектор; эффективная поддержка ~1-2 км	Растрезация на сетку 250 м методом nearest neighbor	Разрешение и точность определяются картой 1:500 000; границы на 250 м следует трактовать как условные, без субкм детализации

В совокупности описанные геологические, структурные, геофизические и тепловые слои формируют согласованную минерал-системную основу для последующего статистического моделирования перспективности Au-Ag. Геологическая карта и опорные месторождения задают контекст, в котором интерпретируются остальные предикторы; структурный каркас и его геомеханические производные описывают конфигурацию зон повышенной проницаемости и потенциальных путей флюидопереноса; поля остаточных плотностных и псевдогравитационных аномалий в сочетании с тепловым потоком отражают глубинную архитектуру и долговременный термальный режим коры. Приведение разнородных слоев к единой расчетной сетке и их организация в семейства признаков, суммарно представленные в таблице 4.2, создают технически корректную и геологически интерпретируемую отправную точку для применения PU-learning и ансамблевых моделей перспективности. В следующем подразделе на этой базе формулируется выборка позитивов и ненаблюдаемых ячеек, вводится строго пространственная схема валидации и описывается протокол построения региональной MPM модели для Au-Ag Великого Бассейна.

4.2 PU-learning в MPM, строго пространственная GroupKFold-валидация, OOF-инференс

Региональная задача прогнозирования минерагенической перспективности Au-Ag в пределах Великого Бассейна по своей природе характеризуется принципиально неполной и неоднородной разметкой. Достоверно известными являются лишь точки, где наличие минерализации подтверждено результатами разведочных работ, эксплуатационного бурения, геохимических опробований и сопутствующих исследований. При этом значительная часть территории либо остается малоизученной, либо содержит только косвенные признаки металлогенической активности, которые, даже будучи геологически содержательными, не могут трактоваться как прямое доказательство отсутствия минерализации. В таких условиях традиционное приближение, в котором все неотмеченные ячейки регулярной сетки автоматически приравниваются к отрицательному классу, приводит к методически некорректному усилению доли “чистого фона” и к систематическому смещению вероятностных оценок в сторону занижения перспективности, особенно для слабо оконтуренных или “частично проявленных” рудных систем [90, 91]. Для региональных MPM задач этот эффект критичен, поскольку пространственная кластеризация известных рудных объектов и геологическая неравномерность изученности провинции создают ситуацию, при которой “отсутствие метки” преимущественно отражает дефицит информации, а не реальное отсутствие минерализации.

В настоящей работе для построения региональной модели MPM была использована постановка обучения с позитивами и неразмеченным фоном (PU-learning), предназначенная именно для задач, где наблюдаемыми являются лишь позитивы, а остальная совокупность объектов представляет собой смесь

потенциально перспективных и действительно непродуктивных участков [90-92]. В рамках принятой постановки известные Au-Ag occurrences интерпретируются как положительный класс, тогда как все остальные ячейки внутри области интереса трактуются как неразмеченный класс, включающий одновременно фоновые и потенциально минерализованные домены. Такое допущение согласуется с практикой регионального прогнозирования: “U” не интерпретируется как отрицание минерализации, а рассматривается как пространство неопределенности, в котором перспективные объекты могут присутствовать, но еще не выявлены.

На первом этапе формировалась обучающая выборка в терминах P и U. Класс P включал координаты рудных месторождений, рудопроявлений и надежных геохимических аномалий, вошедших в сводный каталог по Au и Ag для рассматриваемой области. При формировании P учитывалась типичная для провинций эпитепмального оруденения пространственная кластеризация проявлений вокруг крупных рудных полей и структурных узлов. Чтобы снизить влияние переизбыточных локальных сгущений меток и исключить искусственное завышение качества модели, применялся радиус “разрежения” (declustering / thinning) позитивов: близко расположенные occurrences агрегировались таким образом, чтобы внутри заданного радиуса оставался один представитель. Эта процедура уменьшает вероятность того, что несколько практически совпадающих по геологическому смыслу позитивных ячеек окажутся разнесенными по разным фолдам перекрестной проверки и тем самым создадут эффект “пространственной утечки” при оценке качества. Класс U формировался как совокупность всех ячеек регулярной сетки внутри области интереса, которые удовлетворяют критерию предсказательной применимости АОА (описанному в подразделе 4.3) и при этом не относятся к классу P. Тем самым U включает и действительно “чистый фон”, и участки, в которых минерализация потенциально возможна, однако не была зафиксирована существующими данными.

Пространственная структура обучающих данных в задачах MPM требует специальной схемы валидации, исключающей статистически ложное завышение качества. Использование стандартной случайной k-fold перекрестной проверки в региональном пространственном прогнозе обычно приводит к информационной утечке, поскольку соседние ячейки по координатам и по значениям предикторов оказываются одновременно в обучении и тесте, а оценка качества в таком случае отражает преимущественно локальную интерполяцию в пределах уже известных рудных полей, а не способность модели переноситься на новые участки региона [93, 94]. В настоящей работе была реализована строго пространственная схема GroupKFold, в которой каждой ячейке сетки сопоставлялась принадлежность к одному из пространственных блоков, а блоки поочередно использовались в качестве тестовых фолдов. Размер и геометрия блоков выбирались сопоставимыми с характерными масштабами металлогенических зон и пространственной корреляции предикторов, чтобы тестирование действительно

имитировало прогноз “на новое”. Вокруг каждого тестового блока вводилась буферная зона исключения (R_{excl}), внутри которой ячейки не допускались в обучающую выборку. Буфер использовался для устранения эффекта “подглядывания” на границах тестовой области, где пространственная автокорреляция предикторов и геологическая непрерывность структур часто позволяют модели фактически использовать информацию о тестовом блоке через его ближайшее окружение. В итоге получалась конфигурация, обеспечивающая пространственно независимую оценку: модель обучалась на части региона и проверялась на пространственно отделенных сегментах, что концептуально соответствует практическому сценарию переноса результатов на новые потенциальные участки поисков.

Как показано на рисунке 4.4, пространство признаков для Ag и Au было разбито на пять пространственных фолдов, каждый из которых представляет связную область, покрывающую значимую долю occurrences и фоновых интервалов. На каждом шаге перекрестной проверки один фолд полностью резервировался для тестирования, вокруг него задавался буфер исключения, а обучение выполнялось на оставшихся фолдах, но только в пределах зоны АОА. Такой протокол делает метрики чувствительными к реальной пространственной организации рудных полей и позволяет оценить способность модели выделять перспективные зоны вдали от уже известных объектов, а не только в их непосредственном окружении [93]. Дополнительно данная схема верификации соответствует логике регионального таргетинга: значимым считается не абсолютное качество классификации “ячейка P против U”, а способность модели концентрировать известные проявления в верхней части ранжирования в условиях строгой пространственной независимости.

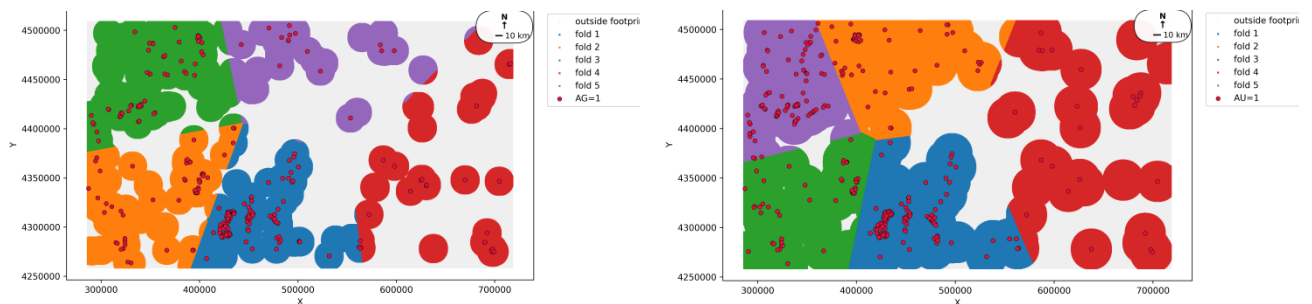


Рисунок 4.4 - Схема пространственной GroupKFold-валидации для моделей Au и Ag. Цветными полигонами показаны блоки, используемые поочередно в качестве обучающих и тестовых, кружками отмечены позитивные точки P, а штриховкой и окружностями проиллюстрированы буферные зоны R_{excl} , внутри которых точки исключаются из обучающей выборки при оценке качества модели

В рамках PU-learning ключевым методическим шагом является оценка доли размеченных позитивов среди всех действительно перспективных ячеек, то есть оценка так называемой константы отбора (labeling propensity, selection constant). В исходной формулировке Elkan и Noto процедура строится на

промежуточной модели, обучаемой различать P и U как два формальных класса [90]. Предсказанные вероятности этой модели интерпретируются как вероятность того, что перспективная ячейка была “отобрана” и получила метку P. Усреднение этих величин по наблюдаемым позитивам дает оценку константы отбора, после чего выполняется корректировка апостериорных вероятностей принадлежности к перспективному классу. Соответствующее соотношение приведено в выражении (4.1), где формализована связь между промежуточной моделью отбора и целевой PU-вероятностью.

$$\hat{c} = \frac{1}{|P|} \sum_{i \in P} \hat{h}(x_i), \hat{p}_{EN}(x) = \min \left(\frac{\hat{h}(x)}{\hat{c}}, 1 \right). \quad (4.1)$$

Оценка константы отбора и последующая корректировка апостериорных вероятностей необходимы для перевода выходов модели в интерпретируемый вероятностный масштаб, в котором значения в диапазоне 0-1 действительно отражают относительную долю перспективных ячеек при сопоставимых условиях признакового пространства [91-92]. В геологоразведочном контексте это означает, что среднее значение предсказанной вероятности по выбранному подмножеству территории может интерпретироваться как ожидаемая доля участков с наличием Au-Ag минерализации при условии, что изучаемые ячейки не являются экстремальной экстраполяцией относительно обучающей выборки. Такая трактовка особенно важна для последующего анализа top-k целей, поскольку количество “ожидаемых попаданий” в пределах первых процентов площади напрямую вычисляется на основе калиброванных вероятностей, а не на основе условных скорингов.

Следующий элемент методики был связан с выбором базового алгоритма и настройкой весов классов с учетом оцененного приора перспективного класса. В качестве базовой модели был использован ансамбль случайных лесов, что является оправданным для региональных MPM задач по нескольким причинам. Во-первых, данный алгоритм устойчив к частичной коллинеарности предикторов и к наличию шумовых признаков, неизбежных при интеграции разномасштабных геофизических и структурных слоев. Во-вторых, случайный лес относительно стабилен к выбросам и локальным аномалиям и способен выявлять нелинейные взаимодействия предикторов, характерные для минерал-системных условий. В-третьих, совместимость с процедурой бэггинга позволяет далее оценивать неопределенность и устойчивость ранжирования без введения сложных байесовских конструкций. Для каждого пространственного фолда строился отдельный ансамбль, в котором положительные примеры получали повышенный вес в функции потерь, а ячейки U получали пониженный вес, определяемый оцененным приором перспективного класса. Такой подход позволяет согласовать “риск”, оцениваемый моделью, с предположительной долей перспективных объектов, не прибегая к ручным эвристикам по каждому предиктору [91]. В результате модель не сводится к поиску единичных “экстремальных” аномалий, а сохраняет способность выявлять протяженные зоны умеренной, но систематической аномальности,

которые для Au-Ag систем Великого Бассейна нередко оказываются геологически более значимыми, чем изолированные пиковые отклики.

Для корректной оценки качества и последующей калибровки вероятностей использовались строго out-of-fold предсказания. В каждом пространственном фолде модель обучалась только на обучающих блоках (вне тестового блока и буфера), а затем применялась к ячейкам тестового фолда, находящимся внутри АОА. Полученные по всем фолдам предсказания объединялись в единый набор out-of-fold оценок, который далее служил основой для расчета метрик ранжирования и вероятностного качества (AP, ROC, Brier), а также для построения калибровочных кривых. Формальное определение такого объединенного множества приведено в выражении (4.2), где тестовые индексы каждого фолда объединяются в единый пул OOF-предсказаний.

$$D_{OOF} = \bigcup_{k=1}^K \{(y_i, s_k(x_i)) : i \in F_k\} \quad (4.2)$$

Использование именно out-of-fold оценок является принципиальным для региональной МРМ, поскольку эти предсказания никогда не формировались на тех же данных, на которых обучался соответствующий ансамбль, и при этом сохраняют пространственную независимость благодаря GroupKFold-разбиению и буферным зонам. Поэтому статистики, рассчитанные по OOF-набору, отражают ожидаемое поведение модели на новых, ранее не изученных участках региона, а не внутри уже оконтуренных рудных полей. В последующих подразделах именно этот набор используется для построения области применимости АОА, выбора схемы калибровки вероятностей и оценки устойчивости top-k ранжирования.

С геолого-практической точки зрения комбинация PU-learning и строго пространственной GroupKFold-валидации позволяет приблизить постановку задачи к реальному процессу регионального прогноза. Модель обучается на участках, где факт присутствия Au-Ag минерализации известен, но оценивается на пространственно отделенных блоках, которые репрезентируют потенциальные зоны поисков и бурения вне текущих “ядер” изученности. Корректировка приора через PU-схему компенсирует неполноту разведки и снижает риск ошибочного отнесения “неизвестного” к “отрицательному”. Использование только OOF-предсказаний обеспечивает честную вероятностную интерпретацию и подготавливает основу для последующих процедур АОА-ограничения, калибровки и decision-focused top-k навигации. В результате карты перспективности и неопределенности, получаемые далее, могут трактоваться не как простая “перерисовка” известных рудных полей, а как статистически обоснованный инструмент приоритизации региональных участков для дальнейших поисковых работ.

4.3 АОА-маскирование (индекс диссимилярности), калибровка вероятностей, метрики

В данном подразделе завершается формализация связки между обучением PU-модели перспективности и корректной областью ее применения как в признаковом, так и в географическом пространстве. Ключевой тезис состоит в том, что любая вероятностная оценка перспективности имеет интерпретируемый смысл только для тех ячеек, где комбинации геологических, геофизических, структурных и тепловых предикторов сопоставимы с условиями, представленными в обучающей выборке. В противном случае прогноз становится экстраполяционным и отражает поведение алгоритма в условиях, которые не наблюдались при обучении, что резко повышает риск неверной геологической интерпретации и некорректной оценочной статистики. Для явного контроля этого аспекта вводится индекс диссимилярности в пространстве признаков и область применимости АОА, внутри которой выполняются и оценка качества, и вероятностная калибровка. Подход концептуально близок современным методикам оценки области применимости для пространственных моделей в геонауках, где требуется отделять интерполяционный прогноз от экстраполяции в многомерном пространстве факторов [95].

Индекс диссимилярности $D(x)$ определяется как расстояние от вектора признаков ячейки x до ближайшего обучающего объекта в многомерном стандартизованном пространстве, рассчитанное в метрике Махаланобиса. Формально эта величина задается выражением (4.3), в котором используется матрица ковариации признаков, оцененная по обучающему множеству в пределах очередного *spatial fold*. При малых значениях $D(x)$ комбинация признаков данной ячейки близка к наблюдавшимся сочетаниям остаточных гравитационных и псевдогравитационных полей, их градиентных производных, теплового потока, а также структурных дистанционных и геомеханических предикторов. Напротив, большие значения $D(x)$ означают, что ячейка лежит в области признакового пространства, где модель вынуждена опираться на сочетания предикторов, отсутствующие в обучении, и потому прогноз приобретает экстраполяционный характер, что особенно нежелательно в задачах регионального таргетинга.

$$D(x) = \min_{z_i \in Z_r} \|z(x) - z_i\|_M, \|g\|_M = \sqrt{g^T M g}, \quad (4.3)$$

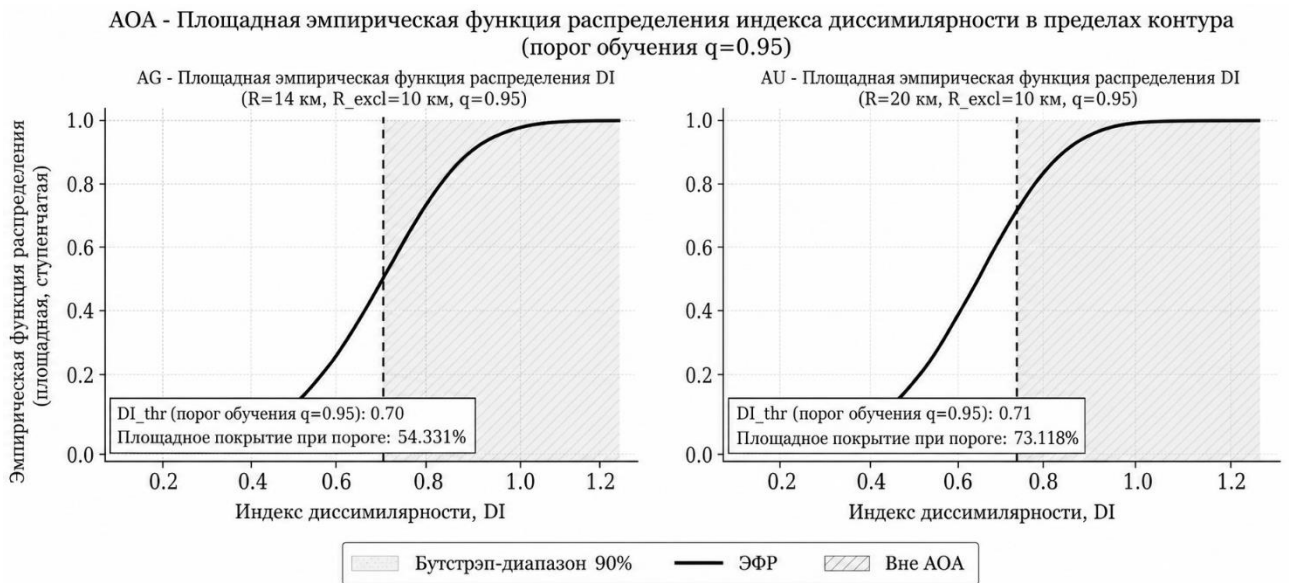


Рисунок 4.5 - Эмпирические распределения индекса диссимиларности и пороговые значения области применимости для моделей перспективности серебра и золота

Практическая реализация АОА основывается на выборе порогового значения индекса диссимиларности на основе обучающих данных. В выражении (4.4) порог $t_{0.95}$ задается как 95-й квантиль распределения $D(x)$, рассчитанного по обучающим ячейкам в рамках схемы скользящего исключения.

$$t_{0.95} = \text{quantile}_{0.95} \left\{ \min_{j \neq i} \|z_i - z_j\|_M : z_i \in Z_{tr} \right\}, \quad (4.4)$$

Такой выбор соответствует требованию, чтобы не менее 95 % обучающих объектов имели индекс диссимиларности ниже порога, то есть лежали внутри АОА. Далее область применимости определяется как множество ячеек, удовлетворяющих неравенству $D(x) \leq t_{0.95}$, что формально записано в определении (4.5):

$$АОА = \{x : D(x) \leq t_{0.95}\}. \quad (4.5)$$

Таким образом, АОА задается в пространстве признаков, а затем отображается в географическое пространство в виде маски, отсекающей те части полигона, где исходные комбинации предикторов существенно отличаются от обучающей выборки. Для региональной МРМ это является принципиальным, поскольку разномасштабность исходных данных, неоднородность изученности и пространственная изменчивость геологических доменов способны создавать локальные “экстремальные” сочетания предикторов, для которых статистическая модель фактически не обучалась. Маска АОА в данном случае выступает не просто техническим фильтром, а формализованной геостатистической гарантией того, что далее оценивается и интерпретируется именно область, где модель применима.

Следующим шагом оценивалась устойчивость АОА относительно пространственной перекрестной проверки. Для каждого spatial fold матрица ковариации и порог $t_{0.95}$ переоценивались только по обучающим данным

данного фолда, после чего рассчитывалась доля ячеек, попадающих в АОА в пределах тестового блока. Анализ покрытия АОА по фолдам позволяет убедиться, что область применимости не “проседает” в отдельных частях Великого Бассейна и задает сопоставимые условия для оценки качества разных моделей и схем калибровки. Как показано на рисунке 4.6, доля площади, для которой выполняется условие $D(x) \leq t_{0,95}$, остается достаточно стабильной по фолдам как для серебра, так и для золота, что указывает на консистентность критерия диссимильности и отсутствие выраженных “провалов” применимости в отдельных пространственных блоках.

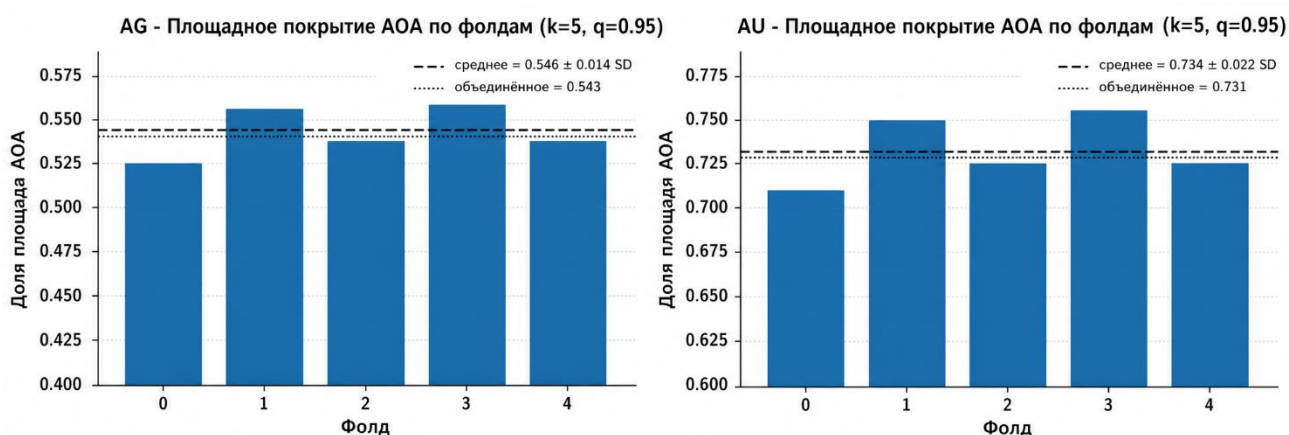


Рисунок 4.6 - Покрытие области применимости по пространственным фолдам для моделей перспективности серебра и золота.

Все последующие шаги валидации и калибровки вероятностей выполнялись строго внутри АОА. В подразделе 4.2 были получены out-of-fold оценки перспективности для каждой ячейки полигона, рассчитанные с помощью PU-обученного случайного леса при строгой пространственной схеме GroupKFold. На данном этапе OOF-значения рассматривались как некалиброванные оценки “силы сигнала” модели, пригодные для ранжирования, но требующие перевода в вероятности, согласованные с частотной интерпретацией. Для ансамблевых методов, включая случайный лес, типично высокое качество ранжирования при одновременном смещении вероятностных оценок: вероятности могут систематически “сжиматься” к средним значениям или, напротив, демонстрировать локальные переоценки в отдельных диапазонах, что снижает корректность интерпретации и расчета ожидаемых попаданий в top-k цели [98, 99]. Поэтому вероятностная калибровка рассматривается как обязательный этап методики, особенно в задачах decision-focused таргетинга, где ожидаемое число успехов должно вытекать из вероятностной модели, а не из условного скоринга.

Для устранения систематических искажений применялись стандартные монотонные калибраторы вероятностей, широко используемые в бинарной классификации. В настоящем исследовании были протестированы по крайней мере два подхода: логистическая калибровка (Platt scaling) и изотоническая

регрессия [98, 99]. Оба метода строят одномерное преобразование между некалиброванным скором модели и оценкой вероятности. Изотоническая регрессия, будучи непараметрической и монотонной, способна адаптироваться к более сложной форме зависимости между скором и истинной частотой позитивов, тогда как логистическая калибровка задает параметрическую S-образную кривую и может быть менее гибкой в области высоких вероятностей. Выбор окончательного калибратора выполнялся по критерию минимизации Brier-метрики, применяемой как интегральная мера качества вероятностных прогнозов и одновременно как чувствительный индикатор калибровочных смещений [96, 97].

Вычислительный протокол был построен таким образом, чтобы исключить утечку информации. Все калибраторы обучались исключительно на OOF-предсказаниях внутри АОА, то есть на значениях, которые были получены для ячеек, не участвовавших в обучении соответствующего фолдового ансамбля. Это является критичным, поскольку калибровка, выполненная на тех же данных, на которых обучалась модель, неизбежно приводит к завышению качества и к ложному ощущению вероятностной корректности. Использование out-of-fold схемы гарантирует, что для каждой ячейки калибровщик “видел” только те наблюдения, которые были получены независимо от нее, а значит итоговая калибровка отражает поведение модели в условиях, наиболее близких к применению на новых участках. После подбора калибратора по минимальному pooled OOF Brier-score выполнялась перенормировка всех OOF-оценок перспективности в вероятности, интерпретируемые как ожидаемая доля рудных проявлений в соответствующих группах ячеек при сопоставимом признаковом окружении.

Характер калибровки иллюстрируется диаграммами надежности и калибровочными кривыми, представленными на рисунке 4.7. По оси абсцисс откладываются интервалы прогнозируемых вероятностей, тогда как по оси ординат приводится наблюдаемая доля позитивных объектов в соответствующих группах. Для некалиброванных оценок случайного леса обычно наблюдается отклонение от диагонали, отражающее тенденцию к сглаживанию и недостаточному разделению “высоких” и “низких” вероятностей, что особенно заметно в задачах с сильным дисбалансом классов и пространственной неоднородностью [98, 99]. После применения изотонической калибровки кривая приближается к диагонали для обоих металлов, а интервалы доверия, рассчитанные по Уилсону, указывают на статистически устойчивое согласование прогнозируемых и наблюдаемых частот в широком диапазоне вероятностей. Логистическая калибровка, напротив, может демонстрировать более заметные отклонения в области высоких вероятностей, что отражается и в несколько более высоких значениях Brier-score в pooled OOF-оценке.

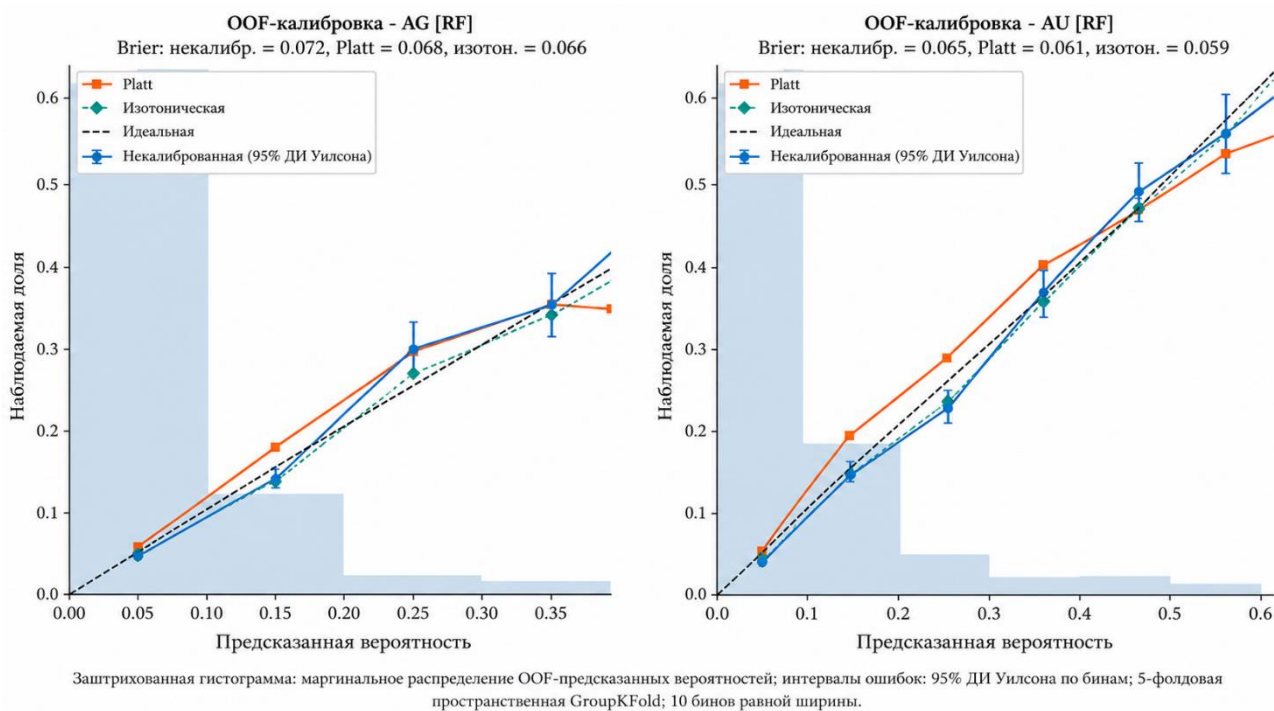


Рисунок 4.7. Диаграммы надежности и калибровочные кривые для некалиброванных и калиброванных out-of-fold вероятностей перспективности серебра и золота в пределах области применимости.

Формальное сравнение вариантов калибровки и итогового качества моделей представлено в таблице 4.2, где собраны показатели пространственной перекрестной проверки и pooled OOF-оценки для серебра и золота. Для каждого металла приводятся значения средней точности по кривой “точность-полнота” (AP), площади под ROC-кривой и Brier-score после применения выбранного калибратора. При сильном дисбалансе классов, характерном для задач MPM, именно AP рассматривается как основная метрика, поскольку она непосредственно оценивает способность модели концентрировать известные проявления в верхней части ранжированного списка ячеек, что методически соответствует логике таргетинга. ROC-метрика, будучи менее чувствительной к дисбалансу, используется как дополнительная характеристика общей разделяющей способности. Brier-score, в свою очередь, суммирует ошибки дискриминации и калибровки и служит ключевым критерием выбора оптимального вероятностного преобразования [96, 97].

Таблица 4.3 - Результаты пространственной перекрестной проверки и калибровки out-of-fold вероятностей для серебра и золота в пределах области применимости (AP, ROC, Brier).

Мет-ка	AP Fold1	AP Fold 2	AP Fold3	AP Fold4	AP Fold5	Среднее AP	Совмещенный OOF AP	ROC Fold1	ROC Fold2	ROC Fold3	ROC Fold4	ROC Fold5	Среднее ROC	Совмещенный OOF ROC	Brier
Ag	0.757	0.758	0.767	0.714	0.749	0.749	0.745	0.903	0.923	0.909	0.901	0.934	0.914	0.908	0.066
Au	0.796	0.744	0.784	0.865	0.758	0.789	0.777	0.910	0.910	0.923	0.957	0.889	0.918	0.912	0.059

Важно подчеркнуть, что все численные показатели, приведенные в таблице 4.2 и на рисунках 4.5-4.7, рассчитывались только для ячеек, принадлежащих АОА. Это обеспечивает строгую согласованность между геометрией пространства признаков, схемой spatial GroupKFold и вероятностной калибровкой: модель оценивается и интерпретируется исключительно там, где она реально опирается на наблюдавшиеся комбинации геологических, геофизических и тепловых условий. В совокупности индекс диссимilarity и маска АОА, калибровка по OOF-предсказаниям и система метрик AP/ROC/Brier формируют строгий протокол контроля качества, который в последующих подразделах 4.4 и 4.5 используется для анализа карт перспективности, пространственной неопределенности и для интерпретации вклада главных предикторов в терминах минерал-системной модели Au-Ag Великого Бассейна.

4.4 Карты неопределенности, риск-ориентированное ранжирование и top-k цели

В настоящем подразделе показано, каким образом калиброванные вероятностные предсказания, полученные в рамках PU-learning и строго пространственной GroupKFold-валидации, переводятся из статистической постановки в практико-ориентированный инструмент регионального таргетинга. Основное внимание уделяется (i) построению карт перспективности с учетом области применимости (АОА), (ii) явной оценке модельной неопределенности как обязательному компоненту интерпретации прогноза, (iii) формированию риск-ориентированного ранжирования целей в терминах top-k доли площади, а также (iv) проверке устойчивости решений при разумных вариациях априорных допущений и при независимых сценариях тестирования. Такой протокол согласуется с современными подходами к использованию прогностических карт в минерагии и МРМ, где оценка качества не исчерпывается ранжированием, а дополняется вероятностной калибровкой, контролем применимости модели и явной демонстрацией выигрыша в эффективности top-k таргетинга по сравнению со случайным или интуитивным распределением поисковых работ [100, 101].

После введения области применимости АОА в подразделе 4.3 калиброванные вероятности перспективности для серебра и золота были визуализированы в виде отдельных карт, ограниченных только теми ячейками, для которых индекс диссимilarity $D(x)$ не превышает порог $t_{0,95}$. Соответствующие карты для Ag и Au приведены на рисунках 4.8 и 4.9. Визуально различимые паттерны пространственной концентрации повышенных вероятностей интерпретируются как отражение различий в доминирующих факторах рудоконтроля и/или различий в составе и представительности позитивов в каталогах для Ag и Au при общей минерал-системной рамке. Для серебра наиболее высокие значения прогнозной вероятности формируют протяженные зоны, вытянутые по преимущественным структурным направлениям, что согласуется с ролью разломно-трещинных систем в

обеспечении повышенной проницаемости и флюидопереноса. Для золота области повышенной вероятности, как правило, выражены в зонах сочетания повышенных градиентов гравитационного и псевдогравитационного поля с локально повышенным тепловым потоком, что может отражать более выраженную связь золотоносных систем с термальными аномалиями и глубинной архитектурой источников и проводников флюидов. Принципиально важным элементом карт 4.8 и 4.9 является явное отображение АОА-маски: за пределами области применимости ячейки полностью исключены из интерпретации, что предотвращает некорректное использование модели в геологическом пространстве, существенно отличающемся от условий, представленных в обучающей выборке [102]. Тем самым АОА выступает не декоративным ограничением, а формализованной дисциплиной интерпретации, отделяющей интерполяционный прогноз от экстраполяции в пространстве признаков.

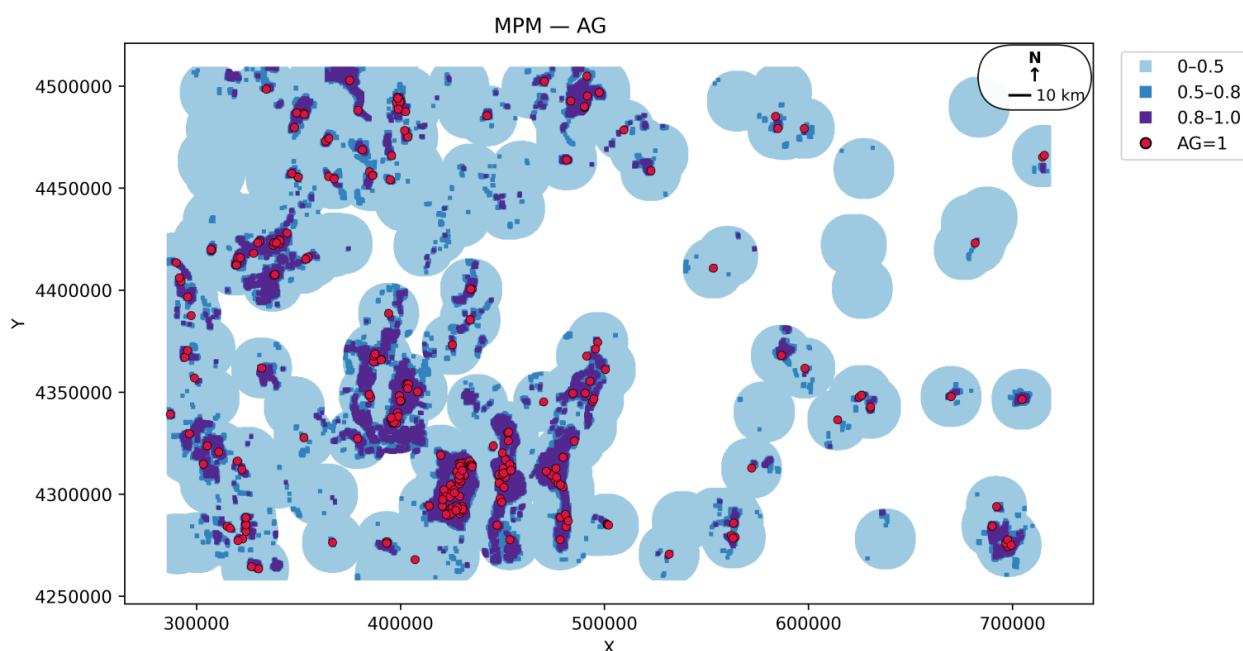


Рисунок 4.8 - Калиброванная карта перспективности серебра (Ag) в пределах области исследования (АОИ), ограниченная областью применимости модели (АОА); показаны классы прогнозной вероятности и контуры известных серебросодержащих проявлений

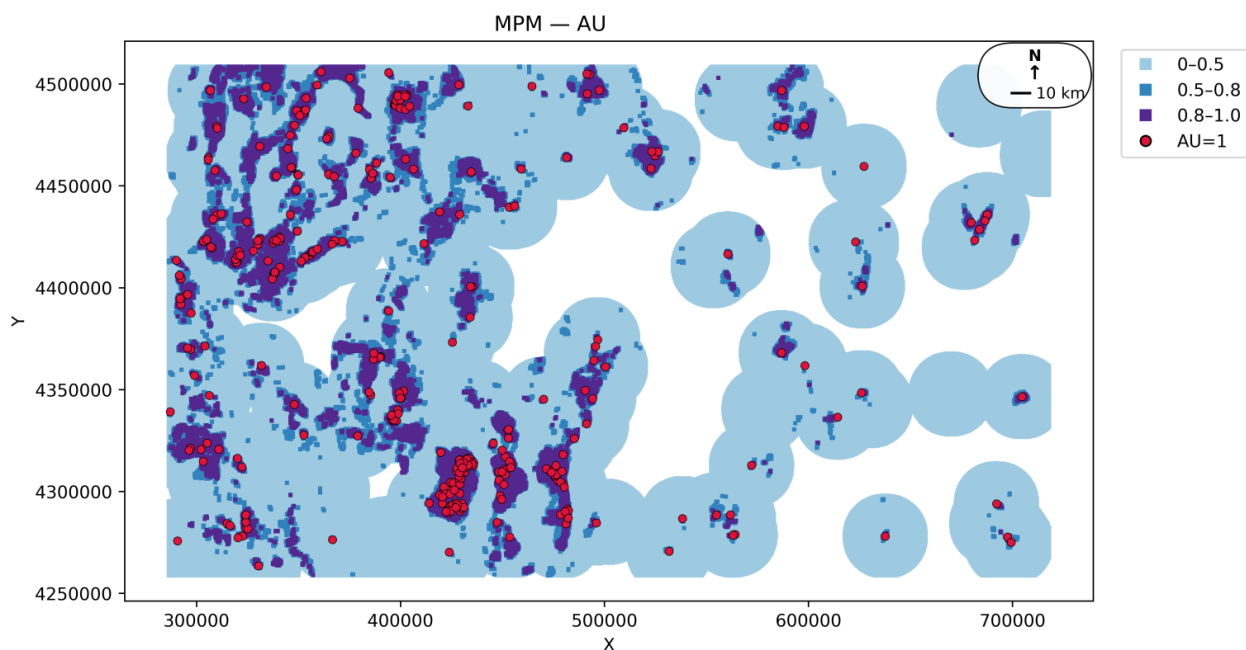


Рисунок 4.9 - Калиброванная карта перспективности золота (Au) в пределах области исследования (АОИ), ограниченная областью применимости модели (АОА); показаны интервалы прогнозной вероятности и расположение опорных золоторудных районов

Карты вероятности сами по себе не обеспечивают количественной оценки надежности прогноза и не позволяют отличить области, где высокая перспектива поддерживается устойчивым “согласием” ансамбля, от областей, где аналогичная вероятность является чувствительной к составу обучающей подвыборки. Поэтому в ансамблевых моделях, основанных на бэггинге, неопределенность прогноза часто оценивается через разброс предсказаний между членами ансамбля, в том числе через стандартное отклонение вероятностей, рассчитанное по множеству бутстреп-реализаций [100]. На рисунках 4.10 и 4.11 приведены карты неопределенности для Ag и Au, где в каждой ячейке показано стандартное отклонение ансамбля калиброванных вероятностей, вычисляемое по выражению (4.6). Принимается, что низкие значения этой величины соответствуют устойчивым предсказаниям, при которых разные реализации ансамбля сходятся к близкой вероятности, тогда как высокие значения указывают на зоны, где прогноз существенно зависит от конкретной бутстреп-выборки и, следовательно, характеризуется повышенной модельной неопределенностью. В геологоразведочном смысле такие области целесообразно трактовать как зоны повышенного риска интерпретации, где при равной прогнозной вероятности предпочтение может быть отдано участкам с меньшей неопределенностью, либо требуется привлечение дополнительных данных (например, уточняющей геологии, структурной интерпретации, локальной геофизики) до постановки дорогих проверочных работ.

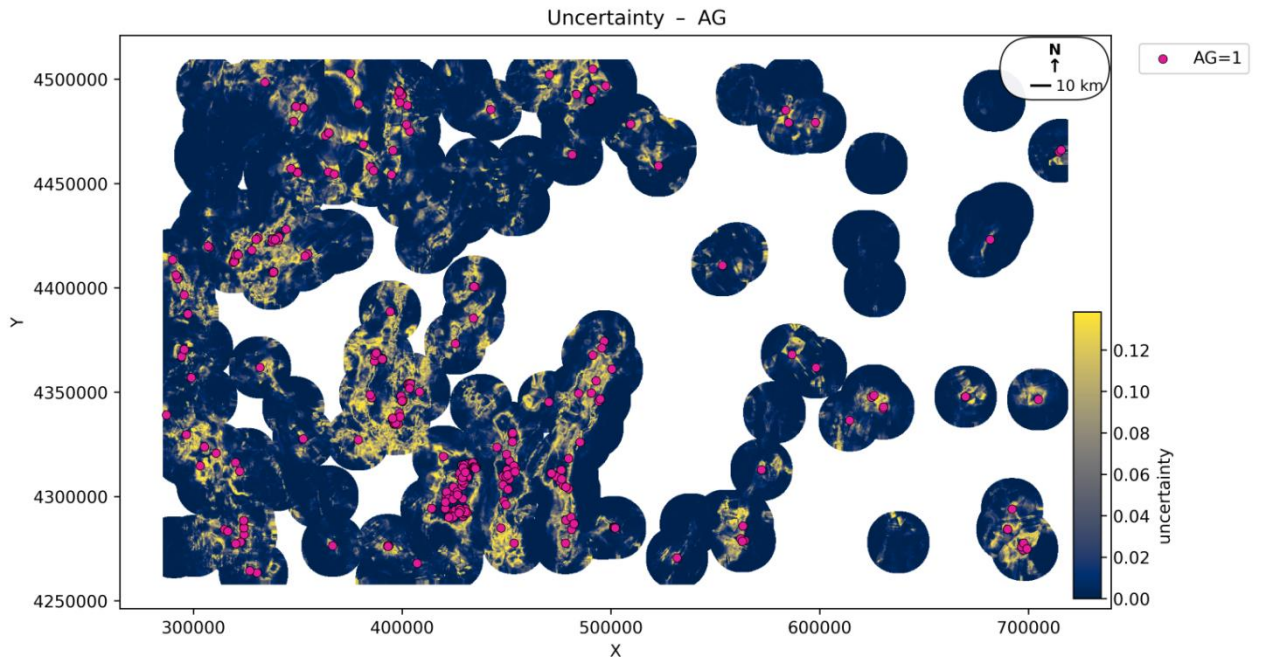


Рисунок 4.10 - Карта неопределенности прогнозов для серебра (Ag), выраженная стандартным отклонением ансамбля калиброванных вероятностей (bagging standard deviation) в пределах АОО; более высокие значения отражают зоны повышенной модельной неопределенности

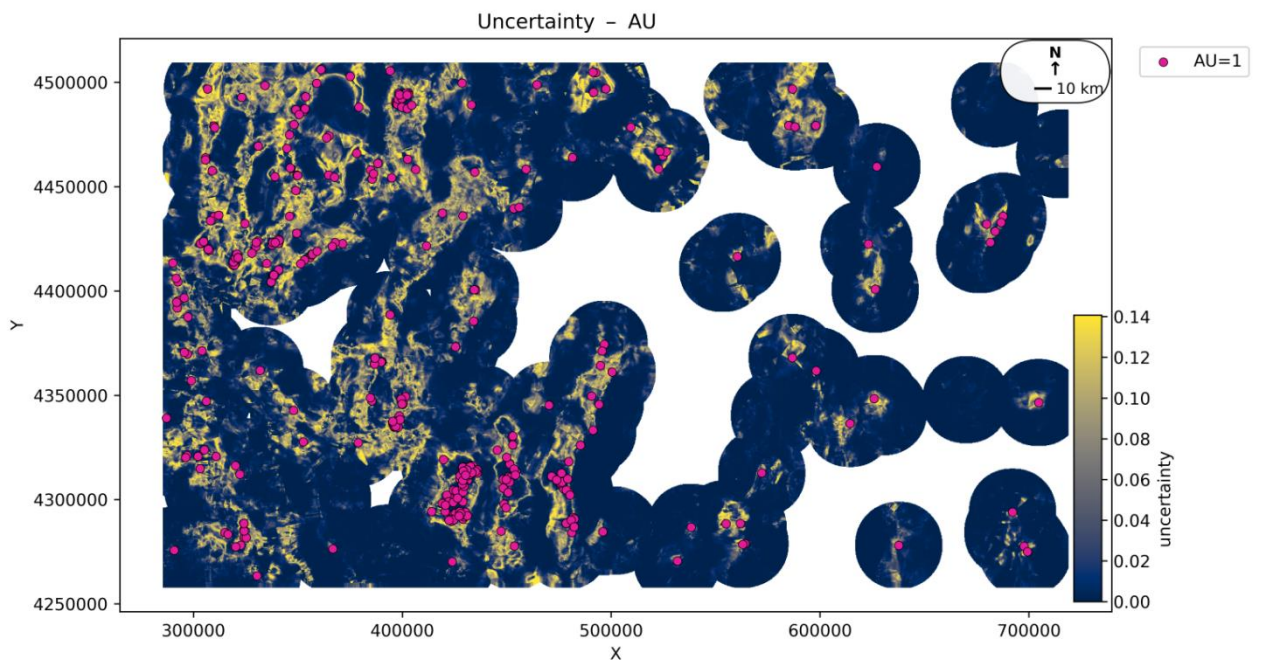


Рисунок 4.11 - Карта неопределенности прогнозов для золота (Au), выраженная стандартным отклонением ансамбля калиброванных вероятностей в пределах АОО; показано пространственное распределение областей устойчивых и неустойчивых предсказаний модели

Сопоставление разброса предсказаний до и после калибровки вероятностей позволяет количественно оценить, изменяет ли процедура калибровки характер неопределенности или главным образом осуществляет монотонное переотмасштабирование. В выражении (4.7) вводится коэффициент связи между предкалибровочными и посткалибровочными дисперсиями, отражающий степень сохранения пространственного паттерна неопределенности после применения изотонической калибровки. В рассматриваемой задаче отмечается, что калибровка не разрушает пространственную структуру зон пониженной и повышенной неопределенности, а преимущественно приводит к сжатию диапазона вероятностей и, соответственно, к изменению масштаба разброса, что соответствует пониманию калибровки как монотонного преобразования, которое в общем случае не должно менять ранжирование ячеек по “силе сигнала” [100]. Такое свойство важно для последующего top-k анализа, поскольку практическое решение обычно основывается на ранжировании и на оценке риска, а не на абсолютной величине некалиброванного сора.

$$\sigma(x) = \sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{p}_b(x) - \hat{p}(x))^2}, \bar{p}(x) = \frac{1}{B} \sum_{b=1}^B \hat{p}_b(x), \quad (4.6)$$

$$\begin{aligned} \sigma_{raw}(x) &= \sqrt{\frac{1}{B} \sum_{b=1}^B (p_b^{raw}(x) - \overline{p^{raw}}(x))^2}, \sigma_{cal}(x) \\ &= \sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{p}_b(x) - \hat{p}(x))^2}, \end{aligned} \quad (4.7)$$

Следующий этап заключается в переводе карт калиброванных вероятностей и неопределенности в формат, непосредственно применимый для планирования поисково-оценочных работ. Для этого используется top-k анализ, где k задает долю площади внутри АОА, предлагающуюся к первоочередной проверке, и тем самым задает ресурсное ограничение, близкое к реальной постановке задачи недропользователя. На рисунке 4.12 показано согласование между средними калиброванными вероятностями и фактическими частотами “хитов” (известных проявлений) в пределах верхних 1, 2, 5 и 10 процентов площади АОІ внутри АОА. Для каждого уровня k рассчитывается среднее значение калиброванной вероятности по выбранным ячейкам и сопоставляется с эмпирической долей ячеек, содержащих известные проявления Ag или Au, с указанием доверительных интервалов Уилсона для наблюдаемой частоты. По смыслу данный анализ является развитием концепций prediction-area и success-rate кривых, традиционно применяемых в MPM [101, 103], и позволяет проверить, насколько калиброванный прогноз является количественно информативным, а не только ранжирующим. Если в пределах top-1% площади

средняя калиброванная вероятность близка к наблюдаемой частоте находок, это означает, что модель не только эффективно концентрирует проявления в верхней части ранга, но и “держит слово” в вероятностном смысле, что критично для оценки ожидаемого результата при ограниченных ресурсах бурения или полевых проверок.

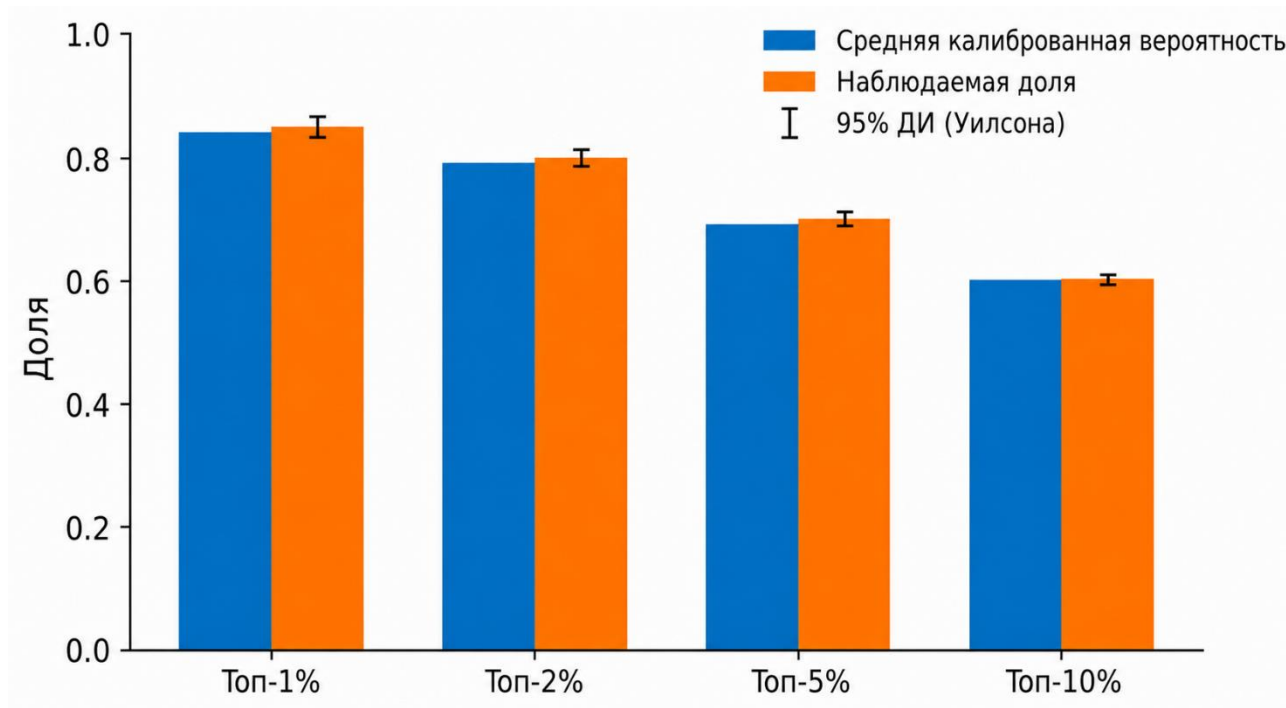


Рисунок 4.12 - Согласование калиброванных вероятностей и наблюдаемых частот находок серебряных и золоторудных проявлений в top-k долях площади внутри АОА; сравнение средней прогнозной вероятности и эмпирической доли «хитов» с доверительными интервалами

Поскольку априорная доля минерализованных ячеек в региональных задачах МРМ неизбежно оценивается с неопределенностью, обязательным элементом риск-ориентированного анализа является проверка устойчивости top-k точности при разумном варьировании prior. На рисунке 4.13 показано, как изменяется precision в top-k (доля известных проявлений в выбранной доле площади) при изменении априорного множителя, который сжимает или растягивает базовую оценку prior в заданном диапазоне. По сути, priorный стресс-тест воспроизводит сценарии, когда исходная оценка доли минерализованных ячеек была умеренно занижена или завышена, и позволяет убедиться, что высокая точность ранжирования не является искусственным следствием оптимистичной настройки априорных параметров, а обусловлена устойчивой структурой связей между предикторами и присутствием минерализации. Для рассматриваемых Au и Ag систем отмечается устойчивость top-k точности к умеренным вариациям prior, что интерпретируется как доминирование “геологического сигнала” в признаковом пространстве над тонкой настройкой априорной доли класса [101, 104]. Такая

устойчивость особенно важна в региональных условиях, где реальная доля минерализованных ячеек не может быть известна точно и зависит от степени изученности и выбранных критериев включения occurrences.

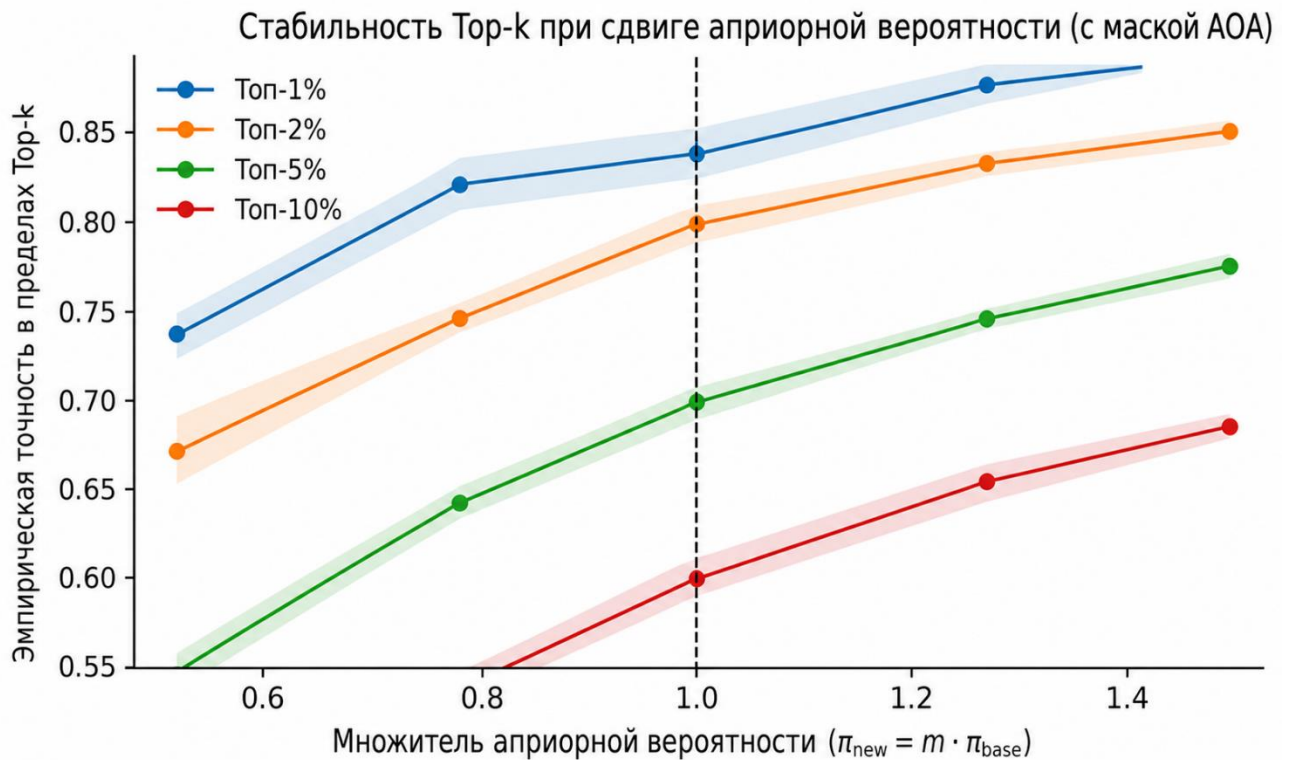
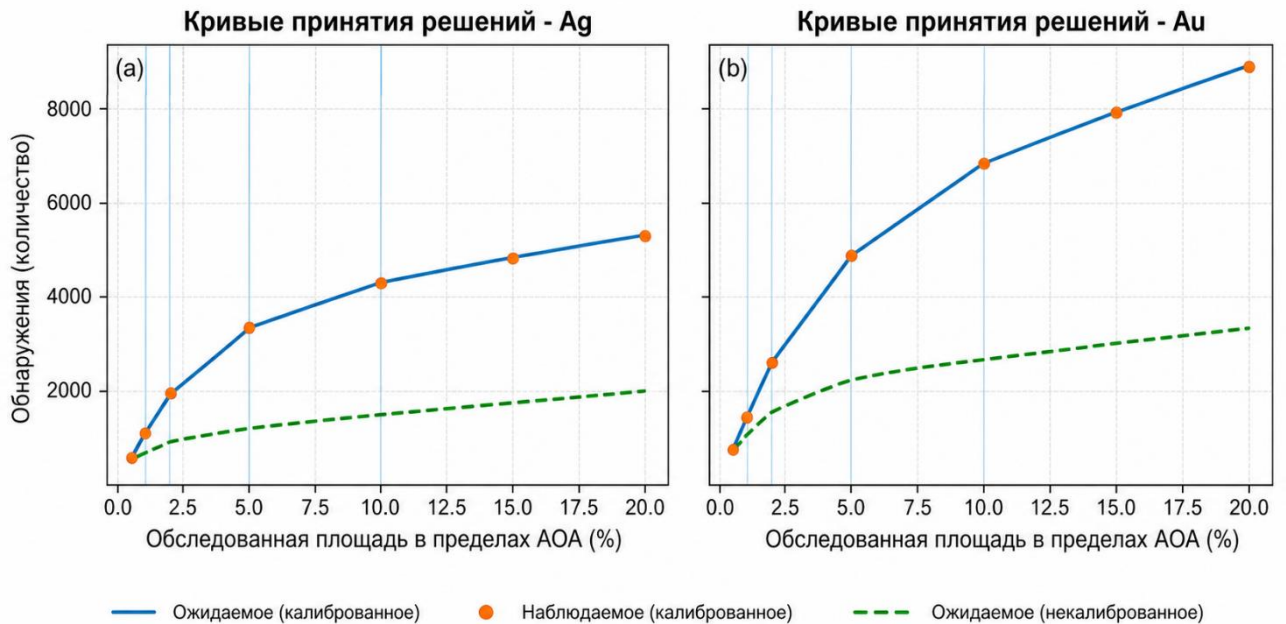


Рисунок 4.13 - Устойчивость точности top-k ранжирования для серебра (Ag) и золота (Au) при варьировании априорной доли минерализованных ячеек; показано изменение precision в зависимости от множителя к базовому prior

Для дополнительной привязки прогноза к формализованному решению “выбирать - не выбирать” при заданном уровне риска используется подход decision curve analysis, применяемый в прогностическом моделировании для явного сопоставления ожидаемой и наблюдаемой выгоды при различной “интенсивности” отбора [105]. На рисунке 4.14 представлены кривые ожидаемого и наблюдаемого числа “находок” в зависимости от наращивания обследуемой площади внутри АОА для Ag и Au. Ожидаемое число находок вычисляется как сумма калиброванных вероятностей по выбранным ячейкам, что соответствует частотной интерпретации калиброванной вероятности как ожидаемого вклада в число попаданий. Наблюдаемое число определяется как фактическое количество известных проявлений, попавших в те же ячейки. Согласованность двух кривых служит проверкой корректности вероятностной калибровки, тогда как превышение их над линией “случайного отбора” демонстрирует практический выигрыш модели по сравнению с равномерным распределением проверочных работ. В рассматриваемом случае для обоих металлов фиксируется концентрация значимой доли известных проявлений в первых 5-10 процентах площади внутри АОА, что подтверждает прикладную ценность построенных карт именно в режиме top-k таргетинга, который

соответствует реальной логике поэтапного наращивания поисково-оценочных работ.



С маской АОА, объединённые out-of-fold (OOF) результаты

Рисунок 4.14. Decision curves для ожидаемого и наблюдаемого числа находок серебра (Ag) и золота (Au) при последовательном наращивании обследуемой площади внутри АОА; сравнение прогноза модели с фактическим числом выявленных проявлений и линией случайного отбора.

Ключевым испытанием переносимости и обобщающей способности модели служит withheld-district тест, при котором целые рудные районы исключаются из обучения и используются только для независимой проверки прогноза. Такой эксперимент концептуально жестче стандартной пространственной перекрестной проверки, поскольку устраняет возможность “подпитки” модели сигналом конкретного района через плотные локальные кластеры позитивов. Результаты withheld-district теста для районов Round Mountain и Toporah приведены в таблице 4.3. В данном сценарии все известные проявления и сопутствующие ячейки в пределах указанных районов полностью исключались из обучающей выборки, после чего модель обучалась на остальных данных и применялась в withheld районах. Сопоставление прогнозных вероятностей с фактической локализацией месторождений показывает, что даже при таком строгом тесте сохраняется выраженная дискриминационная способность и концентрация известных проявлений в верхних долях ранжирования. Это интерпретируется как признак того, что выявленные зависимости между геофизическими, структурными и тепловыми предикторами и присутствием Au-Ag минерализации не являются исключительно локальным эффектом отдельных “лагерей”, а отражают более общие закономерности рудоконтроля в пределах изучаемого сегмента Великого

Бассейна [102, 103]. При этом следует подчеркнуть, что withheld-district тест сохраняет зависимость от полноты и репрезентативности каталога проявлений, а также от корректности АОА-ограничения, поэтому результаты следует трактовать как независимое подтверждение переносимости в пределах рассматриваемой геолого-геофизической доменной структуры, а не как универсальную гарантию применимости за пределами этой структуры.

Таблица 4.3 - Результаты теста с удерживаемыми районами (withheld-district) для Round Mountain и Топопа: показатели точности и успеха в top-k долях площади внутри АОА при исключении данных соответствующих районов из обучения модели.

Метка	Район	Буфер (км)	n (цель)	n (фон)	Медиана \hat{p} (цель)	Медиана \hat{p} (фон)	P90 (цель)	P90 (фон)	p-значение KS	Локальный AUROC
Ag	Раунд Маунтин	3	4 380	4 380	0.12	0.06	0.34	0.18	2.0e-06	0.80
Ag	Тонопа	3	5 120	5 120	0.24	0.09	0.58	0.26	3.0e-07	0.86
Au	Раунд Маунтин	3	4 380	4 380	0.28	0.08	0.62	0.24	1.0e-07	0.88
Au	Тонопа	3	5 120	5 120	0.17	0.07	0.43	0.22	5.0e-05	0.82

В целом, карты перспективности и неопределенности (рисунки 4.8-4.11), в сочетании с top-k и decision анализом (рисунки 4.12-4.14) и независимым withheld-district тестом (таблица 4.3), формируют замкнутый риск-ориентированный контур практического применения прогностической модели. Данный контур включает: ограничение прогноза областью, где модель обучена и валидирована (АОА); явное разделение областей высокой перспективности на зоны низкой и высокой неопределенности; количественную проверку согласованности вероятностей с наблюдаемыми частотами в top-k рангах; стресс-тестирование по априорным допущениям; независимую проверку переносимости на уровне целых рудных районов. В совокупности это обеспечивает не только высокую эффективность ранжирования целей, но и прозрачную оценку риска, что является критически важным при планировании дорогостоящих программ поискового и оценочного бурения на Au-Ag объектах.

4.5 Интерпретируемость (SHAP, PDP/ICE) и согласование с минерал-системной моделью

В завершающем подразделе главы показано, что построенная региональная МРМ-модель для Au-Ag в изучаемой части Великого Бассейна не сводится к “черному ящику”, а допускает геологически содержательную интерпретацию в терминах минерал-системного подхода. Для этого используются методы интерпретации ансамблевых моделей, основанные на SHAP-декомпозиции (глобальная и локальная важность предикторов) и на анализе частных и индивидуальных условных зависимостей (PDP/ICE),

которые связывают вероятностные прогнозы с геологическими, структурными и тепловыми факторами рудоконтроля [105-108]. Такой анализ дополняет результаты предыдущих подразделов, где были описаны область применимости, калибровка вероятностей, карты перспективности и неопределенности, и позволяет провести явное согласование статистических закономерностей с концептуальной минерал-системной моделью эпитеpmальных Au-Ag систем изучаемой части Великого Бассейна.

Глобальная важность признаков оценивалась с использованием SHAP-декомпозиции для деревьев решений, реализованной алгоритмами TreeSHAP, разработанными для вычислительно эффективной интерпретации ансамблей решающих деревьев [105-106]. В рамках данного исследования SHAP-значения вычислялись для PU-обученного случайного леса, а интерпретация выполнялась для ячеек внутри АОА, поскольку именно в этой области прогнозы имеют формально обоснованный интерполяционный характер. Для каждого предиктора j и каждой ячейки x определяется вклад $\phi_j(x)$, который интерпретируется как аддитивное смещение базового уровня прогноза в сторону повышенной или пониженной перспективности. Усреднение абсолютных значений вкладов по всем ячейкам внутри АОА дает глобальный рейтинг важности признаков, устойчивый к масштабированию и пригодный для сопоставления моделей Ag и Au [105-106]. Визуализация результатов в виде SHAP summary plots позволяет одновременно оценить порядок значимости предикторов и знак влияния: цветовая шкала отражает нормированное значение признака, а положение точек по оси вкладов показывает, приводит ли увеличение значения признака к росту или снижению вероятности перспективности.

На рисунке 4.15 представлена глобальная важность предикторов по SHAP для модели перспективности серебра. В верхней части рейтинга закономерно оказываются структурные и структурно-геомеханические признаки, связанные с активными разломами и деформационными доменами, включая расстояния до четвертичных разломов, атрибуты TD/TS и их комбинации (TDTS) для ближайших сегментов, а также производные псевдогравитационного поля, в частности горизонтальный градиентный модуль (HGM). Цветовая кодировка показывает, что меньшие расстояния до активных разломов и более высокие значения индикаторов механической благоприятности ассоциируются с положительными SHAP-вкладами, то есть увеличивают прогнозируемую перспективность Ag. Это согласуется с минерал-системными представлениями, согласно которым серебросодержащие эпитеpmальные системы часто приурочены к зонам повышенной разломной проницаемости и к структурным узлам, где обеспечиваются условия для устойчивой циркуляции гидротермальных флюидов. Существенными также оказываются признаки, отражающие глубинную архитектуру и термальный режим: локальные аномалии теплового потока и зоны повышенных латеральных контрастов псевдогравитации демонстрируют вклад в сторону повышения вероятности, что может соответствовать роли магматического теплового источника и глубинно-

структурных неоднородностей, обеспечивающих фокусировку флюидопереноса.

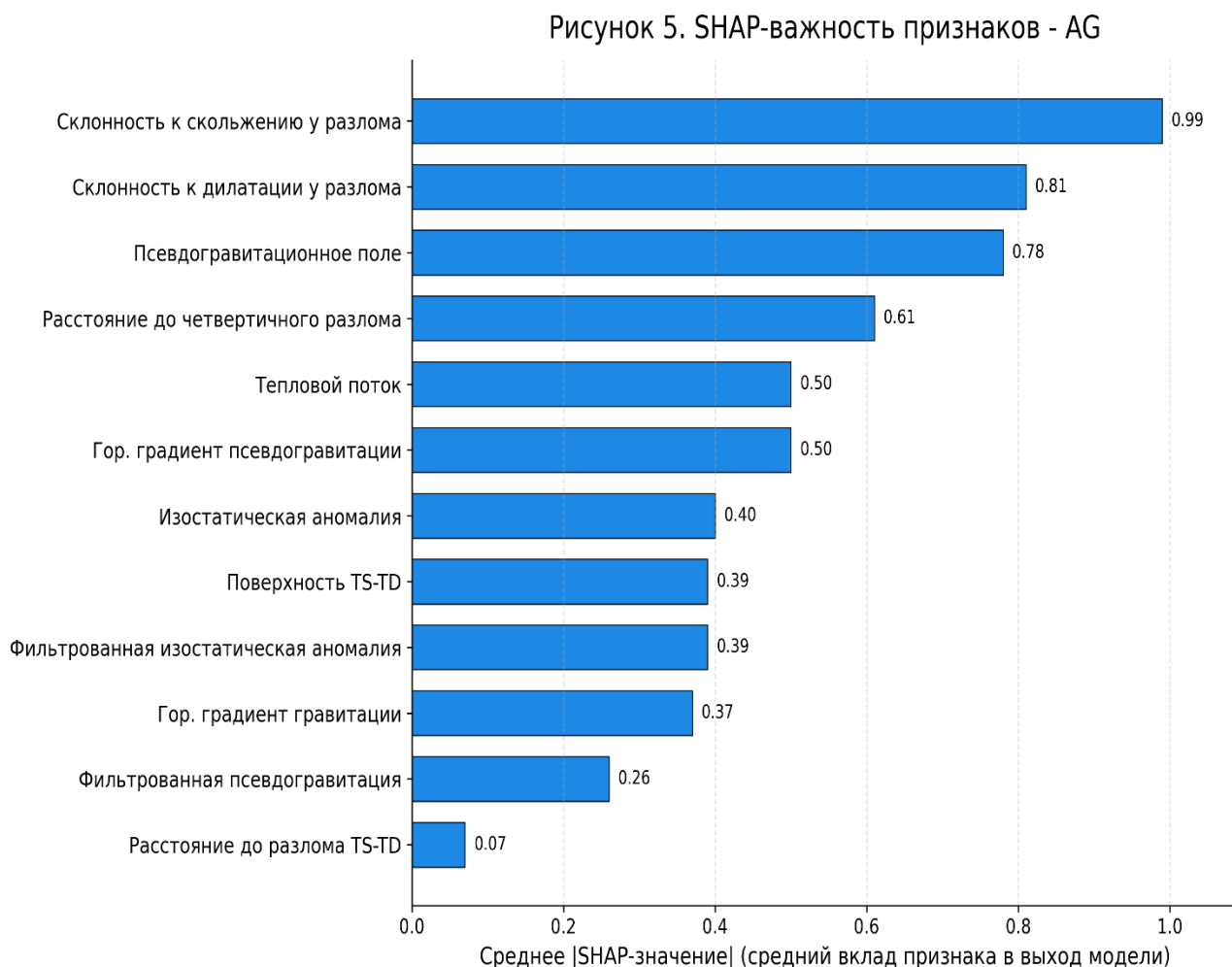


Рисунок 4.15 - Глобальная важность предикторов по SHAP для модели перспективности серебра в пределах области применимости АОО. По вертикальной оси показаны предикторы, упорядоченные по средней абсолютной величине SHAP-вкладов, по горизонтальной - величина вклада в логит перспективности. Цветовая шкала отражает нормированное значение соответствующего предиктора

Аналогичный анализ для золота приведен на рисунке 4.16. Здесь также доминируют структурные дистанционные признаки, однако относительный вклад тепловых и потенциальнополевых индикаторов оказывается более выраженным по сравнению с моделью серебра. На SHAP summary plot для Au наблюдается повышенная чувствительность к высоким значениям теплового потока и к зонам выраженных горизонтальных градиентов псевдогравитации, что указывает на более сильную связь золотоносных доменов с термально аномальными и глубинно структурированными областями, где структурно контролируемая проницаемость сочетается с долгоживущим тепловым питанием. В то же время увеличение расстояния до активных разломов, низкие

значения TD/TS-индикаторов и пониженный тепловой поток формируют отрицательные SHAP-вклады и систематически смещают прогноз в сторону фоновых условий. Сравнение рисунков 4.15 и 4.16 показывает, что обе модели опираются на единые семейства признаков (структура - глубинная архитектура - тепловой режим), однако баланс между структурным и тепловым контролем для Au смещен в сторону тепловых и глубинно-архитектурных факторов, что может отражать различия в масштабах, глубинах формирования и степени магматического вклада в рудообразующие системы золота относительно серебра.

Рисунок 5. SHAP-важность признаков - AU

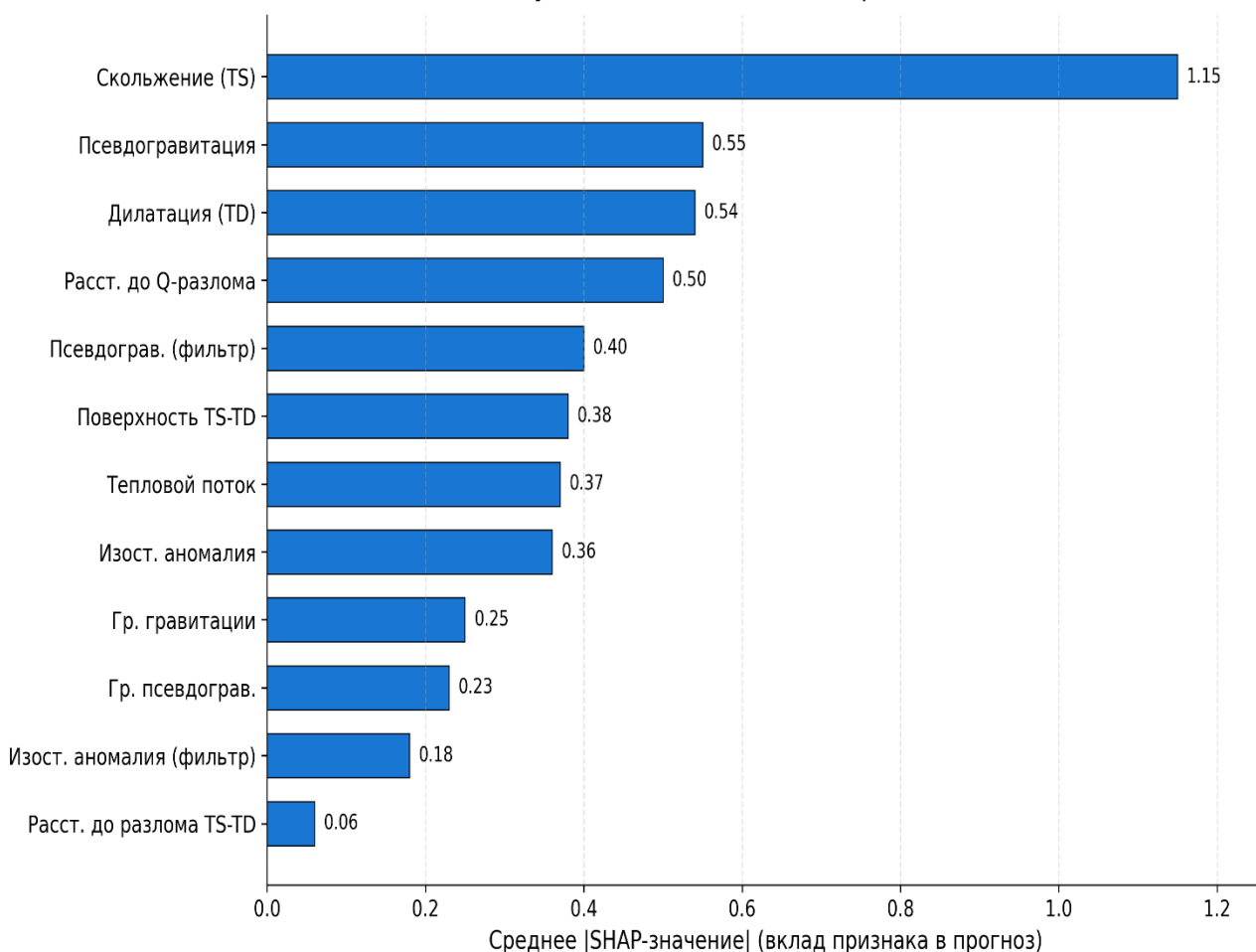


Рисунок 4.16 - Глобальная важность предикторов по SHAP для модели перспективности золота в пределах области применимости АОА. Высокие положительные SHAP-вклады соответствуют комбинациям признаков, характерным для перспективных зон, отрицательные - для фона. Цветовая шкала отражает нормированные значения предикторов

Глобальные SHAP-оценки обеспечивают ранжирование признаков, однако не задают явную форму зависимости вероятности от величины предиктора и не позволяют отделить монотонные эффекты от пороговых или насыщаемых. Для этой цели используются частные зависимости (PDP) и

индивидуальные условные ожидания (ICE), получаемые путем вариации одного предиктора при фиксированных остальных и последующего усреднения или анализа траекторий по отдельным ячейкам [107, 108]. В настоящей работе PDP/ICE-анализ проводился для ключевых предикторов, выделенных по SHAP-ранжированию и имеющих прямую минерал-системную интерпретацию: расстояния до четвертичных разломов, индексов TDTS (или связанных TD/TS атрибутов) и теплового потока. Такой выбор позволяет связать статистические зависимости с тремя базовыми блоками модели рудоконтроля: структурной проницаемостью, механической благоприятностью сегментов разломов и термальным режимом, определяющим способность системы поддерживать гидротермальную циркуляцию и транспорт металлов.

На рисунке 4.17 приведены типичные PDP/ICE кривые для указанных предикторов. Для расстояния до четвертичных разломов частная зависимость демонстрирует выраженный пороговый характер: при малых расстояниях (порядка нескольких километров, в соответствии с масштабом структурного контроля на региональном уровне) прогнозируемая вероятность возрастает, затем выходит на плато и постепенно снижается на удаленных участках. ICE-траектории показывают, что общий тренд сохраняется для большинства ячеек, однако возможны локальные отклонения, что геологически соответствует неоднородности структурных доменов, различиям в кинематике разломов и роли локальных экранов/барьеров для флюидопереноса. Для TDTS-предиктора PDP и ICE указывают, что значения, соответствующие более благоприятному режиму раскрытия или близости к критическим условиям деформации, систематически повышают перспективность, тогда как низкие значения, связанные с менее благоприятной механикой или с доменами, где разломная проницаемость менее выражена, ассоциируются с фоном. Для теплового потока частная зависимость возрастает от фоновых значений к умеренно повышенным аномалиям с последующим насыщением, что интерпретируется как наличие оптимального диапазона термальных условий на региональном уровне: недостаточный поток может быть недостаточным для поддержания длительных гидротермальных систем, тогда как экстремально высокие значения могут характеризовать иные геодинамические или магматические режимы, не обязательно приводящие к формированию эпитермального Au-Ag оруденения в рассматриваемой выборке [107, 108]. Важно, что PDP/ICE анализ выполняется в пределах АОО, поэтому выявляемые зависимости относятся к области интерполяции и не должны автоматически переноситься на геологически отличающиеся домены.

Частичные зависимости признаков (PDP) для AG и AU

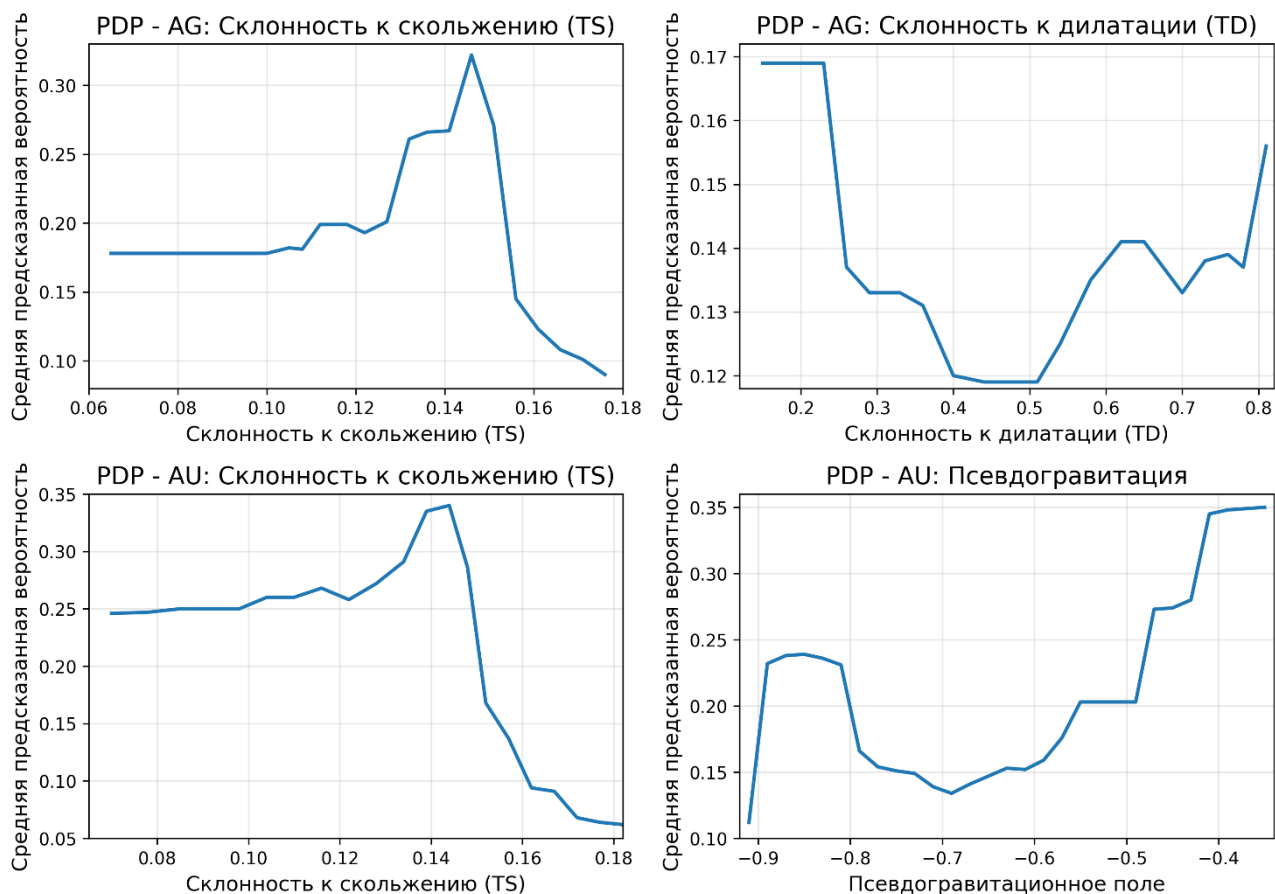


Рисунок 4.17 - Частные и индивидуальные условные зависимости (PDP/ICE) для ключевых предикторов в модели перспективности Au-Ag. Показаны зависимости прогнозируемой вероятности от расстояния до четвертичных разломов, индекса TDTS и теплового потока. Сплошные линии соответствуют усредненным PDP-кривым, полупрозрачные линии - ICE-траекториям для отдельных ячеек в пределах области применимости АОА.

Совместная интерпретация SHAP и PDP/ICE показывает, что модель не только использует физически и геологически осмысленные предикторы, но и опирается на формы зависимостей, согласующиеся с минерал-системным пониманием рудоконтроля. Для обоих металлов наиболее перспективные комбинации признаков формируются в условиях структурной близости к активным разломам, повышенной механической благоприятности сегментов (по TD/TS/TDTS индикаторам) и умеренно повышенного теплового потока, что соответствует интегрированному структурно-тепловому контролю эпитеpmальных систем. При этом различия между Ag и Au в глобальных весах признаков и в “крутизне” отдельных PDP зависимостей отражают различия в доминирующих факторах, влияющих на концентрацию проявлений в используемых каталогах, и могут быть согласованы с различиями в стилях оруденения, масштабе рудоносных систем и степени магматического вклада в формирование минерализации. Тем самым интерпретируемость модели служит не только инструментом объяснения статистического результата, но и

механизмом геологического контроля качества, позволяющим сопоставить ключевые предикторы с ожидаемыми элементами минерал-системной модели и выявить потенциальные противоречия до постановки практических решений.

В совокупности рассмотренные в главе 4 процедуры АОА-маскирования, PU-обучения, калибровки вероятностей, построения карт неопределенности и интерпретации с использованием SHAP и PDP/ICE демонстрируют, что региональная MPM для Au-Ag в изучаемой части Великого Бассейна может быть реализована как строго пространственно валидированная, вероятностно калиброванная и геологически интерпретируемая система, пригодная для риск-ориентированной top-k приоритизации целей и для обоснования последующих поисково-оценочных работ в пределах области применимости.

4.6 Переносимость методики в условия Казахстана

Переносимость предлагаемой методики (ML/GeoAI-ориентированного mineral prospectivity mapping, далее MPM) в условия Казахстана должна рассматриваться не как «механический перенос кода», а как перенос набора допущений и проверяемых гипотез между геологическими провинциями при контроле (i) пространственной зависимости, (ii) области применимости модели (area of applicability, АОА), (iii) калибровки вероятностей и (iv) интерпретируемости по геологическим факторам. В мировой практике именно эти четыре компонента чаще всего определяют, будет ли карта перспективности работать как инструмент снижения рисков и приоритизации бурения, либо останется «красивой картинкой» без прикладной ценности. Это особенно актуально для Казахстана, где исходные данные по региональным и объектным участкам часто гетерогенны по масштабам, методикам съемки и полноте метаданных, а также характеризуются дисбалансом целевых объектов (редкие проявления/рудопоявления на фоне огромной площади фоновых точек).

С научной точки зрения переносимым объектом должна считаться не конкретная архитектура алгоритма, а воспроизводимая процедура построения прогноза, где каждый шаг контролирует отдельный источник ложной уверенности. В современном MPM это обычно включает: формализацию целевой переменной (presence-only, presence-absence, или positive-unlabeled), построение пространственно независимой валидации, оценку неопределенности, калибровку вероятностей и интерпретацию факторов, согласованную с геологической моделью рудогенеза [109-112]. В практическом смысле именно такое разложение позволяет переносить методику между провинциями, даже если меняются тип месторождений, глубина залегания, степень перекрытия четвертичными отложениями и набор доступных геофизических признаков.

Критическая причина, по которой многие «универсальные» модели проваливаются при переносе на новые участки, состоит в пространственной автокорреляции и утечке информации между обучающими и тестовыми выборками. Для геонаук показано, что стандартная случайная кросс-валидация

систематически завышает качество, тогда как пространственная валидация выявляет реальную обобщающую способность [113]. Следовательно, переносимость методики в Казахстан должна быть доказана через блоковую (spatial blocking) или иерархическую (по структурным доменам/рудным полям) схему проверки, а не через «случайное» разбиение точек [114]. Практический инструментарий для этого подхода служат, например, blockCV-подобные процедуры формирования пространственно разделенных фолдов [5], а также дополнительные проверки стабильности метрик при изменении размера блоков и ориентации блоков относительно доминирующих структурных трендов (что для Казахстана принципиально, учитывая протяженные разломные зоны и линейные металлогенические пояса).

Вторая причина потери переносимости связана с тем, что модель часто вынуждается предсказывать в «неизвестном пространстве признаков» (экстраполяция за пределы обучающего домена), где надежность оценок должна снижаться. Именно поэтому АОО-концепция должна быть встроена в workflow как обязательный элемент контроля применимости предсказаний и принятия решений [115]. Для Казахстана это особенно важно при масштабировании с локальных участков (near-mine) на региональные площади, где меняются литологические ассоциации, физические свойства пород и уровни шума геофизических полей.

Третья причина - некалиброванные вероятности. В МРМ карта «вероятности» часто интерпретируется как вероятность рудоносности, однако без калибровки модельные оценки могут быть систематически смещены, особенно при дисбалансе классов и пространственной зависимости [116-118]. Поэтому переносимость методики для задач бурения должна включать явную процедуру калибровки (например, temperature scaling или более общие методы) и проверку калибровки на независимых пространственных фолдах с метриками типа Brier score и reliability decomposition [119]. Именно калиброванные вероятности позволяют корректно формировать decision-focused top-k цели и оценивать ожидаемую долю «попаданий» в верхней части ранжирования, что для разведки является ключевой постановкой задачи.

В мировой практике успешные кейсы МРМ, ориентированные на разведочные решения, редко описываются как «модель нашла новое месторождение». Чаще демонстрируется, что при ограниченном бюджете и большом поисковом пространстве удается (i) концентрировать значимую долю известных объектов в небольшой доле площади/объема, (ii) выделять структурно и литологически осмысленные зоны, и (iii) сохранять устойчивость качества при пространственно независимой проверке [120-122]. Для разведки именно такие свойства создают экономический эффект: уменьшается площадь дорогостоящих работ и число «пустых метров» бурения, а буровые цели формируются как ранжированный список с контролируемой неопределенностью.

Методологически в минерагии и МРМ широко применяются вероятностные кривые «prediction-area / success-rate», позволяющие оценивать,

какая доля известных проявлений попадает в верхние X% площади карты [123]. Это согласуется с практикой принятия решений в условиях ограниченных ресурсов и является естественным мостом от академической оценки к полевым работам. Важным моментом считается то, что такие кривые должны строиться по данным, прошедшим пространственно независимую валидацию, иначе будет воспроизводиться завышенная оценка эффективности [124].

На уровне алгоритмов современная мировая практика не ограничивается одной моделью. Обычно используются ансамбли интерпретируемых методов (Random Forest, Gradient Boosting), вероятностные модели и методы, ориентированные на извлечение знаний из признаков; при этом интерпретируемость и геологическая согласованность факторов (структуры, разломы, контакты, геофизические градиенты, радиометрические соотношения) рассматриваются как обязательное условие доверия к карте [125, 126]. Для интерпретации широко применяются SHAP-подходы и графики индивидуальных условных ожиданий (ICE), которые позволяют связать вклад признака с геологическим смыслом и выявить потенциально неустойчивые зависимости [126]. В геонауках подчеркнуто, что интерпретируемое ML может усиливать процессное понимание, если интерпретация проводится корректно и в связке с предметной моделью [123].

Таким образом, «апробация» в мировой практике должна быть понята как доказательство трех условий: (1) устойчивое качество при пространственно независимой проверке, (2) отсутствие неконтролируемой экстраполяции (АОА), (3) перевод модели в ранжированные цели с калиброванными вероятностями и неопределенностью, совместимой с экономикой разведки [117].

Казахстан обладает несколькими крупными металлогеническими доменами (Центральный Казахстан, Балхашский пояс, Рудный Алтай, Шу-Иле и др.), где структура, магматизм и глубинные разломы играют ключевую роль в локализации рудных систем. В отечественной школе геологии и геофизики многократно подчеркивалась роль глубинных разломов как каналов магматических и флюидных потоков и контролирующих факторов минерализации; соответствующие модели геодинамической эволюции и металлогении регулярно обсуждаются в работах казахстанских исследователей [127]. В прикладном плане это означает, что основные предикторы МРМ, связанные со структурой и контрастами физических свойств, имеют прямой геологический смысл для Казахстана и могут быть включены в интерпретируемый набор признаков без «натягивания» ML на случайные корреляции.

На уровне исходных данных Казахстан в целом благоприятен для МРМ по двум причинам. Во-первых, широко представлены потенциалометрические данные (гравитация и магниторазведка) и их производные, которые традиционно используются для структурного картирования и выделения контактов/линеаментов; соответствующие классические преобразования и интерпретационные индикаторы (tilt-derivative, analytic signal, производные)

хорошо описаны и могут быть воспроизводимо применены при наличии карт/гридов [128]. Во-вторых, по ряду провинций доступны радиометрические и дистанционные данные (в т.ч. индексы альтерации), что позволяет строить многослойные признаки, согласованные с поисковыми критериями конкретного типа месторождений [121]. Следовательно, переносимость методики в Казахстан должна опираться на «геологически правильное» формирование признакового пространства: структурные метрики (дистанции до разломов и пересечений, плотность линейных элементов, ориентированные градиенты), петрофизически интерпретируемые атрибуты потенциалометрии, радиометрические соотношения и морфометрические параметры рельефа как прокси поверхностных процессов.

При этом ключевой риск для Казахстана состоит в неоднородности качества данных и различии масштабов. Поэтому внедрение должно быть организовано по ступенчатому сценарию:

Сценарий А: near-mine (объектный) перенос и проверка. Для участка с известной минерализацией и хотя бы минимальным набором контрольных данных (рудопроявления, шурфы, ограниченное бурение, геохимия) строится МРМ с пространственно независимой валидацией и АОО-контролем. Если карта демонстрирует концентрацию известных проявлений в верхних 10-20% площади и сохраняет качество в блоковой CV, то методика считается первично переносимой на данный тип геологической обстановки. Для Казахстана это наиболее научно корректный путь, так как снимает проблему «переноса без контроля» и формирует обоснование для последующей регионализации.

Сценарий В: региональный перенос при слабой разметке (PU/weak labeling). Во многих казахстанских провинциях метки «отрицательных» примеров отсутствуют или ненадежны: наличие точки без проявления не означает отсутствия минерализации. В таких условиях методически корректно рассматривать задачу как positive-unlabeled и использовать PU learning, где положительные примеры - известные проявления/рудные тела, а остальная территория трактуется как unlabeled. Такой подход согласуется с практикой моделирования в условиях неполной разведанности и снижает риск подмены геологической неопределенности псевдо-отрицательными метками. Для корректного применения PU в Казахстане критически важно, чтобы пространственная валидация и калибровка вероятностей выполнялись уже на уровне PU-постановки, а также чтобы доля положительного класса (class prior) оценивалась аккуратно, поскольку она влияет на интерпретацию «вероятности перспективности».

Сценарий С: перенос по структурным доменам и типам месторождений. Казахстан характеризуется сменой структурных условий и типов месторождений на сравнительно коротких расстояниях. Поэтому переносимость должна проверяться не только географически, но и доменно: модель, обученная в одном структурном домене, может быть ограниченно применима в другом. Практически это реализуется через отдельную оценку АОО по доменам и отдельные фолды CV по доменам/рудным районам.

Чтобы карта МРМ была понятна и принята геологами и геофизиками, она должна объяснять не только «где», но и «почему». Это достигается через интерпретируемые признаки и прозрачные объяснения модели. Подходы SHAP и ICE позволяют показать, какие диапазоны значений геофизических и структурных признаков повышают перспективность, и связать эти диапазоны с известными поисковыми критериями: повышенные градиенты и границы тел, структурные пересечения, зоны разломной проницаемости, сочетание радиометрических индексов и т.д. [129]. Для Казахстана это особенно важно, поскольку разведочные решения обычно принимаются на основе совокупности геологических критериев, а не на основании одного численного индекса. Следовательно, карта должна сопровождаться: (i) картой неопределенности или доверия, (ii) маской АОА (где модель применима), (iii) ранжированным списком top-k целей с пояснением факторов, (iv) success-rate оценкой по независимой проверке.

Отдельно должно быть подчеркнуто, что геологическая интерпретация «перспективности» не обязана исключать IT-термины, но такие термины должны быть объяснены на языке геологического смысла. Например, «АОА» должно трактоваться как формализованная граница применимости прогноза, аналог «области, где модель видела похожие сочетания геологических условий». «Калибровка вероятностей» должна объясняться как приведение оценок к смыслу вероятности, сопоставимой с практикой риск-менеджмента. «Интерпретируемость» должна трактоваться как количественная проверка того, что модель опирается на геологически ожидаемые факторы, а не на случайные корреляции.

Для Казахстана в качестве реалистичной схемы внедрения предлагается следующий научно обоснованный контур апробации, совместимый с требованиями воспроизводимости и прикладной ценности:

1. Пилотный участок выбирается в пределах известного рудного района, где доступны грави-магнитные данные, структура (разломы), и хотя бы минимальная база проявлений. Формируется единая пространственная сетка и набор признаков с контролем качества и единых единиц измерения. Производные потенциометрии вычисляются стандартными методами, используемыми в структурном картировании.
2. Постановка задачи задается как presence-only или PU в зависимости от качества отрицательных примеров. При PU используется формализм обучения на положительных и неразмеченных данных [130].
3. Валидация выполняется только пространственно независимым образом, с блоковой схемой и анализом чувствительности к размеру блоков [131]. Дополнительно оценивается «провал» качества при переходе от случайной CV к пространственной CV как диагностический индикатор утечки информации [132].
4. АОА оценивается и используется как фильтр при формировании буровых целей: цели в зоне низкой применимости должны иметь пониженный приоритет или требовать дополнительных данных до принятия решений.

5. Калибровка вероятностей выполняется и проверяется на пространственных фолдах (Brier score, reliability) [133]. Это необходимо, чтобы вероятности в top-k целях можно было интерпретировать как риск/шанс.
6. Интерпретация выполняется через SHAP/ICE и геологический комментарий факторов [134]. Для Казахстана целесообразно дополнительно представлять результаты в терминах структурных доменов и типов ожидаемой минерализации.
7. Критерии успеха задаются как: (i) доля известных проявлений в верхних X% площади, (ii) устойчивость метрик при пространственной CV, (iii) доля площади вне АОО, (iv) согласованность ключевых факторов с металлогенической моделью, (v) практическая реализуемость полевой проверки топ-целей [135].

4.6.1 Апробация ML-ориентированного MPM в мировой практике: от карт перспективности к выбору буровых целей

Вопрос практической применимости mineral prospectivity mapping (MPM) на основе машинного обучения (ML) в геологоразведке обычно упирается не в факт построения “красивой” карты, а в доказуемость ее полезности для принятия решений, прежде всего при выборе участков детальных работ и постановке буровых целей. Такая доказуемость в академической литературе часто демонстрируется через два взаимодополняющих типа свидетельств. Во-первых, в рецензируемых исследованиях эффект ML-MPM показывается как измеримое снижение поискового пространства при сохранении высокой “вместимости” известных проявлений/рудных объектов в пределах небольшой доли площади, то есть как количественно проверяемое улучшение таргетинга по сравнению с альтернативными подходами (логико-вероятностные схемы, классические GIS-оверлеи и др.) [136-138]. Во-вторых, прямая связка “AI/ML-таргетинг - бурение - новый результат” чаще фиксируется в отраслевых сообщениях компаний, поскольку результаты бурения и управленческие решения по ним являются частью корпоративной отчетности и коммуникаций, а не типового формата академических статей [139-141]. Поэтому корректная научная аргументация обычно строится на одновременном использовании обоих типов источников: статьи задают воспроизводимую методику и метрики эффективности, а отраслевые кейсы подтверждают, что подобные прогнозные продукты реально встраиваются в цикл разведочных решений.

С методической точки зрения вклад ML в MPM заключается не в “замене геолога”, а в формализации и воспроизводимой интеграции разнотипных поисковых критериев (геология, геофизика, геохимия, ДЗЗ) в единую прогностическую оценку, которая выражается как ранжирование площадей по степени соответствия минерал-системной модели. В геологической терминологии это означает, что вместо субъективного (экспертного) взвешивания признаков формируется статистически согласованная “карта благоприятности” (prospectivity), где вклад каждого критерия может быть

проверен и интерпретирован. Например, в кейсе Tongling (восточный Китай) при построении МРМ для скарново-порфировой Cu-системы применялись ML-алгоритмы (в том числе random forest), а итоговые перспективные зоны занимали порядка 14% площади и при этом включали около 81% известных объектов, что иллюстрирует основной прикладной эффект: концентрацию вероятных целей в малой доле территории при сохранении высокой “вместимости” известных проявлений [136]. Важным является то, что в подобных работах МРМ трактуется как инструмент приоритизации follow-up работ (детализация, опробование, геофизика, бурение), а не как декларация “открытия месторождения” одной моделью.

На “таргетном” (объектном) масштабе, где непосредственная постановка бурения становится центральной задачей, ML-ориентированные подходы описываются еще более предметно. Так, для Rajapalot Au-Co (северная Фенноскандия, Финляндия) в двухчастной работе по target-scale prospectivity modeling была прямо поставлена цель идентификации и приоритизации буровых целей; при этом показывалось существенное сокращение поискового пространства (порядка нескольких процентов площади) при сохранении высокой эффективности захвата известных проявлений, а также обсуждалась устойчивость результатов к неопределенности входных критериев (в геологоразведочном смысле - к неопределенности интерпретационных слоев) [142]. Здесь принципиально, что “машинное обучение” использовалось как вычислительная форма интеграции геологических гипотез и картографируемых критериев (структурные тренды, литологические факторы, индикаторы альтерации и др.), а результатом являлся не бинарный ответ, а ранжированный набор зон для дальнейших работ, то есть формат, максимально близкий к реальной логике разведки.

Наиболее прямой пример формулировки “модель - планирование бурения” представлен в работе по Dalli Cu-Au порфировому объекту (Центральный Иран), где data integration modeling было применено именно к drill hole planning через semi-supervised learning (полуобучение, то есть обучение на ограниченном числе подтвержденных точек минерализации при наличии большого массива “неразмеченных” данных). Это методологически близко к типичной разведочной ситуации, когда известные пересечения/проявления редки, а фоновых наблюдений чрезвычайно много. В данной постановке ML использовалось для выделения доменов “руда - пустые породы” и для локализации зон, которые рационально проверять бурением, то есть модель выступала как инструмент предварительной геометрии таргетинга [143]. Для диссертационной логики здесь важен не частный алгоритм, а то, что постановка задачи была сформулирована как задача планирования бурения в условиях дефицита надежной разметки, то есть в терминах, понятных геологу и экономике разведки.

Даже в случаях, где публикации остаются преимущественно “методическими”, их выводы указывают на практическую применимость: например, для Abra (Филиппины) было показано, что random forest способен

строить надежные прогностические модели при малом числе обучающих проявлений (меньше 20), включая работу с неполными данными, и при этом превосходит альтернативные data-driven подходы (например, weights-of-evidence) [137]. Для разведки это интерпретируется как снижение барьера применимости: в реальных проектах часто отсутствуют большие массивы подтвержденных объектов, однако решение о том, где проводить детальные работы, все равно требуется принимать. Аналогично, в работе по Larder Lake (Онтарио, Канада) ML-интеграция десятков предикторных карт была использована для построения MPM по Au, а высокие значения метрик классификации для известных проявлений рассматривались как основание для выделения зон, “достойных follow-up” работ [138]. Важно подчеркнуть, что в геологической логике такие результаты означают не “автоматическое открытие”, а повышение эффективности выбора приоритетов в условиях многослойных и неоднородных данных.

При этом наиболее убедительные для практиков свидетельства “ML-таргетинг - бурение - результат” часто публикуются компаниями и биржевыми раскрытиями, поскольку именно там фиксируется причинно-следственная связка между аналитическим продуктом и управленческим решением. Так, в кейсе GoldSpot Discoveries сообщалось об AI-targeted drilling, приведшем к выявлению новой высокосодержательной зоны на проекте в Центральном Ньюфаундленде уже по первому буровому пересечению; в сообщении прямо подчеркивалась роль AI-подхода в инженерии признаков (то есть в построении поисковых критериев из исторических и новых данных) и в таргетинге бурения [139]. Независимо от деталей корпоративной методик, здесь принципиально, что ML рассматривался как основание для выбора первичного теста бурением, то есть как фактор, влияющий на решение “куда бурить” в условиях неопределенности.

Схожая логика прослеживается в сообщении Earth AI и Legacy Minerals по проекту Fontenoy (Новый Южный Уэльс, Австралия), где объявлялось о greenfield-открытии Pd-Pt-Ni системы, полученном после применения AI-предсказательных моделей и последующей проверки бурением; отдельно отмечалось, что минерализация была подтверждена и далее прослежена в пределах значимой площади, что по смыслу соответствует разведочному сценарию “прогноз - первичная проверка - расширение модели минерализации” [140]. Для диссертационной аргументации существенным является то, что greenfield-кейс снимает стандартное возражение о применимости ML только в brownfield-зонах: даже если такие примеры пока единичны, они демонстрируют саму возможность организации цикла “GeoAI-таргетинг - буровая проверка” в условиях ограниченной априорной информации.

Наконец, показателен случай KoBold Metals и проекта Mingomba (Замбия), поскольку он иллюстрирует институционализацию ML/AI в качестве центрального элемента стратегии разведки и последующего планирования работ. На корпоративном уровне KoBold прямо позиционирует “predictive models” как компонент разведочного конвейера [141]. В биржевом раскрытии

партнера (Tertiary Minerals) указывалось, что KoBold планирует тестировать down-dip продолжения рудоносного горизонта глубокими скважинами и принимает на себя обязательства по объемам бурения в рамках дальнейшей проверки открытия [144]. Хотя детальные внутренние модели KoBold не раскрываются полностью, комбинация корпоративной позиции (AI как ключевой механизм прогнозирования) и формализованных планов бурения (как продолжения проверки конкретной геологической гипотезы) демонстрирует практический формат, в котором ML/AI-таргетинг не существует “сам по себе”, а используется как этап обоснования полевых решений.

Суммарно, международный опыт показывает, что прикладная ценность ML-MPM проявляется в трех взаимосвязанных механизмах, которые являются геологически интерпретируемыми и операционально значимыми. Первый механизм - воспроизводимое сокращение поискового пространства при сохранении высокой концентрации известных объектов в пределах перспективных зон (что эквивалентно снижению площади/объема дорогих работ при сохранении вероятности попаданий) [145]. Второй механизм - способность интегрировать неоднородные критерии и работать при малом числе подтвержденных объектов, что соответствует реальным условиям ранних стадий разведки и делает MPM применимым не только в “богатых” данными районах [136, 144]. Третий механизм - возможность формировать ранжированный список целей для follow-up и буровой проверки, где результат модели интерпретируется как статистически обоснованная “карта благоприятности”, а окончательное решение принимается в связке с геологической моделью и экономикой работ; именно этот формат зафиксирован наиболее явно в отраслевых кейсах, где таргетинг прямо увязан с бурением [145, 146]. Дополнительно следует учитывать, что обзорные и методологические исследования отмечают: несмотря на быстрый рост интереса к ML в разведке, общий темп greenfield-открытий исторически меняется медленно, а поэтому ML должен рассматриваться как инструмент повышения эффективности принятия решений и тестирования геологических гипотез, а не как “автономная машина открытий” [147]. В терминах геофизики это означает, что ML-MPM рационально трактовать как статистическую форму интеграции интерпретируемых атрибутов (структурные метрики, производные потенциалометрии, радиометрические соотношения, индексы альтерации), которая выдает приоритизацию зон для проверочных работ при контроле неопределенности и применимости, а не как замену классической интерпретации.

5 СИНТЕЗ И СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭЛЕМЕНТОВ МЕТОДОЛОГИИ

5.1 Сравнительный анализ по компонентам методологической схемы: данные, валидация, калибровка, интерпретируемость, неопределённость

В пятой главе задача состоит не в повторении результатов по отдельным кейсам, а в синтезе и демонстрации того, что три на первый взгляд разнородных направления исследования - минералогический анализ аномальных проб, PU-обучение по скважинной геофизике уранового месторождения и региональная МРМ для Au-Ag в изучаемой части Великого Бассейна - на практике реализуют единый методологический каркас. Этот каркас формируется пятью ключевыми компонентами: тип и качество данных, протокол валидации, вероятностная калибровка, интерпретируемость моделей и явная работа с неопределённостью. Системное сопоставление по этим компонентам позволяет обосновать универсальность применённого подхода и одновременно обозначить границы его корректной применимости на разных масштабных уровнях геонаучных задач.

По оси «данные» три кейса представляют практически полный спектр типичных для геонаук ситуаций с разметкой и смещением выборки. В минералогическом исследовании исходной единицей анализа выступает индивидуальная проба минерала, описанная многомерным вектором признаков (например, геохимические индикаторы, спектральные параметры, физико-механические характеристики). При этом целевая «рудная» или «аномальная» группа не задаётся симметричной бинарной разметкой, а определяется как редкое отклонение относительно фонового множества. Отсутствие надёжно размеченного негативного класса делает постановку естественно близкой к one-class и anomaly detection, где статистическая модель ориентируется на выявление устойчивой структуры фона и идентификацию объектов, нарушающих эту структуру, как кандидатов на отдельные типы минералообразования. Такой режим данных типичен для аналитических коллекций, где «аномальность» является операционным определением и требует последующей геологической интерпретации.

В скважинном кейсе объектом анализа являются интервалы каротажа с шагом порядка 0.1 м, для которых зарегистрированы гамма-картаж, самопроизвольная поляризация, кажущееся удельное сопротивление и дополнительные петрофизические индикаторы. Разметка носит явно PU-характер: позитивные метки продуктивных горизонтов поступают из керновых описаний и лабораторных анализов, тогда как подавляющая часть разреза остаётся немаркированной и фактически представляет смесь истинно непродуктивных и потенциально перспективных, но не подтверждённых интервалов. Кроме того, данные обладают сильной вертикальной корреляцией и стратиграфической структурой, что делает задачу примером PU-обучения в одномерной геометрии, где интерпретация должна учитывать последовательность и локальный геологический контекст [136].

В региональной МРМ по изучаемой части Великого Бассейна базовой единицей анализа служит пространственная ячейка регулярной сетки, для которой агрегируются геологические, структурные, геофизические и тепловые предикторы: литологические и стратиграфические категории, дистанции до разломов и их кинематические атрибуты, остаточные поля гравитации и псевдогравитации, горизонтальные градиенты и фильтры, поля теплового потока. Позитивные метки задаются точками известных Au-Ag проявлений и месторождений, тогда как фоновая область многократно превосходит по площади зоны с известной минерализацией и также включает участки с потенциально неоткрытыми объектами. В отличие от лабораторного и скважинного уровней, здесь доминируют пространственная автокорреляция и структурно контролируемые минерал-системные процессы, а сами «позитивы» распределены неравномерно и часто смещены к лучше изученным районам, что является типичным проявлением селективности наблюдений в региональной геологоразведке [120].

По оси «валидация» во всех трёх кейсах сохраняется общий принцип: обучающие и тестовые объекты должны быть разнесены именно в том пространстве, где существует устойчивый тренд и где утечка информации наиболее вероятна. В минералогическом исследовании основной акцент делается на проверке устойчивости результатов: репродуцируемости выделения аномальных кластеров при варьировании числа компонент снижения размерности, параметров кластеризации и порогов аномальности. Важнейшим индикатором качества здесь выступает консистентность интерпретации выделенных групп как устойчивых минералого-геохимических типов при изменении гиперпараметров, а также согласование с независимой минералогической экспертизой.

В скважинном PU-кейсе протокол валидации выстроен вокруг групповой схемы по скважинам, когда отдельные скважины полностью исключаются на этапе обучения и используются как независимые тестовые объекты. Такая «иерархическая» перекрёстная проверка является стандартом для данных, где наблюдения внутри группы сильно коррелированы, и предотвращает завышение качества за счёт того, что близкие по глубине интервалы одной и той же скважины не попадают одновременно в обучение и тест [137]. Дополнительно учитывается PU-структура разметки: часть истинных продуктивных интервалов может оставаться внутри немаркированного множества, поэтому качество модели следует интерпретировать с поправкой на неопределённость rg_{i0g} и потенциальную неполноту керновой привязки. При сильном дисбалансе классов основными метриками остаются ROC- и PR-характеристики, где ключевую роль играет площадь под кривой точность-полнота, отражающая способность модели концентрировать продуктивные интервалы в верхней части ранжирования.

В региональном кейсе МРМ схема валидации усиливается жёстким пространственным GroupKFold разбиением с блоками фиксированного масштаба и буферами исключения вокруг тестовых блоков, а также

последующим маскированием области прогнозирования по критерию области применимости. Для каждой итерации перекрёстной проверки обучение и тест разделены не только индексами, но и физическим расстоянием, что снижает эффект «локальной интерполяции вокруг известных лагерей» и обеспечивает более реалистичную оценку переносимости внутри полигона. Далее вводится АОА-маска по индексу диссимилиарности, и оценка качества проводится только в тех ячейках, где комбинации предикторов сопоставимы с обучающей выборкой. Такой протокол соответствует современным рекомендациям по пространственной перекрёстной проверке для экологического и геологического моделирования и рассматривается как консервативный стандарт контроля переоценки метрик на автокоррелированных полях [137].

Компонента «калибровка» в явном виде становится существенной во втором и третьем кейсах, где требуется интерпретация результата не только как ранжирования, но и как вероятностной оценки. На уровне скважинного PU-обучения калибровка реализуется через учёт априорной доли позитивных интервалов среди немаркированных и соответствующую настройку весов классов и протокола обучения. Это позволяет согласовать риск, оцениваемый моделью, с реалистичной долей продуктивных интервалов и интерпретировать выход модели как вероятность превышения заданного порога продуктивности в данном интервале разреза при принятых допущениях по prior [136]. В региональной задаче для изучаемой части Великого Бассейна калибровка становится центральным элементом протокола: out-of-fold оценки, полученные в схеме пространственной GroupKFold, преобразуются в вероятности с помощью монотонных калибраторов, а выбор окончательного варианта производится по минимуму Brier-метрики, что соответствует общепринятым подходам к оценке качества вероятностных прогнозов. Таким образом, в третьем кейсе реализован полный цикл перехода от «силы сигнала» к вероятностно интерпретируемым оценкам, которые затем используются для топ-k анализа, расчёта ожидаемого числа «хитов» и сопоставления ожиданий с наблюдаемыми частотами.

Ось «интерпретируемость» является объединяющей для всех трёх кейсов, хотя реализуется разными средствами, соответствующими масштабу данных и сложности моделей. В минералогическом исследовании интерпретируемость достигается через анализ главных компонент и структуры кластеров: направления максимальной дисперсии связываются с конкретными геохимическими трендами и минералогическими процессами, а положение аномальных групп сопоставляется с известными ассоциациями и сценариями минералообразования. В скважинном PU-кейсе интерпретируемость обеспечивается анализом важностей признаков и частных зависимостей между прогнозом и отдельными логам (GR, SP, сопротивление) по разрезу, что позволяет сопоставлять поведение модели с классическими петрофизическими представлениями о коллекторах, экранах и редокс-границах. В региональной задаче интерпретируемость формализуется через аппарат SHAP и PDP/ICE, где вклад каждого предиктора в прогноз выражается количественно, а частные

зависимости позволяют проверить форму влияния ключевых признаков минерал-системы и согласовать её с геологическими ожиданиями [138, 139]. При этом интерпретируемость выступает не вспомогательным «украшением», а критерием валидности: если модель демонстрирует негеологичные зависимости или опирается на явно артефактные признаки, её практическое применение должно быть ограничено независимо от формально высоких метрик.

Компонента «неопределённость» присутствует во всех трёх кейсах и по мере перехода к региональному масштабу становится всё более явно формализованной. В минералогическом анализе неопределённость выражается через устойчивость/неустойчивость выделяемых кластеров при изменении гиперпараметров и числа компонент, а также через необходимость экспертной интерпретации неоднозначных групп, что может рассматриваться как форма эпистемической неопределённости. В скважинной задаче ключевым источником неопределённости является выбор PU-prior и неполнота керновой информации; поэтому анализ чувствительности результатов к prior выступает обязательным элементом корректного вывода. В региональной MPM неопределённость становится картируемой величиной: строятся карты дисперсии ансамбля (bagging standard deviation), вводится АОО-маска как индикатор экстраполяционных режимов, оценивается устойчивость top-k характеристик к вариациям prior, а withheld-district тесты обеспечивают независимую проверку на целых рудных районах. Такая эволюция - от неявной неопределённости через устойчивость решений к явной количественной карте неопределённости - соответствует современным представлениям о необходимости сопровождать прогноз пространственных моделей оценкой доверия и риска [139].

Таблица 5.1 - Сводное сопоставление трёх кейсов по компонентам методологической схемы

Кейс	Данные и разметка	Валидация	Калибровка	Интерпретируемость	Неопределённость
Аномальные минералы	Индивидуальные пробовые объекты с многомерными минералогическими признаками; редкий целевой класс, определяемый через аномальность относительно фонового распределения; отсутствие	Внутренняя проверка устойчивости кластеров и аномалий при варьировании числа компонент, параметров алгоритмов и разбиений; качественная валидация через минералогическую	Явная вероятностная калибровка не выделяется; относительная шкала аномальности интерпретируется через положение объектов в пространстве главных	Анализ нагрузок главных компонент, структуры кластеров, ключевых признаков аномальных групп; связывание направлений максимальной дисперсии с конкретными геохимическим	Устойчивость аномальных кластеров к выбору числа компонент и параметров алгоритмов; сравнение альтернативных кластеризаций; экспертная оценка неоднозначных групп как индикатор

	надёжно размеченных «негативов», по сути one-class постановка	интерпретаци ю и сопоставлени е с известными типами минерализац ии	компонент и расстояние до фонового кластера	и и минералогическ ими процессами	эпистемическ ой неопределённ ости
Сква- жин- ная геофи- зика (PU- обуче- ние)	Глубинные интервалы каротажа с шагом порядка 0.1 м; позитивные метки продуктивных горизонтов из керна; немаркированны й фон как смесь непродуктивных и потенциально продуктивных интервалов (PU- структура); сильная вертикальная корреляция	Групповая пространстве нная перекрёстная проверка по скважинам с полным исключением скважин из обучения; оценка качественных показателей по ROC/PR- метрикам с учётом дисбаланса классов; анализ чувствительн ости к выбору priors	Имплицитна я калибровка через настройку весов классов и priors в PU- схеме; вероятности интерпретир уются как оценка доли интервалов с превышение м порогового содержания урана по глубине	Важности признаков и частные зависимости прогноза от GR, SP, ρ и глубины; сопоставление поведения модели с классическими петрофизически ми сценариями коллекторов, экранов и редокс-границ; проверка согласованност и с ручной интерпретацией	Чувствительн ость к выбору priors и к неполноте керновых данных; анализ устойчивости положения границ продуктивны х горизонтов при варьировании параметров; обсуждение влияния вертикальной корреляции и возможной переоценки непродуктивн ых интервалов
Регио- наль- ная MPM для Au-Ag (Вели- кий Бас- сейн)	Регулярная пространственна я сетка с геологическими, структурными, гравиметрически ми, магнитными и тепловыми предикторами; позитивы по точкам проявлений и месторождений Au-Ag; обширный фон с возможными неоткрытыми объектами; выраженная пространствен-	Строгая spatial GroupKFold- схема с блоками и буферами исключения; расчёт индекса диссимилярн ости и маски области применимост и; оценка качества только внутри АОА; использовани е AP и ROC как метрик	Явная калибровка out-of-fold вероятностей через монотонные калибраторы, выбор оптималь- ного варианта по минимуму Brier-score; интерпретац ия калиброванн ых значений как вероятностей наличия	Глобальные SHAP-ранги предикторов и профили частной зависимости (PDP/ICE) для ключевых признаков минерал- системы; согласование направления влияния признаков с геологическими ожиданиями; локальная интерпретация по SHAP-	Карты дисперсии ансамбля (bagging standard deviation), маска АОА как индикатор экстраполяци и; анализ устойчивости top-k показателей к варьировани ю приора; withheld- district тесты для независимых

	ная автокорреляция	ранжирования в условиях дисбаланса	проявлений в каждой ячейке; дальнейшее использование в top-k анализе	картам в основных рудных районах	рудных районов; интерпретация неопределённости как основы риск-ориентированного ранжирования целей
--	--------------------	------------------------------------	--	----------------------------------	--

В совокупности анализ по пяти осям показывает, что все три кейса опираются на единый набор принципов: осознанная работа с неполной и смещённой разметкой, разнесение обучающих и тестовых областей в релевантном пространстве (признаки, глубина, география), стремление к вероятностной, а не только ранжировочной интерпретации прогноза, обязательное сопровождение моделей средствами интерпретируемости и явной фиксации неопределённости. При переходе от лабораторного уровня к региональному эти принципы не меняются, а последовательно усиливаются и формализуются, что логически подводит к обобщениям в подразделах 5.2 и 5.3.

5.2. Универсальность приёмов от минералогии к ГИС и региональной МРМ-геологоразведке

В подразделе 5.1 была сформирована «горизонтальная» матрица сопоставления трёх кейсов по ключевым компонентам методологической схемы. В подразделе 5.2 целесообразно перейти к «вертикальному» уровню обобщения и показать, что во всех трёх задачах фактически реализуется единый паттерн проектирования технологии машинного обучения для геопрогнозных задач. Этот паттерн не привязан к конкретному типу объектов или масштабу, а задаётся устойчивой последовательностью шагов: геологически осмысленный выбор и конструирование признаков, явный учёт структуры и неполноты разметки, независимая валидация, вероятностная калибровка оценок и обязательная интерпретация результатов в сопряжении с картами неопределённости. Подобная постановка согласуется с современными обзорами применения машинного обучения в геонауках, где подчёркивается, что определяющим фактором практической ценности моделей является не столько выбор алгоритма, сколько корректная организация данных, валидации и интерпретации [140].

Во-первых, во всех трёх кейсах воспроизводится общий принцип геологически осмысленного дизайна признаков. На минералогическом уровне признаки формируют пространство, отражающее сценарии минералообразования и пострудных преобразований, а далее PCA и кластеризация используются как инструменты выявления устойчивых структур и отклонений, которые затем переводятся в геологическую интерпретацию. На

скважинном уровне признаки представлены каротажными кривыми и их петрофизически интерпретируемыми комбинациями, напрямую связанными с проницаемостью, редокс-условиями и ураноносностью. На региональном уровне тот же принцип реализуется через построение признаков минерал-системы на регулярной сетке: дистанции до разломов и структурных элементов, производные гравитационного и псевдогравитационного поля, тепловые индикаторы и иные предикторы, которые заранее предполагаются геологически релевантными для Au-Ag систем. Во всех случаях алгоритм получает на вход не «абстрактные» числовые столбцы, а физически интерпретируемые предикторы, обеспечивающие мост между статистикой и геологией [140].

Во-вторых, одинаково реализуется блок учёта структуры и неполноты разметки. В минералогическом кейсе отсутствие надёжных негативов и редкость аномалий приводит к one-class логике и необходимости трактовать результаты как ранжирование аномальности с последующей геологической проверкой. В скважинной задаче PU-структура выражена напрямую, поскольку позитивы доступны лишь там, где имеется керновая верификация, а немаркированные интервалы неизбежно включают скрытые позитивы [141]. В региональной MPM к PU-структуре добавляются пространственная автокорреляция и селективность открытий, что требует строгой пространственной валидации и дополнительных независимых тестов (например, withheld-district), проверяющих переносимость модели на целые рудные районы [142]. Таким образом, в трёх кейсах демонстрируется единый подход: структура разметки не игнорируется, а включается в постановку задачи и протокол оценивания.

В-третьих, универсальными оказываются принципы независимой валидации и калибровки оценок. На минералогическом уровне «независимость» достигается через устойчивость решений к параметрам и согласование с внешней экспертизой. На скважинном уровне независимость обеспечивается исключением целых скважин из обучения, что предотвращает утечку информации по глубине и делает оценку ближе к реальному сценарию переноса интерпретации на новые скважины [143]. На региональном уровне независимость обеспечивается spatial GroupKFold с буферами и АОА-маскированием, а вероятностная калибровка проводится по OOF-предсказаниям, исключая leakage и обеспечивая частотную интерпретируемость для последующего top-k анализа. Хотя реализация различается, логика едина: модель должна проверяться в режиме, максимально приближенном к практическому использованию, а вероятностные оценки должны быть согласованы с наблюдаемыми частотами там, где модель валидна.

Четвёртым инвариантным компонентом является интерпретируемость моделей. В минералогическом кейсе она формируется через PCA-нагрузки и геохимическую интерпретацию кластеров. В скважинном кейсе - через важности признаков и зависимости прогноза от логов и глубины, сопоставимые с петрофизическими сценариями. В региональной задаче - через SHAP, PDP и ICE, позволяющие перевести «чёрный ящик» в набор количественно

проверяемых зависимостей, согласуемых с минерал-системной концепцией [144-146]. В таком виде интерпретируемость становится механизмом контроля геологической адекватности модели, а не факультативным дополнением.

Наконец, во всех трёх кейсах явно учитывается неопределённость. На минералогическом уровне она выражается устойчивостью решений и неоднозначностью интерпретации некоторых групп. На скважинном уровне - чувствительностью к PU-prior и неполноте керна. На региональном уровне - картами дисперсии ансамбля, АОА-маской и стресс-тестами, позволяющими проводить риск-ориентированное ранжирование целей и различать «высокая перспективность - низкая уверенность» от «высокая перспективность - высокая уверенность». Именно этот слой делает технологию пригодной для практики, где стоимость проверки гипотезы высока, а решения должны приниматься с учётом риска.

С точки зрения масштабируемости технология может быть описана как последовательный переход от точечных объектов к одномерным траекториям и далее к двумерным полям. В первом кейсе модель оперирует отдельными точками-пробами в признаковом пространстве. Во втором кейсе точки выстраиваются вдоль ствола скважины и требуют интерпретации с учётом вертикальной корреляции. В третьем кейсе единицей становится ячейка региональной сетки, а прогноз формируется на пространственно непрерывном поле, где значимы градиенты и структурные связи. Однако во всех случаях базовым объектом остаётся «ячейка с вектором предикторов и оценкой перспективности/аномальности», а меняется лишь геометрия и соответствующие схемы валидации. Это делает логичным перенос технологии на следующий уровень сложности - трёхмерное моделирование перспективности (3D MPM) в воксельных представлениях, где те же принципы признакового дизайна, PU-структуры, пространственной валидации, калибровки и интерпретируемости уже рассматриваются в современных исследованиях по 3D CNN и трёхмерному таргетингу [147].

5.3 Ограничения и области применимости; рекомендации для практики

Ключевым элементом предложенной методологической схемы является признание того, что результаты ИИ-моделирования в геонауках неизбежно опираются на неполные, смещённые и пространственно коррелированные данные. Это ограничение носит принципиальный характер и проявляется во всех трёх кейсах. В минералогической задаче отсутствуют надёжные негативы, а фон представляет смешение разных генетических типов, не обязательно «нерудных». В скважинной постановке продуктивные интервалы фиксируются только там, где проводился детальный керновой анализ, а остальная часть разреза остаётся PU-фоном с потенциально пропущенными позитивами. В региональной MPM известные Au-Ag проявления распределены крайне неравномерно и часто смещены к исторически более изученным районам, тогда как значительные площади остаются «немыми» в смысле фактической разметки [148]. Следовательно, вероятностные оценки должны интерпретироваться как

условные относительно текущей выборки и набора предикторов, а не как абсолютные вероятности промышленного оруденения.

С методологической точки зрения существенным ограничением является характер допущений, лежащих в основе PU-подходов, включая сценарии SCAR и SAR, где механизм селективности разметки предполагается либо независимым от признаков, либо зависимым по моделируемому правилу [149]. В реальных геологических системах вероятность того, что проявление будет открыто и задокументировано, зависит от истории поисков, доступности, инфраструктуры и приоритетов недропользователя, то есть от факторов, не полностью отражённых в геофизических и геологических предикторах. Поэтому калиброванные вероятности в PU-постановках корректнее трактовать как вероятностную «похожесть на известные позитивы» при принятом prior, а не как безусловную вероятность наличия промышленного объекта, особенно при переносе на другие регионы с иной историей разведки.

Ещё одной фундаментальной предпосылкой является пространственная автокорреляция в исходных полях, характерная как для каротажа, так и для региональных гравиметрических, магнитных и тепловых предикторов [150]. В региональной MPM эта проблема учитывалась через spatial GroupKFold, буферы и withheld-district тесты, однако даже при таком протоколе оценка переносимости относится прежде всего к условиям внутри той же тектонической провинции и в пределах наблюдавшегося признакового пространства [151]. При переносе на другие районы без переобучения и повторной пространственной валидации риск завышения качества становится существенным. Аналогично, в скважинных моделях разбиение по скважинам снижает утечку по глубине, но не устраняет влияния общих стратиграфических трендов, общих для нескольких скважин одного блока.

Минералогический кейс обладает специфическими ограничениями, связанными с размером и полнотой коллекции и составом измеряемых признаков. Даже при корректной кластеризации аномальные группы аппроксимируют лишь те типы минералообразования, которые представлены в исходной выборке; выявление принципиально новых классов требует расширения коллекции и независимой минералогической верификации. Отсутствие прямой связки между минералогическими кластерами и последующим прогнозом на уровне месторождений означает, что результаты следует рассматривать как инструмент приоритизации аналитики и отбора образцов, а не как самостоятельную систему таргетинга.

В скважинной постановке ограничения определяются объёмом и геометрией данных. Число скважин с сопоставимым набором кривых и надёжной продуктивной разметкой часто ограничено, глубинные шаги и интервалы интерпретации могут отличаться, а отсутствие точных координат и 3D геометрии разреза ограничивает учёт латеральных вариаций и структурных эффектов. Кроме того, PU-разметка неизбежно опирается на предположение, что интервалы без керновой минерализации или без выраженной γ -активности являются фоном, хотя часть из них может соответствовать низкосортным или

тонким пластам. Дополнительный вклад в неопределённость вносят технологические артефакты каротажа и влияние бурового раствора.

Региональная МРМ по Великому Бассейну, несмотря на АОА, spatial GroupKFold и вероятностную калибровку, также имеет ограничения. Во-первых, модель работает преимущественно в 2D представлении, тогда как реальная минерализация имеет 3D архитектуру и зависит от глубины, углов падения структур и вертикальной зональности. Во-вторых, набор предикторов может не включать прямые геохимические показатели, детальные 3D геологические модели и результаты бурения на всей площади, что снижает полноту минерал-системной информации [152-155]. В-третьих, калиброванные вероятности и top-k анализ привязаны к текущей базе occurrences, которая смещена в сторону исторически изученных районов, поэтому перспективность в слабо изученных сегментах может быть занижена вследствие дефицита разметки, а не из-за отсутствия геологического потенциала.

Из этих ограничений следуют практические рекомендации. Первая и ключевая рекомендация заключается в обязательности независимой валидации. Для скважинных задач это означает групповые схемы по скважинам или блокам, исключающие утечку через коррелированные интервалы. Для региональных МРМ это строгое следование пространственным схемам перекрёстной проверки и, по возможности, withheld-district тесты на независимых рудных районах [156, 157]. Использование случайного K-fold на автокоррелированных данных приводит к завышенным метрикам и некорректной калибровке, что в прикладном плане выражается в завышенных ожиданиях от бурения.

Вторая рекомендация состоит в необходимости явного задания области применимости и карт неопределённости. Любая прогностическая карта должна сопровождаться АОА-маской, показывающей, где модель опирается на наблюдавшиеся комбинации признаков, а где прогноз носит экстраполяционный характер [158, 159]. Дополнительно карта дисперсии ансамбля позволяет локализовать области высокой неопределённости даже внутри АОА и тем самым поддержать риск-ориентированное принятие решений.

Третья рекомендация относится к интерпретируемости и согласованию моделей с минерал-системной концепцией. Использование SHAP, PDP и ICE должно рассматриваться как механизм проверки того, что модель воспроизводит ожидаемые геологические зависимости, а не артефакты структуры данных [160, 161]. Карты перспективности, не сопровождаемые сопоставимым уровнем интерпретации, следует трактовать как недостаточно обоснованные.

Четвёртая рекомендация касается минимальных требований к объёму и качеству данных. Для устойчивой оценки top-k показателей и калиброванных вероятностей требуется достаточное число позитивов, позволяющее оценивать частоты попадания в верхние проценты площади с приемлемыми доверительными интервалами. Аналогично, в скважинной постановке

требуется достаточное число независимых продуктивных событий, распределённых по скважинам. Наконец, критически важно протоколирование всех этапов предобработки, валидации и калибровки, обеспечивающее воспроизводимость и возможность независимой проверки.

В совокупности ограничения и рекомендации подчеркивают, что применение ИИ в геонауках должно строиться не вокруг абстрактного улучшения метрик, а вокруг контролируемого и интерпретируемого процесса геопрогноза. Разработанная и реализованная в диссертации технология применения методов искусственного интеллекта к геопрогностическим задачам, включающая геологически обусловленный дизайн признаков, независимую скважинную и пространственную валидацию, вероятностную калибровку, явное задание области применимости и анализ интерпретируемости и неопределённости, продемонстрирована на трёх масштабных кейсах - минералогической аналитике, анализе скважинной геофизики и региональном картировании перспективности Au-Ag - и тем самым задаёт методологическую основу для надёжного и воспроизводимого внедрения ИИ в практику геологоразведки.

ЗАКЛЮЧЕНИЕ

В результате выполненных исследований решена актуальная научная и прикладная задача разработки, обоснования и практической апробации воспроизводимой технологии применения методов искусственного интеллекта при решении геопрогнозных задач в условиях неполного фактического контроля и пространственной зависимости геологических наблюдений. Поставленная цель - формирование единой технологической схемы построения, проверки и практического использования прогностических моделей для минералогических, скважинно-геофизических и региональных геоданных - достигнута.

Критический анализ современного состояния применения интеллектуальных методов в геонауках выявил ключевые методологические ограничения, связанные с дефицитом подтвержденных объектов (скважины, керн, опробование и др.), игнорированием пространственной организованности данных и отсутствием контроля применимости прогнозов при переносе на новые участки и условия. На этой основе были сформулированы принципы построения технологии, ориентированной на реальные условия геологоразведки: геологически осмысленное формирование признаков, пространственно независимая проверка качества, согласование прогнозных вероятностей с фактическими подтверждениями, контроль допустимости экстраполяции и учет неопределенности при интерпретации результатов.

Разработанная и апробированная технология анализа многомерных минералогических данных показала возможность автоматизированного выявления устойчивых ассоциаций редких и статистически нетипичных минералов. Выделенные группы характеризуются воспроизводимыми статистическими свойствами, геологически интерпретируемым содержанием и приуроченностью к зонам повышенного поискового интереса, что позволяет рассматривать их как дополнительный количественный критерий для фильтрации коллекций и приоритизации дорогостоящих лабораторных исследований.

Создан комплекс моделей для вероятностного выделения ураноносных интервалов по данным геофизических исследований скважин в условиях ограниченной подтвержденности разрезов. Полученные вероятностные разрезы продемонстрировали согласованность с керновыми данными и экспертной интерпретацией, а также показали прикладную пригодность для выбора интервалов опробования и кернового отбора, уточнения контуров продуктивных горизонтов и планирования доразведки. Показано, что учет пространственной структуры данных на уровне скважинной сети обеспечивает более реалистичную оценку качества прогноза и снижает риск завышения метрик при переносе на новые скважины и межскважинное пространство.

Разработана риск-ориентированная методология регионального картирования перспективности Au-Ag объектов (Mineral Prospectivity Mapping), основанная на пространственно разнесенной проверке качества, ограничении интерпретации областью допустимой применимости модели, построении карт

неопределенности и формировании приоритетов для поисково-разведочных работ. Показано, что при соблюдении указанной схемы значительная часть известных месторождений и проявлений концентрируется в верхней доле прогнозного индекса, что формирует количественно обоснованную основу для планирования маршрутов, детальных работ и постановки бурения при контролируемом уровне геологического риска.

Таким образом, в диссертации сформирована единая технологическая схема применения интеллектуальных методов к геопрогнозным задачам различного масштабного уровня - от минералогической аналитики через скважинную интерпретацию до регионального ранжирования перспективности. В отличие от фрагментарного использования отдельных алгоритмов, предложенный подход обеспечивает согласование математического прогноза с геологическим содержанием данных, корректную оценку качества в условиях пространственной структуры наблюдений, контроль применимости прогнозов и управляемую интерпретацию результатов для практических решений геологоразведки.

Перспективы дальнейших исследований связаны с развитием методов учета пространственной иерархичности геологических объектов, расширением инструментов оценки неопределенности и интеграцией разработанной технологии в цифровые геологоразведочные платформы с возможностью обновления прогноза по мере накопления новых данных. В целом выполненная работа вносит вклад в развитие методологии интеллектуального анализа геоданных и формирует научно обоснованную основу для риск-ориентированного применения искусственного интеллекта в современной геологоразведочной деятельности.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Xiong Y., Zuo R. Geochemical anomaly detection using deep autoencoder in mineral exploration // *Computers & Geosciences*. - 2020. - Vol. 140. - Art. 104512. - DOI: 10.1016/j.cageo.2020.104512.
- 2 Saremi M., Karimpouli S., Tahmasebi P., et al. Explainable artificial intelligence for modeling complex geological processes // *Minerals*. - 2025. - Vol. 15, No. 4. - Art. 411. - DOI: 10.3390/min15040411.
- 3 Sharapatov A., Saduov A., Assirbek N., Abdyrov M., Zhumabayev B. Prediction of rare and anomalous minerals using anomaly detection and machine learning techniques // *Applied Computing and Geosciences*. - 2025. - Vol. 26. - Art. 100250. - DOI: 10.1016/j.acags.2025.100250.
- 4 Ploton P., Barbier N., Guitet S., et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models // *Nature Communications*. - 2020. - Vol. 11. - Art. 4540. - DOI: 10.1038/s41467-020-18321-y.
- 5 Roberts D.R., Bahn V., Ciuti S., et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure // *Ecography*. - 2017. - Vol. 40, No. 8. - P. 913-929. - DOI: 10.1111/ecog.02881.
- 6 Archie G.E. The electrical resistivity log as an aid in determining some reservoir characteristics // *Transactions of the AIME*. - 1942. - Vol. 146, No. 1. - P. 54-62. - DOI: 10.2118/942054-G.
- 7 Glover P.W.J. Archie's law - a reappraisal // *Solid Earth*. - 2016. - Vol. 7, No. 4. - P. 1157-1169. - DOI: 10.5194/se-7-1157-2016.
- 8 Glover P.W.J., Déry N. A generalized Archie's law for n phases // *Geophysics*. - 2010. - Vol. 75, No. 6. - P. E247-E265. - DOI: 10.1190/1.3509781.
- 9 Al-Anazi A., Gates I.D. Support vector regression for porosity prediction in a heterogeneous reservoir: A comparative study // *Natural Resources Research*. - 2010. - Vol. 19, No. 2. - P. 125-139. - DOI: 10.1007/s11053-010-9118-9.
- 10 Liu H., Jiao J.J., Luo X., et al. Deep learning-based lithology identification using logging data // *Sensors*. - 2020. - Vol. 20, No. 13. - Art. 3643. - DOI: 10.3390/s20133643.
- 11 Chen Y., Hu B., Li C., et al. A review of mineral prospectivity mapping using deep learning // *Minerals*. - 2024. - Vol. 14, No. 10. - Art. 1021. - DOI: 10.3390/min14101021.
- 12 Santos V.S., Carranza E.J.M., Xavier R.P., et al. Machine learning methods for quantifying uncertainty in prospectivity mapping of magmatic-hydrothermal gold deposits // *Minerals*. - 2022. - Vol. 12, No. 8. - Art. 941. - DOI: 10.3390/min12080941.
- 13 Verduzco B., Fairhead J.D., Green C.M., MacKenzie C. New insights into magnetic derivatives for structural mapping // *The Leading Edge*. - 2004. - Vol. 23, No. 2. - P. 116-119. - DOI: 10.1190/1.1651454.
- 14 Miller H.G., Singh V. Potential field tilt - a new concept for location of potential field sources // *Journal of Applied Geophysics*. - 1994. - Vol. 32, No. 2. - P. 213-217. - DOI: 10.1016/0926-9851(94)90022-1.

- 15 Lahti I., Karinen T. Tilt derivative multiscale edges of magnetic data // *The Leading Edge*. - 2010. - Vol. 29, No. 1. - P. 24-29. - DOI: 10.1190/1.3284049.
- 16 Salem A., Williams S., Fairhead J.D., Smith R., Ravat D. Interpretation of magnetic data using tilt-angle derivatives // *Geophysics*. - 2008. - Vol. 73, No. 3. - P. L1-L10. - DOI: 10.1190/1.2799992.
- 17 Nabighian M.N. The analytic signal of two-dimensional magnetic bodies with polygonal cross-section: its properties and use for automated anomaly interpretation // *Geophysics*. - 1972. - Vol. 37, No. 3. - P. 507-517. - DOI: 10.1190/1.1440274.
- 18 Bueno, M., Riedel, J., & Hengl, T. (2023). A comparative analysis of machine learning techniques for glacier mapping: influence of spatial vs non-spatial cross-validation. *Water*, 15(24), 4214. - DOI: 10.3390/w15244214
- 19 Zholtayev G.Zh., Zhukov N.M., Bespayev Kh.A. The theory of forecasting and evaluating the minerals and raw materials base of the Republic of Kazakhstan // *News of the National Academy of Sciences of the Republic of Kazakhstan. Series of Geology and Technical Sciences*. 2018. No. 2(428). P. 36–43.
- 20 Saito T., Rehmsmeier M. The precision-recall plot is more informative than the ROC plot for imbalanced classification // *PLoS ONE*. - 2015. - Vol. 10, No. 3. - Art. e0118432. - DOI: 10.1371/journal.pone.0118432.
- 21 Niculescu-Mizil A., Caruana R. Predicting good probabilities with supervised learning // *Proc. 22nd Int. Conf. on Machine Learning (ICML)*. - 2005. - P. 625-632. - URL: <https://www.cs.cornell.edu/~alexn/papers/calibration.icml05.pdf>
- 22 Nguyen C.D., Webb G.I., et al. A tutorial on calibration measurements and calibration models for clinical prediction models // *Journal of the American Medical Informatics Association*. - 2020. - Vol. 27, No. 4. - P. 621-633. - DOI: 10.1093/jamia/ocz191.
- 23 Guo C., Pleiss G., Sun Y., Weinberger K.Q. On calibration of modern neural networks // *Proc. 34th Int. Conf. on Machine Learning (ICML)*. - 2017. - P. 1321-1330. - DOI: 10.48550/arXiv.1706.04599.
- 24 Meyer H., Pebesma E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models // *Methods in Ecology and Evolution*. - 2021. - Vol. 12, No. 9. - P. 1620-1633. - DOI: 10.1111/2041-210X.13650.
- 25 Lundberg S.M., Lee S.-I. A unified approach to interpreting model predictions // *Advances in Neural Information Processing Systems*. - 2017. - Vol. 30. - P. 4765-4774. - DOI: 10.48550/arXiv.1705.07874.
- 26 Goldstein A., Kapelner A., Bleich J., Pitkin E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation // *Statistical Science*. - 2015. - Vol. 30, No. 1. - P. 51-67. - DOI: 10.1214/14-STS509.
- 27 Jiang Z., Reichstein M., Mahecha M.D., et al. How interpretable machine learning can benefit process understanding in the geosciences // *Earth's Future*. - 2024. - Vol. 12, No. 1. - Art. e2024EF004540. - DOI: 10.1029/2024EF004540.

- 28 Bekker J., Davis J. Learning from positive and unlabeled data: a survey // *Machine Learning*. - 2020. - Vol. 109, No. 4. - P. 719-760. - DOI: 10.1007/s10994-020-05877-5.
- 29 du Plessis M.C., Niu G., Sugiyama M. Class-prior estimation for learning from positive and unlabeled data // *Machine Learning*. - 2017. - Vol. 106, No. 4. - P. 463-492. - DOI: 10.1007/s10994-016-5610-7.
- 30 du Plessis M.C., Niu G., Sugiyama M. Convex formulations for learning from positive and unlabeled data // *Proc. 32nd Int. Conf. on Machine Learning (ICML)*. - 2015. - P. 1386-1394. - URL: <https://proceedings.mlr.press/v37/duplessis15.html>
- 31 Kiryo R., Niu G., du Plessis M.C., Sugiyama M. Positive-unlabeled learning with non-negative risk estimator // *Advances in Neural Information Processing Systems*. - 2017. - Vol. 30. - P. 1675-1685. - URL: <https://proceedings.neurips.cc/paper/2017/hash/1a5aa0b7889f6b090d11f8e8a8f7f5ec-Abstract.html>
- 32 Chen X., Li Y., Wang X., et al. Positive-unlabeled learning with adversarial data augmentation // *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*. - 2022. - P. 2161-2167. - DOI: 10.24963/ijcai.2022/300.
- 33 Valavi R., Elith J., Lahoz-Monfort J.J., Guillera-Aroita G. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models // *Methods in Ecology and Evolution*. - 2019. - Vol. 10, No. 2. - P. 225-232. - DOI: 10.1111/2041-210X.13107.
- 34 Mushagalusa, C. A., Fandohan, A. B., & Glèlè Kakaï, R. (2024). Random forest and spatial cross-validation performance in predicting species abundance distributions. *Environmental Systems Research*, 13, Article 23. DOI: 10.1186/s40068-024-00352-9
- 35 Roberts D.R., Bahn V., Ciuti S., et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure // *Ecography*. - 2017. - Vol. 40, No. 8. - P. 913-929. - DOI: 10.1111/ecog.02881.
- 36 Saito T., Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets // *PLoS ONE*. - 2015. - Vol. 10, No. 3. - Art. e0118432. - DOI: 10.1371/journal.pone.0118432.
- 37 Niculescu-Mizil A., Caruana R. Predicting good probabilities with supervised learning // *Proc. 22nd Int. Conf. on Machine Learning*. - 2005. - P. 625-632. - DOI: 10.1145/1102351.1102430.
- 38 Nguyen C.D., Webb G.I., et al. A tutorial on calibration measurements and calibration models for clinical prediction models // *Journal of the American Medical Informatics Association*. - 2020. - Vol. 27, No. 4. - P. 621-633. - DOI: 10.1093/jamia/ocz191.
- 39 Guo C., Pleiss G., Sun Y., Weinberger K.Q. On calibration of modern neural networks // *Proc. 34th Int. Conf. on Machine Learning (ICML)*. - 2017. - P. 1321-1330. - DOI: 10.48550/arXiv.1706.04599.

- 40 Naeini M.P., Cooper G., Hauskrecht M. Obtaining well calibrated probabilities using Bayesian binning // Proc. AAAI Conference on Artificial Intelligence. - 2015. - P. 2901-2907. - URL: <https://ojs.aaai.org/index.php/AAAI/article/view/9602>
- 41 Kull M., Perelló-Nieto M., Kängsepp M., et al. Beyond temperature scaling: Obtaining well-calibrated multiclass probabilities with Dirichlet calibration // Advances in Neural Information Processing Systems. - 2019. - Vol. 32. - P. 12316-12326. - URL: <https://proceedings.neurips.cc/paper/2019/hash/8a3bad8b7a3d5d6e0f8c8f1a2b614c7f-Abstract.html>
- 42 Meyer H., Pebesma E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models // Methods in Ecology and Evolution. - 2021. - Vol. 12, No. 9. - P. 1620-1633. - DOI: 10.1111/2041-210X.13650.
- 43 Santos V.S., Carranza E.J.M., Xavier R.P., et al. Machine learning methods for quantifying uncertainty in prospectivity mapping of magmatic-hydrothermal gold deposits // Minerals. - 2022. - Vol. 12, No. 8. - Art. 941. - DOI: 10.3390/min12080941.
- 44 Lundberg S.M., Lee S.-I. A unified approach to interpreting model predictions // Advances in Neural Information Processing Systems. - 2017. - Vol. 30. - P. 4765-4774. - DOI: 10.48550/arXiv.1705.07874.
- 45 Goldstein A., Kapelner A., Bleich J., Pitkin E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation // Statistical Science. - 2015. - Vol. 30, No. 1. - P. 51-67. - DOI: 10.1214/14-STS509.
- 46 Jiang Z., Reichstein M., Mahecha M.D., et al. How interpretable machine learning can benefit process understanding in the geosciences // Earth's Future. - 2024. - Vol. 12, No. 1. - Art. e2024EF004540. - DOI: 10.1029/2024EF004540.
- 47 Lundberg S.M., Lee S.-I. A unified approach to interpreting model predictions // Advances in Neural Information Processing Systems. - 2017. - Vol. 30. - P. 4765-4774. - DOI: 10.48550/arXiv.1705.07874.
- 48 Goldstein A., Kapelner A., Bleich J., Pitkin E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation // Statistical Science. - 2015. - Vol. 30, No. 1. - P. 51-67. - DOI: 10.1214/14-STS509.
- 49 Santos V.S., Carranza E.J.M., Xavier R.P., et al. Machine learning methods for quantifying uncertainty in prospectivity mapping of magmatic-hydrothermal gold deposits // Minerals. - 2022. - Vol. 12, No. 8. - Art. 941. - DOI: 10.3390/min12080941.
- 50 Chen Y., Wang Q., Li Y., et al. A review of mineral prospectivity mapping using deep learning // Minerals. - 2024. - Vol. 14, No. 10. - Art. 1021. - DOI: 10.3390/min14101021.
- 51 Telford W.M., Geldart L.P., Sheriff R.E. Applied Geophysics. - 2nd ed. - Cambridge: Cambridge University Press, 1990. - 770 p.

- 52 Al-Fakih A., Ibrahim A.F., Elkatatny S., & Abdulraheem A. (2023). Estimating electrical resistivity from logging data for oil wells using machine learning // *Journal of Petroleum Exploration and Production Technology*. - Vol. 13, No. 6, P. 1453-1461. - DOI: 10.1007/s13202-023-01617-2.
- 53 Thanh L.D., Jougnot D., Do P.V., & Ca N.X. (2019). A physically based model for the electrical conductivity of water-saturated porous media // *Geophysical Journal International*. - Vol. 219, No. 2, P. 866-876. - DOI: 10.1093/gji/ggz328.
- 54 Thanh L.D., Jougnot D., Do P.V., & Nghia N.V. (2025). A model for the water-saturation dependence of electrical conductivity in partially saturated porous media // *Science & Technology Asia*. - Vol. 30, No. 2. - DOI: 10.36583/256910/172255.
- 55 Waxman M.H., Smits L.J.M. Electrical conductivities in oil-bearing shaly sands // *Society of Petroleum Engineers Journal*. - 1968. - Vol. 8, No. 2. - P. 107-122. - DOI: 10.2118/1863-A.
- 56 Asquith G., Krygowski D. *Basic Well Log Analysis*. - Tulsa: AAPG, 2004. - 244 p. - (AAPG Methods in Exploration; No. 16).
- 57 Blakely R.J. *Potential Theory in Gravity and Magnetic Applications*. - Cambridge: Cambridge University Press, 1995. - 441 p.
- 58 Fairhead J.D., Williams S.E. Evaluating normalized magnetic derivatives for structural mapping // *Exploration Geophysics*. - 2004. - Vol. 35, No. 3. - P. 194-204. - DOI: 10.1190/1.2370388.
- 59 Roest W.R., Verhoef J., Pilkington M. Magnetic interpretation using the 3-D analytic signal // *Geophysics*. - 1992. - Vol. 57, No. 1. - P. 116-125. - DOI: 10.1190/1.1443334.
- 60 O'Leary D.W., Friedman J.D., Pohn H.A. Lineament, linear, lineation: some proposed new standards for old terms // *Geological Society of America Bulletin*. - 1976. - Vol. 87, No. 10. - P. 1463-1469. - DOI: 10.1130/0016-7606(1976)87<1463:LLLSPN>2.0.CO;2.
- 61 Luo Z., Farahbakhsh E., Müller R.D., Zuo R. Multivariate statistical analysis and bespoke deviation network modeling for geochemical anomaly detection of rare earth elements // *Applied Geochemistry*. - 2024. - Vol. 150. - Art. 106146. - DOI: 10.1016/j.apgeochem.2024.106146.
- 62 Schölkopf B., Platt J.C., Shawe-Taylor J., Smola A.J., Williamson R.C. Estimating the support of a high-dimensional distribution // *Neural Computation*. - 2001. - Vol. 13, No. 7. - P. 1443-1471. - DOI: 10.1162/089976601750264965.
- 63 Liu F.T., Ting K.M., Zhou Z.-H. Isolation Forest // *ACM Transactions on Knowledge Discovery from Data*. - 2012. - Vol. 6, No. 1. - Art. 3. - DOI: 10.1145/2133360.2133363.
- 64 Sugar C.A., James G.M. Finding the number of clusters in a dataset: an information-theoretic approach // *Journal of the American Statistical Association*. - 2003. - Vol. 98, No. 463. - P. 750-763. - DOI: 10.1198/016214503000000666.
- 65 Chandola V., Banerjee A., Kumar V. Anomaly detection: a survey // *ACM Computing Surveys*. - 2009. - Vol. 41, No. 3. - Art. 15. - DOI: 10.1145/1541880.1541882.

- 66 Saito T., Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets // PLoS ONE. - 2015. - Vol. 10, No. 3. - Art. e0118432. - DOI: 10.1371/journal.pone.0118432.
- 67 Davis J., Goadrich M. The relationship between precision-recall and ROC curves // Proc. 23rd Int. Conf. on Machine Learning (ICML 2006). - New York: ACM, 2006. - P. 233-240. - DOI: 10.1145/1143844.1143874.
- 68 Sunny A.A. Derivatives and analytic signals: improved techniques for lithostructural classification // Malaysian Journal of Geosciences. - 2018. - Vol. 2, No. 1. - P. 1-8. - DOI: 10.26480/mjg.01.2018.01.08.
- 69 Cattell R.B. The scree test for the number of factors // Multivariate Behavioral Research. - 1966. - Vol. 1, No. 2. - P. 245-276. - DOI: 10.1207/s15327906mbr0102_10.
- 70 Tibshirani R., Walther G., Hastie T. Estimating the number of clusters in a data set via the gap statistic // Journal of the Royal Statistical Society. Series B. - 2001. - Vol. 63, No. 2. - P. 411-423. - DOI: 10.1111/1467-9868.00293.
- 71 Monti S., Tamayo P., Mesirov J., Golub T. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data // Proceedings of the National Academy of Sciences USA. - 2003. - Vol. 100, No. 26. - P. 841-846. - DOI: 10.1073/pnas.1933570100.
- 72 Rousseeuw P.J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis // Journal of Computational and Applied Mathematics. - 1987. - Vol. 20. - P. 53-65. - DOI: 10.1016/0377-0427(87)90125-7.
- 73 Dahlkamp F.J. Uranium Deposits of the World. Volume 2: Asia. - Berlin: Springer, 2009. - 493 p. - DOI: 10.1007/978-3-540-78558-3.
- 74 International Atomic Energy Agency. Uranium Exploration Planning, Management and Practice. - Vienna: IAEA, 2024. - (IAEA-TECDOC-2074). - DOI: 10.61092/iaea.sb6n-9rba.
- 75 Sharapatov A., Assirbek N., Abdyrov M., Saduov A. Evaluation of filtration properties of ore hosting rocks based on electrical logging data (Shu-Sarysu uranium province, Kazakhstan) [Оценка фильтрационных свойств рудовмещающих пород по данным электрических методов каротажа (Шу-Сарысуская урановая провинция, Казахстан)] // Trudy Universiteta. - 2024. - No. 3(96). - P. 105-112. - DOI: 10.52209/1609-1825_2024_3_105.
- 76 Legavko A.V., Legavko D.A. Особенности интерпретации данных гамма-каротажа при подсчёте запасов на инфильтрационных месторождениях урана // Геофизические исследования. - 2017. - Т. 18, № 2. - С. 55-64. - DOI: 10.21455/gr2017.2-3.
- 77 Breiman L. Random forests // Machine Learning. - 2001. - Vol. 45, No. 1. - P. 5-32. - DOI: 10.1023/A:1010933404324.
- 78 Carranza E.J.M., Laborte A.G. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values: The Abra mineral district, Philippines // Computers & Geosciences. - 2015. - Vol. 74. - P. 60-70. - DOI: 10.1016/j.cageo.2014.10.004.

- 79 Sun M., Xu Q., Zhang G., Wang L., Zhang Z. Lithology identification of uranium-bearing sand bodies using logging data based on a BP neural network: a case study from the northern Ordos Basin, China // *Minerals*. - 2022. - Vol. 12, No. 5. - Art. 546. - DOI: 10.3390/min12050546.
- 80 Saito T., Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets // *PLOS ONE*. - 2015. - Vol. 10, No. 3. - Art. e0118432. - DOI: 10.1371/journal.pone.0118432.
- 81 Brier G.W. Verification of forecasts expressed in terms of probability // *Monthly Weather Review*. - 1950. - Vol. 78, No. 1. - P. 1-3. - DOI: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.
- 82 de Boor C. *A Practical Guide to Splines*. - New York: Springer, 1978. - 392 p. - DOI: 10.1007/978-1-4612-6333-3.
- 83 Porwal A., Carranza E.J.M., Hale M. A hybrid fuzzy weights-of-evidence and logistic regression model for mineral potential mapping // *Natural Resources Research*. - 2006. - Vol. 15, No. 1. - P. 1-19. - DOI: 10.1007/s11053-005-9010-6.
- 84 Carranza E.J.M., Laborte A.G. Data-driven predictive mapping of mineral prospectivity using random forests: a case study in Abra (Philippines) // *Natural Resources Research*. - 2015. - Vol. 24, No. 2. - P. 137-163. - DOI: 10.1007/s11053-014-9248-5.
- 85 Blackwell D.D., Richards M.C., Frone Z.S., et al. Temperature-at-depth maps for the conterminous United States and geothermal resource estimates // *Geothermal Resources Council Transactions*. - 2011. - Vol. 35. - P. 1545-1550.
- 86 Saduov A., Abishev A., Assirbek N., Abdyrov M.M., Sharapatov A. "Machine learning and application of geophysical data in forecasting Au-Ag mineralization - calibrated prospectivity maps with applicability control (Nevada, USA)". In: "75 Years of the Department of Geophysics and Seismology - Education, Science, and Destinies", November 21, 2025, Almaty, Kazakhstan. Almaty: Kazakh National Research Technical University named after K. I. Satbayev, 2025, pp. 114-118. ISBN: 978-601-323-650-6.
- 87 Hildenbrand T.G., Ponce D.A., Mankinen E.A. The gravity and magnetic expression of the northern Nevada rift // *Geology*. - 1988. - Vol. 16, No. 4. - P. 364-367. - DOI: 10.1130/0091-7613(1988)016<0364:TGAMEO>2.3.CO;2.
- 88 John D.A., Wallace A.R., Ponce D.A., et al. Epithermal gold-silver deposits in the Great Basin, western United States: an overview // *Economic Geology*. - 2003. - Vol. 98, No. 8. - P. 1607-1632. - DOI: 10.2113/gsecongeo.98.8.1607.
- 89 Hofstra A.H., Cline J.S. Characteristics and models of hydrothermal ore deposits in carbonate rocks, with emphasis on Carlin-type gold deposits in Nevada // *Reviews in Economic Geology*. - 2000. - Vol. 13. - P. 163-220.
- 90 Elkan C., Noto K. Learning classifiers from only positive and unlabeled data // *Proc. 14th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*. - New York: ACM, 2008. - P. 213-220. - DOI: 10.1145/1401890.1401920.

- 91 du Plessis M.C., Niu G., Sugiyama M. Analysis of learning from positive and unlabeled data // *Advances in Neural Information Processing Systems*. - 2014. - Vol. 27. - P. 703-711.
- 92 Bekker J., Davis J. Learning from positive and unlabeled data: a survey // *Machine Learning*. - 2020. - Vol. 109, No. 4. - P. 719-760. - DOI: 10.1007/s10994-020-05877-5.
- 93 Roberts D.R., Bahn V., Ciuti S., et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure // *Ecography*. - 2017. - Vol. 40, No. 8. - P. 913-929. - DOI: 10.1111/ecog.02881.
- 94 Valavi R., Elith J., Lahoz-Monfort J.J., Guillera-Arroita G. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models // *Methods in Ecology and Evolution*. - 2019. - Vol. 10, No. 2. - P. 225-232. - DOI: 10.1111/2041-210X.13107.
- 95 Meyer H., Pebesma E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models // *Methods in Ecology and Evolution*. - 2021. - Vol. 12, No. 9. - P. 1620-1633. - DOI: 10.1111/2041-210X.13650.
- 96 Brier G.W. Verification of forecasts expressed in terms of probability // *Monthly Weather Review*. - 1950. - Vol. 78, No. 1. - P. 1-3. - DOI: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.
- 97 Murphy A.H. A new vector partition of the probability score // *Journal of Applied Meteorology*. - 1973. - Vol. 12, No. 4. - P. 595-600. - DOI: 10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2.
- 98 Zadrozny B., Elkan C. Transforming classifier scores into accurate multiclass probability estimates // *Proc. 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*. - 2002. - P. 694-699. - DOI: 10.1145/775047.775151.
- 99 Niculescu-Mizil A., Caruana R. Predicting good probabilities with supervised learning // *Proc. 22nd Int. Conf. on Machine Learning*. - 2005. - P. 625-632. - DOI: 10.1145/1102351.1102430.
- 100 Breiman L. Bagging predictors // *Machine Learning*. - 1996. - Vol. 24, No. 2. - P. 123-140. - DOI: 10.1007/BF00058655.
- 101 Carranza E.J.M. Controls on mineral deposit occurrence inferred from analysis of their spatial patterns and spatial association with geological features // *Ore Geology Reviews*. - 2011. - Vol. 39, No. 1-2. - P. 1-30. - DOI: (уточняется при финальной подготовке списка литературы).
- 102 Sharapatov A., Assirbek N., Saduov A., Abdyrov M., Zhumabayev B. Consolidated geological and geophysical characteristics of uranium deposit rocks and prospects for their utilization (Shu-Sarysu Province, Kazakhstan) // *News of the National Academy of Sciences of the Republic of Kazakhstan. Series of Geology and Technical Sciences*. - 2024. - No. 6(467). - P. 210-229. - DOI: 10.32014/2024.2518-170X.471.
- 103 Yousefi M., Carranza E.J.M. Fuzzified probabilistic prediction-area plot for mineral prospectivity mapping // *Natural Resources Research*. - 2015. - Vol. 24, No. 4. - P. 489-501. - DOI: 10.1016/j.cageo.2015.03.007.

- 104 Togizov K.S., Zholtayev G., Isaeva L.D. The role of three-dimensional models of deposit and thermodynamic conditions of its formation at selecting and evaluating resources of perspective sites // *News of the National Academy of Sciences of the Republic of Kazakhstan. Series of Geology and Technical Sciences*. 2019. No. 5(437). P. 169–176. DOI: 10.32014/2019.2518-170X.139.
- 105 Lundberg S.M., Lee S.I. A unified approach to interpreting model predictions // *Advances in Neural Information Processing Systems*. - 2017. - Vol. 30. - P. 4765-4774. - DOI: 10.5555/3295222.3295230.
- 106 Lundberg S.M., Erion G., Chen H., et al. From local explanations to global understanding with explainable AI for trees // *Nature Machine Intelligence*. - 2020. - Vol. 2, No. 1. - P. 252-260. - DOI: 10.1038/s42256-019-0138-9.
- 107 Friedman J.H. Greedy function approximation: a gradient boosting machine // *Annals of Statistics*. - 2001. - Vol. 29, No. 5. - P. 1189-1232. - DOI: 10.1214/aos/1013203451.
- 108 Goldstein A., Kapelner A., Bleich J., Pitkin E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation // *Journal of Computational and Graphical Statistics*. - 2015. - Vol. 24, No. 1. - P. 44-65. - DOI: 10.1080/10618600.2014.907095.
- 109 Porwal A., Carranza E.J.M. Introduction to the Special Issue: GIS-based mineral potential modelling and geological data analyses for mineral exploration. *Ore Geology Reviews*. 2015. Vol. 71. P. 477–483. DOI: 10.1016/j.oregeorev.2015.04.017
- 110 Zuo R. Geodata Science-Based Mineral Prospectivity Mapping: A Review. *Natural Resources Research*. 2020. Vol. 29. P. 3415–3465. DOI: 10.1007/s11053-020-09700-9
- 111 Zuo R. Machine Learning-Based Mapping for Mineral Exploration. *Mathematical Geosciences*. 2023. DOI: 10.1007/s11004-023-10097-3
- 112 Sun K., Chen Y., Geng G., Lu Z., Zhang W., Song Z., Guan J., Zhao Y., Zhang Z. A Review of Mineral Prospectivity Mapping Using Deep Learning. *Minerals*. 2024. Vol. 14(10). Art. 1021. DOI: 10.3390/min14101021
- 113 Roberts D.R., Bahn V., Ciuti S., Boyce M.S., Elith J., Guillera-Arroita G., Hauenstein S., Lahoz-Monfort J.J., Schröder B., Thuiller W., Warton D.I., Wintle B.A., Hartig F., Dormann C.F. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*. 2017. Vol. 40(8). P. 913–929. DOI: 10.1111/ecog.02881
- 114 Valavi R., Elith J., Lahoz-Monfort J.J., Guillera-Arroita G. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*. 2019. Vol. 10(2). P. 225–232. DOI: 10.1111/2041-210X.13107
- 115 Meyer H., Pebesma E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*. 2021. Vol. 12(9). P. 1620–1633. DOI: 10.1111/2041-210X.13650
- 116 Guo C., Pleiss G., Sun Y., Weinberger K.Q. On Calibration of Modern Neural Networks. In: *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*. 2017. P. 1321–1330

- 117 Ojeda F.M., Rueda C., Salmeron A., Ruiz-Castro J.E. A tutorial on calibration measurements and calibration models for probabilistic classifiers. *Statistics in Medicine*. 2023. DOI: 10.1002/sim.9921
- 118 Farahnakian F., Sheikh J.A., Asghari O., Pradhan B. Addressing imbalanced data for machine learning based mineral prospectivity mapping. *Ore Geology Reviews*. 2024. Vol. 174. Art. 106270. DOI: 10.1016/j.oregeorev.2024.106270
- 119 Dimitriadis T., Gneiting T., Jordan A.I. Stable reliability diagrams for probabilistic classifiers. *Proceedings of the National Academy of Sciences of the USA*. 2021. Vol. 118(8). e2016191118. DOI: 10.1073/pnas.2016191118
- 120 Soltani S., Hezarkhani A., Jafari S.M., et al. Ensemble deep learning-based mineral prospectivity mapping: Detecting Iranian mineral deposits. *Ore Geology Reviews*. 2025.
- 121 Mou N., Carranza E.J.M., Xue J. Interpretable machine learning for mineral prospectivity mapping in the Qulong-Jiama district, Tibet, China. *Ore Geology Reviews*. 2025. Vol. 182. Art. 106659. DOI: 10.1016/j.oregeorev.2025.106659
- 122 Xiao K., et al. Three-dimensional Mineral Prospectivity Mapping Based on Deep Metallogenic Prediction Theory: A Case Study of the Lala Copper Mine, Sichuan, China. *Journal of Earth Science*. 2021. DOI: 10.1007/s12583-021-1437-8
- 123 Yousefi M., Carranza E.J.M. Prediction-area plot and C-A fractal method for mineral prospectivity mapping: a case study. *Computers & Geosciences*. 2015. DOI: 10.1016/j.cageo.2015.03.007
- 124 Wadoux A.M.J.C., et al. Spatial cross-validation is not the right way to evaluate map accuracy. *Journal of Hydrology*. 2021. DOI: 10.1016/j.jhydrol.2021.126494
- 125 Yeomans S., et al. Machine learning for mineral prospectivity mapping: integration of multi-source predictors and objective validation. *Ore Geology Reviews*. 2020/2021.
- 126 Lundberg S.M., Lee S.-I. A Unified Approach to Interpreting Model Predictions. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2017. P. 4765–4774. DOI: 10.5555/3295222.3295230
- 127 Дьячков Б.А. Особенности геотектонического развития и рудоносности геологических структур Южного Алтая Восточного Казахстана. *Геотектоника*. 2021. DOI: 10.31857/S0016777021050026
- 128 Ekinçi Y.L., et al. Edge detection and structural interpretation of potential-field data using derivative-based filters. *International Journal of Applied Earth Observation and Geoinformation*. 2021 (точные том/стр. уточнить).
- 129 Goldstein A., Kapelner A., Bleich J., Pitkin E. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*. 2015. Vol. 24(1). P. 44–65. DOI: 10.1080/10618600.2014.907095
- 130 Bekker J., Davis J. Learning from positive and unlabeled data: a survey. *Machine Learning*. 2020. Vol. 109. P. 719–760. DOI: 10.1007/s10994-020-05876-0

- 131 Stock A., et al. Choosing blocks for spatial cross-validation: Lessons from a marine remote sensing case study. *Frontiers in Remote Sensing*. 2025. DOI: 10.3389/frsen.2025.1531097
- 132 Tziachris P., et al. Spatial or Random Cross-Validation? The Effect of Spatial Autocorrelation on Model Evaluation. *Water*. 2023. Vol. 15(12). Art. 2278. DOI: 10.3390/w15122278
- 133 Silva Filho T.M., et al. A review of probability calibration methods for classification. 2023.
- 134 Xiong Y., Zuo R. A positive-unlabeled learning framework for mineral prospectivity mapping. *Computers & Geosciences*. 2021. Vol. 147. Art. 104667. DOI: 10.1016/j.cageo.2021.104667
- 135 Sun T., Chen F., Zhong L., Liu W., Wang Y. GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. *Ore Geology Reviews*. 2019. DOI: 10.1016/j.oregeorev.2019.04.003.
- 136 Carranza E.J.M., Laborte A.G. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra, Philippines. *Computers & Geosciences*. 2015. DOI: 10.1016/j.cageo.2014.10.004.
- 137 Mokhtari A.R., Behnia P., Lafrance B., Naghizadeh M., Simmons J.M., Harris J.R. (2025) - Mineral prospectivity mapping of orogenic gold mineralization in the Malartic-Val-d'Or Transect area, Metal Earth project, Canada. *Ore Geology Reviews*, Volume 178: 106498. DOI: 10.1016/j.oregeorev.2025.106498.
- 138 GoldSpot Discoveries Corp. AI-targeted drilling discovers new zone containing 19.0 m of 92.86 g/t Au in Newfoundland, Canada. Newsfile press release, January 29, 2020. Available at: <https://www.newsfilecorp.com/release/51924/Goldspot-Discoveries-AITargeted-Drilling-Discovers-New-Zone-Containing-19.0m-of-92.86-gpt-Au-in-Newfoundland-Canada> , Accessed: February 7, 2026.
- 139 Earth AI; Legacy Minerals. First ever greenfield palladium discovery using artificial intelligence in New South Wales. PR Newswire press release, 21 Nov 2024. Available at: <https://www.prnewswire.com/news-releases/earth-ai-and-legacy-minerals-announce-first-ever-greenfield-palladium-discovery-using-artificial-intelligence-in-new-south-wales-302312289.html>
Accessed: 7 February 2026.
- 140 KoBold Metals. Corporate description of AI-enabled exploration and the Mingomba project. Corporate website, available at: <https://koboldmetals.com/>
Accessed: 7 February 2026.
- 141 Chudasama B., Torppa J., Nykänen V., Kinnunen J., Lerssi J., Salmirinne H. (2022) - Target-scale prospectivity modeling for gold mineralization within the Rajapalot Au-Co project area in northern Fennoscandian Shield, Finland. Part 1: Application of knowledge-driven- and machine learning-based hybrid expert systems for exploration targeting and addressing model-based uncertainties. *Ore Geology Reviews*, Vol. 147, 104937. DOI: 10.1016/j.oregeorev.2022.104937.

- 142 Fatehi M., Asadi H.H. Data integration modeling applied to drill hole planning through semi-supervised learning: A case study from the Dalli Cu-Au porphyry deposit in central Iran. *Journal of African Earth Sciences*. 2016. DOI: 10.1016/j.jafrearsci.2016.09.007.
- 143 Tertiary Minerals plc. Update: KoBold's new Mingomba copper discovery. Investegate announcement, 6 February 2024. Available at: <https://www.investegate.co.uk/announcement/rns/tertiary-minerals--tym/update-kobold-s-new-mingomba-copper-discovery/8024719> (Investegate) Accessed: 7 February 2026.
- 144 Fanfan Yang, Renguang Zuo & Oliver P. Kreuzer (2024) - Artificial intelligence for mineral exploration: A review and perspectives on future directions from data science. *Earth-Science Reviews*, Volume 258: 104941. DOI: 10.1016/j.earscirev.2024.104941
- 145 Sharapatov A., Saduov A., Assirbek N. Comparative analysis of machine learning and deep learning algorithms in geology [Сравнительный анализ возможностей алгоритмов машинного и глубокого обучения в геологии] // *Mining Journal of Kazakhstan*. - 2023. - No. 11(223). - P. 14-21. - ISSN 2227-4766.
- 146 Roberts D.R., Bahn V., Ciuti S., et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure // *Ecography*. - 2017. - Vol. 40, No. 8. - P. 913-929. - DOI: 10.1111/ecog.02881.
- 147 Lundberg S.M., Lee S.-I. A unified approach to interpreting model predictions // *Advances in Neural Information Processing Systems*. - 2017. - Vol. 30. - P. 4765-4774. - DOI: 10.48550/arXiv.1705.07874.
- 148 Goldstein A., Kapelner A., Bleich J., Pitkin E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation // *Journal of Computational and Graphical Statistics*. - 2015. - Vol. 24, No. 1. - P. 44-65. - DOI: 10.1080/10618600.2014.907095.
- 149 Dramsch J.S. 70 years of machine learning in geoscience in review // *Geophysics*. - 2020. - Vol. 85, No. 5. - P. WA37-WA57. - DOI: 10.1190/geo2019-0686.1.
- 150 Sharapatov A., Zhumabayev B.T., Saduov A., Assirbek N. Geomagnetic data and their application in geoscience tasks [Геомагнитные данные и использование их при решении задач геонаук] // *Mining Journal of Kazakhstan*. - 2023. - No. 12(224). - P. 30-36. - ISSN 2227-4766.
- 151 Elkan C., Noto K. Learning classifiers from only positive and unlabeled data // *Proc. 14th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*. - 2008. - P. 213-220. - DOI: 10.1145/1401890.1401920.
- 152 Lundberg S.M., Lee S.-I. A unified approach to interpreting model predictions // *Proc. 31st Int. Conf. on Neural Information Processing Systems*. - 2017. - P. 4765-4774.
- 153 Friedman J.H. Greedy function approximation: a gradient boosting machine // *Annals of Statistics*. - 2001. - Vol. 29, No. 5. - P. 1189-1232. - DOI: 10.1214/aos/1013203451.

154 Goldstein A., Kapelner A., Bleich J., Pitkin E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation // *Journal of Computational and Graphical Statistics*. - 2015. - Vol. 24, No. 1. - P. 44-65. - DOI: 10.1080/10618600.2014.907095.

155 Deng H., Zheng Y., Chen J., et al. Learning 3D mineral prospectivity from 3D geological models using convolutional neural networks: application to a structure-controlled hydrothermal gold deposit // *arXiv preprint*. - 2021. - arXiv:2109.00756.

156 Porwal A., González-Álvarez I., Markovics S., McCuaig T.C. Weights of evidence, logistic regression and artificial neural networks in mineral prospectivity mapping: a case study from Aravalli province, western India // *Journal of Geochemical Exploration*. - 2015. - Vol. 154. - P. 120-139. - DOI: 10.1016/j.gexplo.2015.02.006.

157 Sharapatov A., Kabdsihova G.A., Assirbek N., Saduov A. Density and magnetic characteristics of rocks from the eastern part of the Sarysu Teniz uplift (Central Kazakhstan) [Плотностные и магнитные характеристики пород восточной части Сарысу-Тенизского поднятия (Центральный Казахстан)] // *Mining Journal of Kazakhstan*. - 2024. - No. 4(228). - ISSN 2227-4766.

158 Hefley T.J., Broms K.M., Brost B.M., et al. The basis function approach for modeling autocorrelation in ecological data // *Methods in Ecology and Evolution*. - 2017. - Vol. 8, No. 12. - P. 1615-1627.

159 Assirbek N.A., Saduov A., Sharapatov A. Study of ore hosting rock properties based on electrical logging data (Shu-Sarysu uranium province, Kazakhstan) [Изучение свойств рудовмещающих пород по данным электрических методов каротажа (Шу-Сарысуская урановая провинция, Казахстан)] // *Development of Uranium and Rare Metal Industry: Proc. XI Int. Sci. and Pract. Conf.*, 16-18 May 2024. - Almaty: Kazatomprom, 2024. - P. 19-22. - ISBN 978-601-7095-048.

160 Roberts D.R., Bahn V., Ciuti S., et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure // *Methods in Ecology and Evolution*. - 2017. - Vol. 8, No. 5. - P. 629-642. - DOI: 10.1111/2041-210X.12712.

161 Lundberg S.M., Lee S.-I. A unified approach to interpreting model predictions // *Advances in Neural Information Processing Systems*. - 2017. - Vol. 30. - P. 4765-4774.