

ИСПОЛЬЗОВАНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ КОМПОНЕНТОВ ПОЧВ С РАЗЛИЧНОЙ ГИДРОФИЛЬНОСТЬЮ

© 2025 г. О.А. Софинская^{a,*} (<http://orcid.org/0000-0002-8785-4505>),
Ф.А. Муравьев^a, Д. Раконяц^{b, c}, Л.М. Маннапова^a

^aКазанский (Приволжский) федеральный университет, ул. Кремлевская, 4/5, Казань, 420008 Россия

^bБелорусский государственный университет, пр-т Независимости, 4, Минск, 220030 Беларусь

^cМосковский физико-технический институт,
Институтский пер., 7, Московская обл., Долгопрудный, 141701 Россия

*e-mail: ushik2001@mail.ru

Поступила в редакцию 25.05.2024 г.

После доработки 30.09.2024 г.

Принята к публикации 01.10.2024 г.

Предварительно оценены возможности пилотной модели – классификатора, обученного распознаванию следов микробной деятельности на твердых поверхностях, свидетельствующих о развитии почвоподобных тел. Для машинного обучения собрана база данных объемом 500 образцов, описанных самостоятельно и в открытых источниках в период с 1988 г. по настоящее время; среди них 59 образцов представляли почвенные горизонты, 146 материнские породы и почвоподобные образования, а также породообразующие минералы, сопутствующие компоненты почвообразования, ксенобиотики, распространенные в техногенно преобразованных ландшафтах мира. Образцы вошли в базу данных в вариациях дисперсности, покрытия биопленками и пленками иной природы, химической и физической обработки. Массив значимых для машинного обучения признаков образцов включал квантили распределения контактного угла смачивания и обобщающие категориальные показатели геометрии поверхности, минерального состава, состояния органического вещества. Целевой функцией классификации служило наличие устойчивых следов микробной деятельности на твердой поверхности. Недостающие данные реконструировали с помощью процедур Монте-Карло и случайной перевыборки. В результате численных экспериментов по оптимизации качества обучения получен сбалансированный обучающий набор данных, содержащий 1233 элемента псевдовыборок. Обучено и оценено 6 моделей классификаторов с вариацией параметров. Наиболее производительный классификатор – пятислойная нейронная сеть со случайно отключаемыми нейронами – продемонстрировал на тестовой выборке правильность предсказаний 0.74 и ROC AUC 0.80, что выше, чем у более простых и быстродействующих (правильность и ROC AUC 0.70). На основании несогласия классификаций между экспертом-человеком и обученным алгоритмом установлены общие черты сложных для машинной классификации образцов: со следами жизнедеятельности, карбонатные, дисперсные – что позволяет определить направление сбора информации для повышения производительности классификатора. Разработка алгоритма распознавания следов микробной деятельности полезна для уточнения механизмов почвообразования, биогеохимических и биогеотехнологических процессов в грунтах различного происхождения, в том числе терраформирования.

Ключевые слова: биопленки, контактный угол смачивания, алгоритмы классификации, нейронная сеть, подбор гиперпараметров

DOI: 10.31857/S0032180X25020086, **EDN:** COOYQH

ВВЕДЕНИЕ

Увеличение доли техногенных и техногенно преобразованных грунтов (Urbosols, Technosols) в используемой части суши приводит к развитию

почвоподобных тел (ППТ), отличающихся высокой неоднородностью и нестабильностью свойств по сравнению с естественными грунтами и почвами [2]. Эти образования, вскрываясь, например,

эррозионным процессом, могут перезапустить почвообразование. Также они способны серьезно и непредсказуемо влиять на устойчивость почвы или массива грунта к механическим нагрузкам и факторам выветривания [3, 4, 10, 105]. Включения ППТ, несмотря на небольшую долю в объемах, могут оказаться критическими для пластических свойств, содержания органического вещества, влиять на темпы эрозии почвы и эмиссию углекислого газа в масштабах массива. При этом, в зависимости от вида использования земель человеком, ППТ могут выступать и как желательные, и как нежелательные образования. Особенности ППТ во многом определяются направлением биогеохимической деятельности микрофлоры [1, 4, 38, 105]. Управление поведением ППТ в техногенных ландшафтах можно осуществлять с помощью биогеотехнологий [26, 28, 59, 72, 107, 118]. При этом мониторинг хода биогеотехнологических изменений должен быть обеспечен достаточно чувствительными, оперативными и гибко настраиваемыми методами, оперирующими большим количеством параметров.

В первую очередь требуется диагностика наличия ППТ, которую проводят по совокупности визуальных признаков и пространственного распределения веществ и структур [2, 6, 12, 100]. К таковым признакам относят: четкую слоистую архитектуру с закономерной последовательностью самостоятельных слоев или микрослоев, следы вертикального перетока вещества между слоями, определенный диапазон для соотношения микро- и макропористости, присутствие биомаркеров (фосфор, сера, азот) и органического вещества, в количествах и типах отличающегося от вмещающей породы.

Наиболее сложным при исследовании предполагаемого включения ППТ во вмещающей породе оказывается установить наличие в его настоящем или прошлом биологической составляющей. Любые признаки протекания биогеохимических процессов имеют геохимических двойников (в частности, это проявляется, когда скорости процессов неизвестны), поэтому каждый признак в отдельности не может служить доказательством деятельности живой составляющей. В итоге комплекса биогеохимических и геохимических преобразований, в осадочно-метаморфических породах содержится до 9% гумусового вещества и аминокислот и до 1% битумоидов и керогена [9, 13, 15, 47]. Указанные органические вещества способны сохраняться в породе от миллионов до миллиардов лет, маскируя более современные следы жизнедеятельности [13, 17].

Несмотря на частичную субъективность диагностики признаков ППТ, достоверно известен факт, что в условиях биосфера ППТ обязательно должны содержать следы микробной активности. Наиболее стойкие следы остаются после поливидовых

биопленок, которые являются основной формой существования активных микробных сообществ на любой твердой поверхности в пределах биосфера, гипотетически на протяжении всей истории жизни на Земле [19, 42]. В классическом определении биопленки – это микробное сообщество, характеризуемое клетками, необратимо прикрепленными к субстрату, поверхности раздела или друг к другу, встроенными в матрицу продуцируемых ими внеклеточных полимерных веществ и проявляющими измененный фенотип в отношении скорости роста и транскрипции генов. Поскольку поливидовые биопленки в благоприятных условиях способны практически к неограниченному по времени существованию на поверхности, внеклеточный полимерный матрикс постепенно проникает в твердый субстрат, вызывая глубокие преобразования последнего (крайний вариант такого преобразования – строматолиты). ППТ являются продуктом именно таких преобразований, что придает их поверхностям особенности по сравнению с породами и материалами, не подвергнутыми глубокому воздействию живого вещества [99]. Для вскрытия этих особенностей пригодны анализ больших данных и машинное обучение (МО).

Признаки присутствия биопленок на твердых поверхностях более определены и разработаны к настоящему времени, чем признаки, пригодные для диагностики ППТ. В частности, известно, что под действием биопленок изменяются пределы смачиваемости поверхности частиц твердых тел [18, 19, 25, 96]. В течение жизни биопленка меняет свои поверхностные свойства: на определенных стадиях развития и под действием стрессовых факторов она инициирует гидрофобные взаимодействия между твердыми частицами и образование водостойких агрегатов [19, 42]. При этом гидрофобность/гидрофильность биопленки варьирует в зависимости от стадии ее зрелости и стрессовых факторов окружающей среды: в первые 10–20 сут после прикрепления к твердой поверхности биопленка продуцирует гидрофобные вещества, а затем специфические мембранные структуры, приводящие к необратимой адгезии клеток к поверхности [25]. С момента необратимой адгезии поверхность биопленки может становиться и гидрофильной, образуя набухающий внеклеточный полисахаридный матрикс [121]. В итоге на популяциях бактериальных клеток даже одного и того же штамма отмечено одновременное присутствие гидрофильных и гидрофобных пятен [41]. Эти пятна неодинаковой смачиваемости являются характерным признаком деятельности биопленок и, вероятно, способны сохраняться на контактирующем субстрате в течение долгого времени.

Наиболее широко для измерения смачиваемости горных пород и других материалов применяется контактный угол смачивания (КУ). Физически

КУ представляет собой угол, измеряемый в точке периметра контакта трех фаз (твердая поверхность—жидкость—газ), который образует касательная к поверхности раздела жидкость—газ с твердой поверхностью [14]. Материалы с КУ в пределах 0°—90° считают гидрофильтными, а в пределах 90°—180° — гидрофобными. По существующим данным можно заключить, что направление изменений смачиваемости поверхностей под биопленками зависит от стадии жизни биопленки и от свойств самой поверхности [5, 25, 42, 53, 86, 98]. Устойчивая биопленка вызывает гидрофобизацию лишь отдельных участков поверхности, не приводя к их супергидрофобности, так как в противном случае теряется доступ к питательным растворам. Однако идентификация устойчивых следов микробной деятельности на основании одной лишь гидрофобизации осложнена, поскольку сходный процесс идет и после пожаров, и после внесения в почву веществ биологического происхождения: навоза, казеина, хитозана, крахмала, лигнина, агаровой или геллановой камеди и др., а также ряда ксенобиотиков [30, 40, 71, 79]. В таких сложных случаях, когда отдельно взятый индикатор неоднозначен по отношению к явлению, используется подход “отпечатка пальца” (“fingerprinting” в англоязычной литературе). В частности, анализ геохимических “отпечатков пальцев” основан на идее о том, что химический состав минерала или горной породы отражает геологические процессы, связанные с их образованием [57, 61]. “Отпечаток пальца” — это некоторое уникальное сочетание признаков, однозначно указывающее на конкретный класс объектов. Подобным “отпечатком пальца” в случае следов микробной деятельности может оказаться вид распределения КУ [14, 97]. Следовательно, для идентификации следов микробной деятельности необходимо сравнить исследуемую поверхность с аналогичной без такого отпечатка.

Поставленная задача идентификации следов микробного преобразования твердой поверхности по смачиваемости нетривиальна, из-за неоднозначности как изменений смачиваемости под биопленками, так и сохранности таких следов на разных поверхностях, однако видится решаемой с помощью моделей МО. Цель при этом состоит в том, чтобы научить модель различать горную породу с индивидуальным набором признаков, слабо затронутую микробной деятельностью, и ту же породу, но после длительного взаимодействия с биопленками. При этом следует исходить из предположения о том, что некоторое взаимодействие поверхности с биохимическими агентами неизбежно при отборе образцов почв, горных пород и геоматериалов, поэтому оно должно быть признано фоновым шумом. Исходя из постановки задач исследования, шумом предлагается считать биопленку младше 48 ч, не прошедший этап

необратимой адгезии на поверхности, неустойчивую к механическим воздействиям или смыву водой, или ее следы [25].

Модель МО требует оснащения набором индивидуальных признаков поверхностей, которые некоторым образом связаны со смачиваемостью. В минимальном варианте это могут быть дисперсность породы, наличие глинистых и других породообразующих минералов, цемента, частиц сажи, компонентов нефти, синтетических органических соединений, признаков гидроморфизма [63, 103]. Для формирования обучающей базы данных необходимо экспертоное описание указанных признаков, измерение КУ и диагностика следов деятельности биопленок. Последняя традиционно проводится методами метагеномики, метапротеомики, анализа элементного состава поверхности, микроморфологии и др. [5, 19, 25, 26, 33, 46, 53, 80, 82, 86, 97, 99, 100, 110, 117]. Обучающая и тестовая базы данных в идеале должны включать все разнообразие компонентов, которое может содержаться в ППТ. С другой стороны, адекватный обучающий набор признаков, формируемых на основе этих компонентов, ограничивается количеством доступных данных.

В плане наполнения базы о свойствах поверхности и их изменениях под биопленками востребованы данные о почвоподобных новообразованиях в подземных средах, модификации минералов и искусственных материалов, помещенных в почвы, поверхностных свойствах биопленок на разных стадиях жизни и др. [5, 6, 12, 19, 33, 51, 53, 82, 86, 89, 100, 105]. Кроме того, для алгоритмов МО есть необходимость в образцах, где биопленка удалена различными способами обработки, либо, наоборот, покрытыми составами (пленками), маскирующими исходную поверхность.

Некоторые особенности МО в почвоведении предполагают иную расстановку акцентов в работе с данными, чем принято в отраслях, традиционно связанных с МО (рис. 1). Вместо высоких скоростей обработки вертикальных массивов объемом выше 10^7 позиций при работе с информацией о почвах, горных породах и ППТ имеют место небольшие тщательно отобранные базы данных с высокой долей труда экспертов. Техническая сложность определения каждого показателя в таких базах зачастую превосходит сложность построения модели, что является причиной появления множества источников погрешности, включая разные виды ошибок, искусство экспериментатора и др. Систематизация данных внешних источников для дополнения собственных наблюдений во многом исключает автоматизацию процесса, поскольку предполагает структурирование данных, рассеянных по тексту в нестандартном или даже неявном виде. Некоторые признаки при этом приходится реконструировать по косвенным сведениям



Рис. 1. Некоторые принципы предобработки данных для целей МО в почвоведении.

авторов источника либо из соображений здравого смысла. Из сказанного следует, что предсказательные возможности базы данных почвенных свойств во многом зависят от заложенной избыточности признаков и разумного баланса между детальностью и надежностью описания образца.

В настоящем исследовании поставлен вопрос о возможности обучить классификатор для распознавания следов деятельности биопленок на поверхностях, присутствующих в техногенно преобразованных грунтах и свидетельствующих о развитии ППТ. Задача состояла в том, чтобы провести обучение, опираясь, в первую очередь, на признаки, формирующие “отпечаток пальца” биопленки. В исследовании проверяли гипотезу о том, что таким “отпечатком пальца” могут быть характеристики статистического распределения КУ. При этом необходимо учитывать и те признаки поверхности, которые традиционно определяются параллельно смачиваемости, т.е. отражают зависимость смачиваемости от геометрии и химического состава поверхности. Цикл работы с информацией начался с отбора собственных и достоверных внешних источников, а также формирования признаков поверхностей. Данный подход задавал некоторую фрагментарность материала, поступающего на вход моделей, т.е. для успешной классификации от алгоритма в ходе МО требовалось развить высокую обобщающую способность.

Некоторые используемые термины и сокращения:
КУ – контактный угол смачивания твердых поверхностей.

ППТ – почвоподобные тела [2].
МО – машинное обучение.
СКО – среднее квадратическое отклонение.
ROC (receiver operating characteristic) – устойчивая аббревиатура, обычно не переводится – графический подход к анализу производительности классификатора,

ROC AUC (area under curve) – площадь под кривой ROC.

Бинарные признаки – качественные или полукачественные признаки объекта, заданные как 0 – отсутствие и 1 – присутствие.

DT (Decision Tree) – алгоритм МО дерево решений.

RF (Random Forest) – алгоритм МО случайный лес.

GB (Gradient Boosting) – алгоритм МО градиентный бустинг.

KNN (k Nearest Neighbors) – алгоритм МО ближайшего соседа.

LR (Logistic Regression) – алгоритм МО логистическая регрессия.

ANN (Artificial Neural Network) – искусственная нейронная сеть.

ОБЪЕКТЫ И МЕТОДЫ

Общие принципы работы с данными. В основу разработки классификаторов положено две гипотезы: 1) устойчивые следы биопленки могут быть обнаружены на поверхностях по характерному распределению КУ (“отпечатку пальца”); 2) распределение КУ на поверхности нормально, если в источнике явно не утверждается обратное.

Цикл моделирования предполагал: 1) структурирование данных, полученных в ходе обработки источников; 2) разбиение данных на обучающую и тестовую выборки; 3) предобработку данных; 4) выбор нескольких алгоритмов МО; 5) оптимизацию и обучение алгоритмов; 6) выбор наилучшего классификатора по производительности; 7) сравнение прогноза классификатора с альтернативной экспертной диагностикой присутствия биопленки на поверхностях.

Для наполнения базы данных о свойствах поверхностей почв, горных пород, геоматериалов и искусственных материалов, подвергнутых как воздействию биопленок, так и антимикробной обработке, собственные исследования были сведены с данными открытых источников (табл. 1). Согласно задаче исследования сформировано два принципа отбора образцов в сводку. Первый – образцы должны были представлять собой почвы, ППТ или

компоненты, в них входящие; при этом учитывалось, что эти компоненты могут демонстрировать совершенно разные поверхностные свойства, находясь в разных состояниях дисперсности или шероховатости, – для учета этого фактора, например, в базу были внесены порошок кальцита, шлифованная поверхность кальцитовых кристаллов и т.п. На основе собственных наблюдений формировалась тестовая база данных для проверки качества классификаторов, а на основе внешних источников – обучающая база данных.

При обработке данных, построении и оценке моделей за ориентир были приняты следующие документы: ГОСТ Р 24668-2022 (Информационные технологии. Искусственный интеллект. Структура управления процессами для анализа больших данных), ГОСТ Р 59276-2020 (Системы искусственного интеллекта. Способы обеспечения доверия), ГОСТ Р 59898-2021 (Оценка качества систем искусственного интеллекта), ГОСТ Р 70466-2022/ISO/IEC TR 205471:2020 (Информационные технологии. Эталонная архитектура больших данных), ГОСТ Р 70462.1-2022 (Информационные технологии. Интеллект искусственный. Оценка рабочести нейронных сетей), а также документация

языка программирования Python. Все этапы цикла работы с данными от предобработки до визуализации результатов моделирования реализованы в среде для совместной разработки Google Colab (<https://colab.research.google.com>).

Объекты. Рассматривались как естественные, так и техногенные образования и искусственные материалы, имеющие в составе компоненты, присущие почвам и ППТ, распространенным в техногенно преобразованных ландшафтах мира. Образцы почв, материнских пород и ППТ географически относились к разным областям Евразии, Северной Америки, Австралии. В качестве целевой функции моделирования был принят признак наличия устойчивых следов развитой биопленки на поверхности материала. Развитой считалась биопленка, прошедшая этап необратимой адгезии к поверхности и имеющая сформированный матрикс. Такая биопленка не может быть смыта с поверхности образца водой и устойчива к слабым механическим воздействиям [25]. Это ограничение для классификации поверхностей относит образцы, зараженные микроорганизмами не более 48 ч, к незаселенным биопленками, что удобно для отсечения случайного фона. В литературе признается некоторая

Таблица 1. Обзор образцов, на базе свойств которых проведено МО алгоритмов классификации наличия устойчивых следов биопленок (в скобках – количество образцов)

| Группа образцов | Тип образцов | Ссылки на источники |
|---|---|--|
| Почвы (59) | Kastanozem, Cambisol, Chernozem, Entisol, Inceptisol, Luvisol, Oxisol, Spodosol, Podzol, Albeluvisol, Retisol, Mollisol, Alfisol, Fluvisol | С.и.*, [8, 14, 16, 27, 31, 32, 55, 98, 104, 112] |
| Материнские породы (127) | Лёссы, известняки, песчаники, мергели, карбонатные и бескарбонатные глины, пески, ферралиты, аллювиальные отложения, каолин, бентонит, филлит, сланцы, гранит, мрамор | С.и., [14, 21, 22, 24, 31, 32, 36, 45, 52, 55, 63, 74, 75, 78, 79, 87, 89, 90, 98, 103, 104, 109, 112] |
| Породообразующие минералы (71) | Кварц, слюда, гипс, гематит, гетит, ортоклаз, глиноzem, тальк, бемит, каолинит, монтмориллонит, иллит, галлуазит, кальцит, аттапульгит, ферригидрит | С.и., [8, 14, 18, 19, 31, 35, 64, 76, 80, 83, 89, 94, 96, 103, 115] |
| Минорные компоненты (102) | Уголь, бактериальные культуры, твердые органические кислоты, полисахариды, смолы, углеводороды, кремний, кунцит, гидденит, рубин, сапфир, алмаз, стекла, галогениды | С.и., [11, 14, 20, 21, 22, 23, 24, 25, 80, 29, 31, 34, 36, 42, 44, 49, 55, 56, 58, 62, 63, 65, 69, 76, 74, 77, 79, 80, 81, 82, 83, 87, 89, 90, 93, 96, 103, 104, 110, 113, 114, 115, 116, 122] |
| Ксенобиотики, искусственные материалы (122) | Сплавы, пластики, силиконы, графен | [7, 14, 23, 31, 36, 37, 39, 43, 46, 48, 49, 50, 54, 56, 58, 60, 62, 65, 66, 67, 68, 69, 70, 73, 78, 79, 81, 84, 91, 92, 93, 95, 101, 106, 108, 111, 113, 114, 117, 119, 120] |
| Специфические природные образования (19) | Мраморный оникс, пещерные образования | С.и. [97, 98] |

*С.и. – собственные измерения.

неопределенность относительно того, что считать следами биопленок в горных породах, которую разрешить пока не удалось [80, 99, 100]. Эта неопределенность связана с неопределенным сроком жизни поливидовой биопленки на поверхности, а также периодами угасания и возобновления жизнедеятельности в порах и трещинах горной породы, что во многих случаях не позволяет провести границу между следами современных биопленок и сингенетическим органическим веществом. Соответственно, данное положение вещей передает вынужденную неуверенность алгоритмам машинного обучения. Зачастую вывод древних или сильноминерализованных останков биопленки на поверхность образца сопряжен с агрессивной обработкой (абразивами, травлением), которая сама по себе изменяет поверхностные свойства [80]. В связи с этим в базу данных вносили информацию об образцах с различной степенью дисперсности (начиная от шлифов и заканчивая тонкодисперсным материалом, где это позволялось природой образца). Кроме того, учитывали информацию о способах подготовки и преимущественном составе поверхности образца.

Количественная информация об образце выражалась через распределение его КУ. Наиболее часто применяемым в почвоведении методом определения смачиваемости является метод измерения КУ на статической сидячей капле, однако наряду с ним так же применяются методы прикрепленного пузырька, динамической капли, тарелок Вильгельми, времени проникновения капли, высоты капиллярного поднятия, измерения КУ по томограммам образцов, пропитанных рассолами тяжелых элементов (КІ). Из-за важности смачиваемости поверхности для поставленной цели МО метод определения КУ был учтен в базе данных, хотя часто разные методы дают сопоставимые результаты.

Поскольку распределение КУ неочевидно (рис. 2), ряды измерений контактных углов смачивания были описаны с помощью 5 процентилей (10, 25, 50, 75, 90), экстремальных значений и среднего. В случаях, когда в источнике информации не были приведены данные, позволяющие получить указанные процентили КУ, проводили реконструкцию выборки, согласно процедуре, описанной в разделе Предобработка рядов КУ.

В базу данных вносили и другие признаки состояния поверхности, определяющие основные химические соединения в составе. Однако, в силу разнообразия состава как почв, так и ППТ, все уровни содержания всех соединений учесть было невозможно, а излишняя детальность могла повредить обобщающей способности обучаемых алгоритмов. Поэтому из признаков состава поверхности для МО были отобраны максимально обобщающие и представлены в бинарном виде.

Предобработка данных заключалась в стандартных процедурах резервирования данных для тестирования моделей, реконструкции недостающих данных, устранения дисбалансов в категориях признаков, масштабировании значений количественных признаков. Все процедуры предобработки выполняли на языке Python, преимущественно с помощью библиотек Pandas, Scikit-learn и Numpy.

Для того, чтобы обеспечить надежность тестирования модели и, следовательно, ее качество, на начальном этапе обработки данных была ликвидирована возможность их утечки. Утечкой называют частичное смешение данных тестовой базы и базы, на которой происходит МО. В представленном случае тестовая часть базы данных была отделена от обучающей части до манипуляций по реконструкции данных. В итоге для тестовой базы данных было отобрано 95 из 500 образцов по принципу наибольшего разнообразия категорий признаков. Далее для тестовой базы данных проводили только масштабирование отдельных признаков, а для обучающей – реконструкцию данных, балансировку и масштабирование признаков.

Предобработка рядов КУ. В ряде статей приведены не полные ряды измерений КУ, а указаны только средние величины, СКО и, реже, некоторые квантили. В таком случае выполняли реконструкцию ряда измерений по указанным выборочным моментам с помощью библиотеки Python Random, предполагая, вслед за авторами, нормальное распределение данных (метод Монте-Карло). Авторы настоящего исследования понимают, что гипотеза о нормальном распределении КУ очень редко проверяется в практике. Однако перебор различных способов реконструкции требует значительных затрат вычислительных и временных ресурсов, поэтому на этапе создания общего подхода к классификации поверхности на основе МО ограничились минимальным вариантом.

Было проведено масштабирование характеристик КУ с тем, чтобы значения лежали в диапазоне от 0 до 1, с помощью функции масштабирования MinMaxScale библиотеки Scikit-learn (Документация Python Scikit-learn. Предобработка. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>).

Балансировку представленности признаков по категориям производили только для тренировочной базы данных. Балансировку осуществляли с целью выполнения условия: одна категория бинарного признака должна превышать по представленности другую не более, чем в 3 раза (рис. 3). В ходе балансировки производили генерацию псевдовыборок: дублирование образцов с недостающими категориями признаков и случайную перевыборку измерений (или реконструированных значений) КУ в этих дублях с последующим вычислением стати-

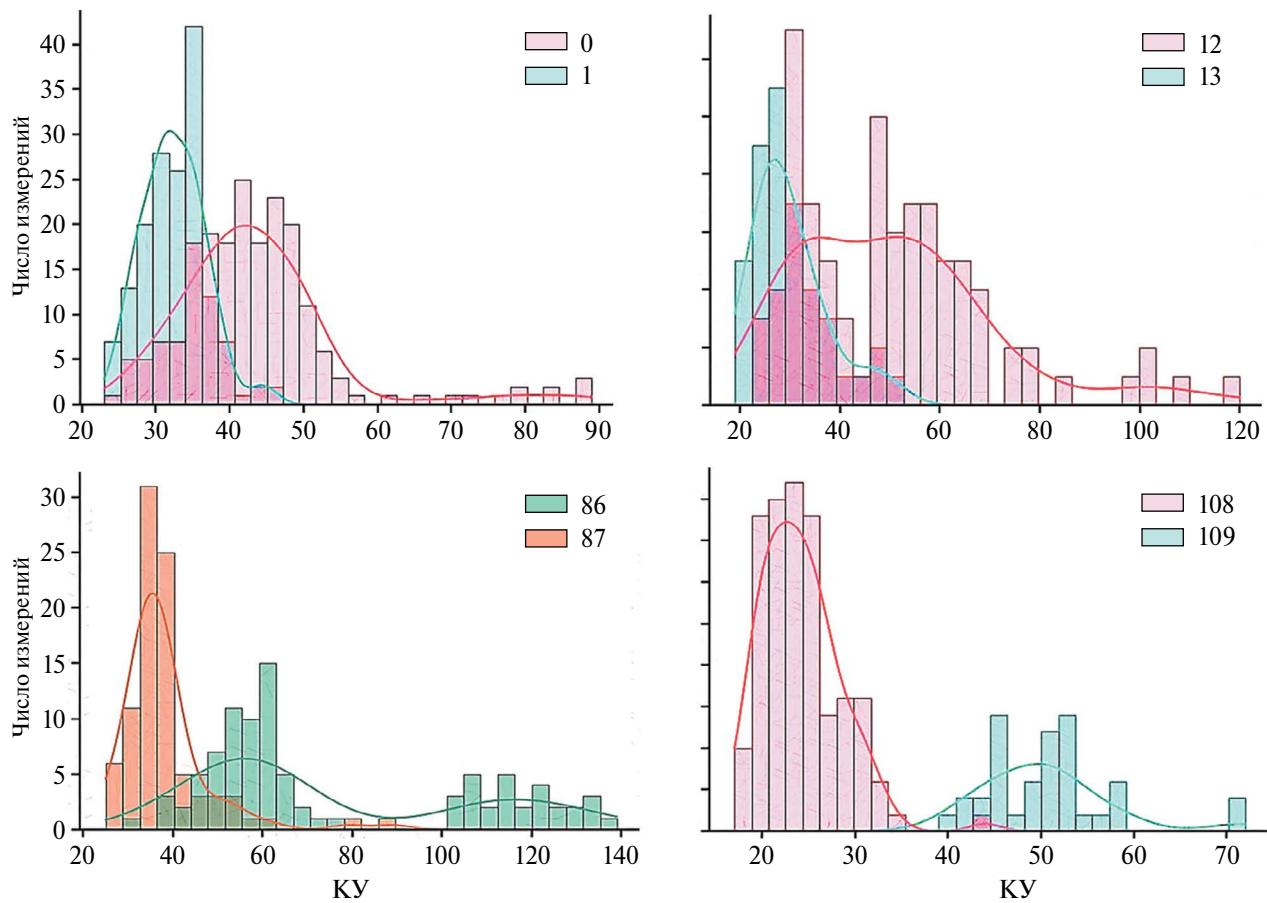


Рис. 2. Примерный вид распределений КУ на поверхностях образцов с устойчивыми следами биопленок (голубые столбики) и без них (розовые столбики) по собственным измерениям: 0 – предметное стекло: дезинфицирующий раствор и прокаливание, 1 – стекло с биопленкой, 12 – шлиф мраморного оникса, 13 – мраморный оникс с биопленкой, 86 – бентонит с биопленкой, 87 – бентонит: прокаливание, 108 – каолин: прокаливание, 109 – каолин с биопленкой.

стик для каждого. Данная процедура была реализована с помощью библиотеки Python Scikit-learn (утилита resample). Весь процесс балансировки повторяли несколько раз с доведением каждой вновь полученной обучающей базы данных до этапа проверки моделей, поскольку заложенная в процессе балансировки случайность должна была отражаться на результатах моделирования.

После устранения признаков, в которых сбалансировать категории не представлялось возможным, в рабочей базе данных осталось 8 характеристик КУ и 8 бинарных признаков поверхности (табл. 2). Более подробное разбиение качественных признаков на категории, а также учет признаков, связанных с методиками измерения КУ, в настоящий момент неосуществимы вследствие недостатка сведений, однако в структуру базы данных эта возможность включена.

Выбор типа и архитектуры модели МО осуществляли, исходя из пригодности для работы с не-

большими наборами данных, простоты реализации, производительности и интерпретируемости результатов моделирования. Всего было протестировано 6 алгоритмов: классификация методом ближайшего соседа (KNN), логистическая регрессия (LR), дерево решений (DT), случайный лес (RF), градиентный бустинг (GB) и глубокая нейронная сеть (ANN). Первые три алгоритма просты и легко интерпретируются, но более склонны к переобучению. Последние три представляют собой “черные ящики”, поэтому сложнее интерпретируются, однако содержат возможности контроля переобучения и, соответственно, способны развить большую обобщающую способность в ходе обучения. Построение нейронной сети реализовано с помощью библиотеки Python TensorFlow Keras, а остальные алгоритмы – с помощью Python Scikit-learn. В качестве базового варианта был выбран классификатор на основе дерева решений (как наиболее простой со всех точек зрения), с которым сравнивался результат работы остальных классификаторов.

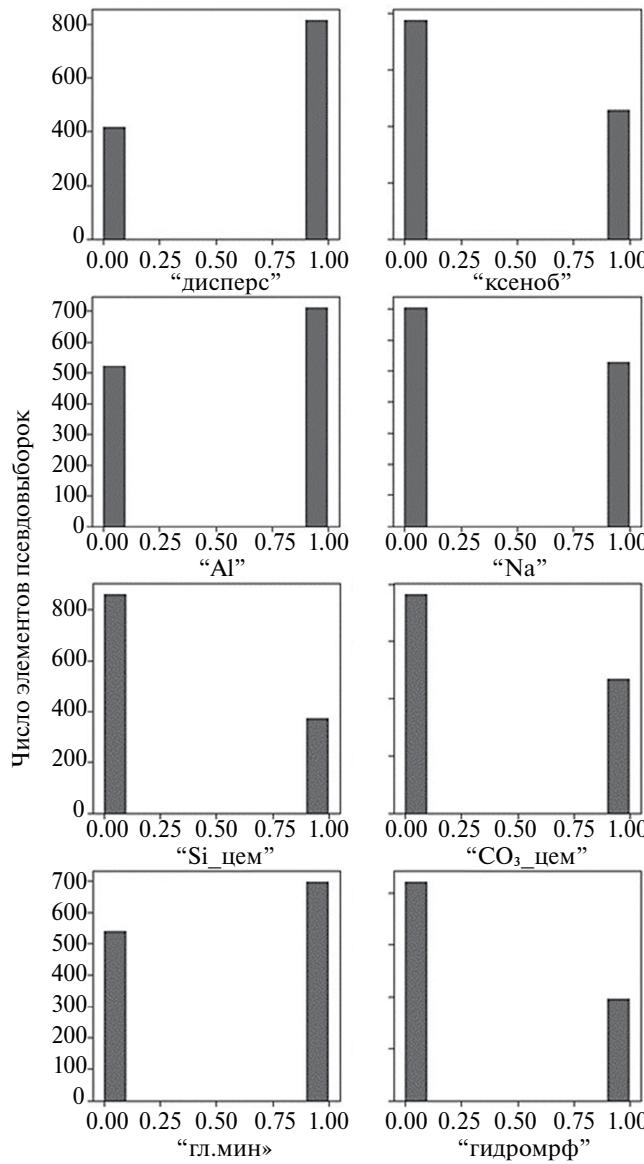


Рис. 3. Сбалансированные признаки поверхности образцов, использованных для МО. Расшифровка обозначений признаков – в табл. 2.

Оптимизацию гиперпараметров классификаторов производили автоматически, исходя из максимизации информационного выигрыша методом уполовиненного перекрестного поиска по сетке гиперпараметров HalvingGridSearchCV (документация Python Scikit-learn. Выбор модели на основе HalvingGridSearchCV. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.HalvingGridSearchCV.html). Для коррекции классификаторов в ходе обучения применяли кросс-валидацию – проверку качества модели в ходе обучения на случайно выделяемой установленной доле обучающей выборки.

Таблица 2. Признаки поверхности образцов почв, горных пород и других материалов, учтенные классификаторами

| Обозначение признака | Расшифровка обозначения признака |
|------------------------|--|
| “следы” | Наличие устойчивых следов биопленки – целевая функция (1 – есть, 0 – нет) |
| “дисперс” | Дисперсность (1 – дисперсное вещество, 0 – слитая поверхность) |
| “гл.мин” | Глинистые минералы в макроколичествах (1 – есть, 0 – нет) |
| “AI” | Соединения алюминия в макроколичествах (1 – есть, 0 – нет) |
| “Si_цем” | Силикатный цемент (1 – есть, 0 – нет) |
| “CO ₃ _цем” | Карбонатный цемент (1 – есть, 0 – нет) |
| “Na” | Соединения натрия в макроколичествах (1 – есть, 0 – нет) |
| “ксеноб” | Органическое вещество, не относящееся к современным биопленкам – уголь, нефть, искусственные полимеры, кремнийорганика (1 – есть, 0 – нет) |
| “гидромрф” | Признаки гидроморфизма: сведения о восстановительной обстановке, затоплении, болотных и аллювиальных почвах и грунтах (1 – есть, 0 – нет) |
| “КУ_ср” | Среднее выборочное значение КУ |
| “СКО” | СКО по выборке КУ |
| “КУ_мин” | Минимальный наблюдаемый КУ для образца |
| “КУ_макс” | Максимальный наблюдаемый КУ для образца |
| “КУ_10” | Процентиль 10 для выборки КУ |
| “КУ_25” | Процентиль 25 для выборки КУ |
| “КУ_50” | Процентиль 50 для выборки КУ (медиана) |
| “КУ_75” | Процентиль 75 для выборки КУ |
| “КУ_90” | Процентиль 90 для выборки КУ |

Далее оценивали значимость каждого фактора для способности модели различать классы образцов: наличие и отсутствие следов биопленки. Для этого использовали уменьшение ошибки предсказания при учете очередного фактора (значимость по Джини) [85].

Архитектуру нейронной сети подбирали, исходя из минимизации числа задействованных нейронов, поскольку база данных невелика. Для этого, в частности, использовали слои типа Dropout, где часть нейронов случайным образом отключалась. При компиляции модели подбирали размер пачки данных (batch), оптимизатор и его гиперпараметры.

Выбор происходил, исходя из устойчивости модели к переобучению, правильности предсказаний и скорости обучения. В процессе обучения происходило случайное перемешивание образцов внутри тренировочной базы данных, поэтому один и тот же образец мог несколько раз попасть в разные пачки данных.

Качество моделей оценивали, во первых, с помощью меры правильности предсказаний (accuracy), которая выражает долю правильных предсказаний в общем количестве предсказаний (Документация Python Scikit-learn https://scikit-learn.org/stable/modules/model_evaluation.html). Во-вторых, процесс работы модели был визуализирован с помощью кривой ROC, и использована метрика площади, ограничивающейся этой кривой – ROC AUC. ROC-кривая служит для определения того, работает ли классификатор лучше, чем случайное угадывание (прямая $y = x$) [102]. ROC AUC модели наилучшего качества стремится к 1. Кроме того, сравнение качества моделирования разными алгоритмами проводили с помощью матриц неточности предсказаний (confusion matrix) [88]. Порог отсечения на матрице неточности подобран таким образом, чтобы доли ложных предсказаний в отрицательном и положительном классе были максимально близкими.

Таким образом, полная последовательность численных экспериментов с обучающей базой данных сводилась к следующему:

- балансировка категорий в признаках, осложненная случайной перевыборкой КУ из исходных выборок по образцам;
- вариация алгоритмов МО и их гиперпараметров; здесь же – вариация архитектуры нейронной сети.

РЕЗУЛЬТАТЫ

Образцы были отражены в настолько различных состояниях, насколько позволила доступная информация: дисперсном, слитом, обработанном отжигом, облучением, химическими реагентами и покрытиями.

После балансировки категорий в каждом из признаков обучающая база данных, приведшая к наилучшему качеству обучения, содержала 1233 объекта.

Результаты численных экспериментов с подбором гиперпараметров для обучающихся алгоритмов приведены в табл. 3.

Наиболее обучаемой оказалась нейронная сеть с 5 скрытыми слоями, из которых 2 – полносвязанных (Dense) и 3 – со случайно выключенными нейронами (Dropout); всего работало 125 нейронов (см. приложение). Функциями активации нейронов в полносвязанных слоях были выбраны

в порядке следования слоев от входного к выходному: ReLU (rectified linear unit, или “линейный выямитель”) – softmax (функция, преобразующая список произвольных чисел в список вероятностей, пропорциональных этим числам) – ReLU – tanh (гиперболический тангенс). В слоях типа Dropout доля выключенных нейронов варьировалась от 0.5 до 0.8.

После экспериментов оказалось, что оптимизатор Adam приводил к наилучшим результатам при использовании с гиперпараметрами, установленными по умолчанию.

При анализе работы классификаторов DT и RF обнаружено, что наиболее значимыми признаками для принятия решений этими алгоритмами являлись: присутствие глинистых минералов, карбонатного цемента, органического вещества, не относящегося к биопленкам, шероховатость. Из характеристик КУ классификаторы в первую очередь опирались на выборочное среднее и процентиля 10, 25, 75. Наименьшее значение для классификации имело наличие макроподколичеств натрия и силикатного цемента.

ОБСУЖДЕНИЕ

Сравнение обученных моделей с помощью метрик качества. По совокупности показателей ROC AUC, правильности, долей верных предсказаний в категориях целевой функции “отсутствие устойчивых следов биопленки” и “наличие следов биопленки” (порог отсечения 0.5), времени обучения, наилучшим можно признать алгоритм нейронной сети с выбранной архитектурой (табл. 4). Этот алгоритм относится к глубокому обучению, т.е. самостоятельно извлекает диагностические признаки из предоставленных данных. Алгоритмы логистической регрессии, градиентного бустинга и ближайшего соседа не имели других преимуществ, кроме быстродействия. Дерево решений относительно медленно проходило оптимизацию гиперпараметров и обучение, при этом демонстрируя наибольшую правильность предсказаний. Случайный лес показал второй результат по ROC AUC при более быстрой обучаемости. Тем не менее с учетом того, что запланировано дальнейшее дообучение модели на вновь поступающих данных, предпочтение отдается ANN.

Кривая ROC для ANN показывает, что алгоритм достаточно сильно отклоняется от случайного угадывания тестовых данных о наличии биопленки (т.е. прямой линии $y = x$). Это позволяет заключить, что обучение прошло успешно (рис. 4).

Несогласие классификации образцов между обученными алгоритмами и экспертами-людьми. Свойства образцов, для которых наблюдалось несогласие между классификацией, данной экспертами-людьми и практически всеми обученными

Таблица 3. Оптимальные характеристики обученных алгоритмов классификации образцов на наличие устойчивых следов биопленок (расшифровка обозначений – в тексте)

| Алгоритм | Оптимизируемые гиперпараметры | Гиперпараметры наилучших классификаторов |
|----------|---|---|
| ANN | Размер пачки данных (batch) для одной итерации обучения | 64 |
| | Тип оптимизатора | tf.keras.optimizers.Adam (документация PyTorch. Оптимизатор Adam. https://pytorch.org/docs/stable/generated/torch.optim.Adam.html) |
| | Скорость обучения | 0.01 |
| | Количество эпох обучения | 300–400 |
| RF | Количество деревьев | 27 |
| | Максимальная глубина дерева | 14–15 |
| | Минимальное число ветвлений | 6–8 |
| | Минимальное число листьев дерева | 1–3 |
| DT | Максимальная глубина дерева | 8 |
| | Минимальное число ветвлений | 2–3 |
| | Минимальное число листьев дерева | 2–3 |
| LR | C – параметр регуляризации | 7 |
| GB | Количество деревьев | 49 |
| | Максимальная глубина дерева | 7 |
| | Скорость обучения | 0.1 |
| | Доля тренировочной выборки для валидации | 0.2 |
| KNN | Количество ближайших соседей | 1–3 |

Таблица 4. Производительность алгоритмов классификации образцов на наличие устойчивых следов биопленок (расшифровка обозначений – в тексте)

| Алгоритм | ROC AUC | Правильность предсказаний | Верно предсказано отсутствие биопленки | Верно предсказано наличие биопленки | Время обучения, с |
|----------|---------|---------------------------|--|-------------------------------------|-------------------|
| | | | % всех предсказаний | | |
| ANN | 0.80 | 0.74 | 73 | 75 | 84.5 |
| RF | 0.73 | 0.72 | 77 | 71 | 34.8 |
| DT | 0.71 | 0.75 | 77 | 66 | 94.2 |
| LR | 0.70 | 0.70 | 70 | 71 | 0.9 |
| GB | 0.69 | 0.72 | 63 | 77 | 0.8 |
| KNN | 0.65 | 0.66 | 63 | 68 | 7.7 |

алгоритмами, приведены в табл. 5. В обобщенном виде сложный для классификации образец представляет собой дисперсный карбонатно-алюминатный грунт со следами микробной деятельности, например, карбонатные почвы, мергели.

В некоторых из учтенных случаев несогласия невозможно однозначно утверждать, что ошибки исходят со стороны алгоритмов МО. Так, экспертная оценка предполагала, что почвы и природные грунты, не подверженные агрессивной обработке,

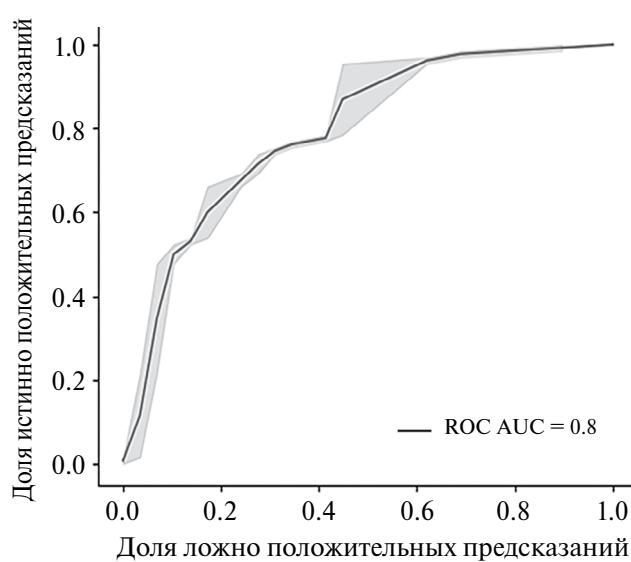


Рис. 4. Кривая ROC для нейросети – классификатора образцов на наличие устойчивых следов биопленок с оптимальными характеристиками.

всегда имеют устойчивые следы биопленок. Однако алгоритмы отнесли к категории “отсутствие следов биопленки” некоторые образцы бентонита, карбонатных глин, аллювиальных почв и грунтов, подпочвенного слоя ферралита, а также образец чернозема, который несколько лет экспонировался со следовыми количествами нефти. Все эти образцы долгое время хранились в воздушно-сухом состоянии, что могло отчасти уничтожить следы биопленок. К случаям несогласия привел и сдвиг условий экспозиции затопленных образцов в сторону анаэробности – без дополнительной стимуляции питательными средами в таких условиях формировалось особое микробное сообщество, что было расценено алгоритмами МО как “отсутствие следов биопленки”. Кроме того, затруднения у

алгоритмов вызвали спелеоотмы, обработанные соляной кислотой, щелочью и спиртом, а также типа “пещерное молочко”. Экспертные оценки квалифицировали эти образцы как содержащие устойчивые следы микробной деятельности, поскольку в ходе анализов было зафиксировано обилие микроорганизмов. Однако вполне возможно, что воздействие на образцы привело к серьезному нарушению целостности и уничтожению компонентов биопленочного матрикса. Или, в случае “пещерного молочка”, имевшего состояние геля, биопленка могла изначально испытывать слишком слабую адгезию к поверхности, которая нарушалась при любом механическом воздействии и прекращала влиять на смачиваемость после подготовки образца к измерениям КУ. Приведенные случаи несогласия оценок нуждаются в целенаправленном и более тщательном изучении образцов со свойствами, аналогичными описанным – например, биоматов.

В то же время в подавляющем большинстве случаев наблюдалось совпадение оценок наличия следов биопленок, данных экспертами-людьми и обученной нейронной сетью. Этот факт говорит о принципиальной возможности МО на уже имеющихся данных о свойствах поверхности почв, горных пород и других материалов, а наблюдаемые сложности классификации помогают конкретизировать аспекты будущих исследований.

ЗАКЛЮЧЕНИЕ

Предложен подход к обучению и проведено пилотное обучение 6 классификаторов для распознавания следов деятельности биопленок на поверхностях, которые могут свидетельствовать о развитии почвоподобных тел внутри и на поверхности грунтов разного происхождения. Разрабатываемый подход опирается на легко определяемые признаки. Обосновано предположение о том, что неоднородность смачиваемости поверхности геологических объектов служит важным индикатором

Таблица 5. Представленность признаков поверхности в тестовой базе данных в случаях несогласия классификаций и в общем числе образцов (расшифровка обозначений в табл. 2)

| Признак поверхности | Доля случаев наличия признака поверхности (значение признака = 1), % | |
|------------------------|---|---|
| | при несогласии классификаций между экспертом-человеком и обученными алгоритмами | в общем числе образцов тестовой базы данных |
| “следы” | 82 | 67 |
| “дисперс” | 77 | 54 |
| “CO ₃ _цем” | 77 | 45 |
| “Al” | 64 | 45 |
| “Si_цем” | 27 | 43 |

следов микробной деятельности. В процесс обучения классификаторов вошли экспертные оценки наличия следов биопленки, 8 характеристик статистического распределения контактного угла смачивания и 9 обобщающих категориальных признаков геометрии, минерального состава, состояния органического вещества поверхности для таких объектов, как почвы, материнские породы, породообразующие минералы, сопутствующие компоненты почвообразования, ксенобиотики. Несмотря на фрагментарность входных данных, удалось достичь хорошей обобщающей способности классификаторов, обученных распознаванию следов микробной деятельности на твердых поверхностях, следующих типов: нейронная сеть со 125 нейронами, дерево решений и случайный лес. Эти классификаторы демонстрировали правильность классификации тестовых данных на уровне 0.72–0.75 и ROC AUC 0.71–0.80. Сформированы ограничения для моделей на современном этапе: все признаки поверхности, кроме смачиваемости, можно выразить как бинарные, исследуемый материал не разрушается под действием воды, обнаруживаемые следы биопленки устойчивы во времени к смыву водой и слабым механическим воздействиям. Разнообразие образцов, на которых обучались модели, достаточно невелико, поэтому есть необходимость в продолжении их обучения для повышения качества классификации. Наиболее перспективна в плане прогресса обучения модель нейронной сети со случайно отключаемыми нейронами. Определено, что сбор новых данных в целях продолжения обучения классификатора должен быть акцентирован на дисперсных карбонатных образцах в различных вариантах химической, физической и биологической обработки. В рамках развития подхода к диагностике почвоподобных тел рациональной представляется замена бинарного разбиения присутствия/отсутствия следов микробной деятельности на несколько категорий, соответствующих типам следов биопленок на поверхности, в зависимости от стадии зрелости, функционального состава и условий жизнедеятельности. Таким образом, данный подход можно использовать для уточнения механизмов почвообразования, биогеохимических и биогеотехнологических процессов в грунтах различного происхождения, в том числе терраформирования.

БЛАГОДАРНОСТЬ

Авторы выражают благодарность Дмитрию Колдезеву, лектору Open Data Science, за ценную информацию о разработке систем машинного обучения.

ФИНАНСИРОВАНИЕ РАБОТЫ

Данная работа финансировалась Программой стратегического академического лидерства Казанского федерального университета (ПРИОРИТЕТ-2030).

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

В данной работе отсутствуют исследования человека или животных.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют о том, что у них нет конфликта интересов.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Онлайн-версия содержит дополнительные материалы, доступные по адресу
<https://doi.org/10.31857/S0032180X25020086>

СПИСОК ЛИТЕРАТУРЫ

1. Алексеев И.В. Развитие комплексного инженерно-геологического и микробиологического мониторинга на Яковлевском руднике для повышения безопасности ведения очистных работ под неосущенными водоносными горизонтами. дис. канд. геол.-минерал. наук. СПб., 2015.
2. Горячkin С.В., Мергелов Н.С., Таргульян В.О. Генезис и география почв экстремальных условий: элементы теории и методические подходы // Почвоведение. 2019. № 1. С. 5–19.
<https://doi.org/10.1134/S0032180X19010040>
3. Дацко Р.Э., Котюков П.В. Инженерно-геологическое обеспечение эксплуатационной надежности подземных транспортных сооружений в Санкт-Петербурге // Записки Горного института. 2011. Т. 190. С. 71–77.
<http://elibrary.ru/item.asp?id=20876963>
4. Дацко Р.Э., Норова Л.П., Руденко Е.С. Эволюция геоэкологического состояния подземного пространства Санкт-Петербурга // Разведка и охрана недр. 1998. № 7–8. С. 57–59.
5. Зорина А.С. Биопленки нитрилгидролизующих бактерий *Alcaligenes faecalis* 2 и *Rhodococcus ruber* Gt 1 в процессах трансформации нитрилов и амидов карбоновых кислот. Дис. канд. биол. наук. Пермь, 2020.
6. Иноземцев С.А., Таргульян В.О. Верхнепермские палеопочвы: свойства, процессы, условия формирования. М.: ГЕОС, 2009. 188 с.
7. Кабов О.А., Зайцев Д.В. Влияние гистерезиса смачивания на растекание капли под действием гравитации // Доклады Академии Наук. 2013. Т. 451. № 1. С. 37–40.
<https://doi.org/10.7868/S0869565213190122>

8. Матвеева Н.В., Милановский Е.Ю., Рогова О.Б. Способ подготовки образцов почв для определения контактного угла смачивания методом сидячей капли // Бюл. Почв. ин-та им. В.В. Докучаева. 2019. Вып. 97. С. 91–112. <https://doi.org/10.19047/0136-1694-2019-97-91-112>
9. Муравьев Ф.А., Винокуров В.М., Галеев А.А., Булка Г.Р., Низамутдинов Н.М., Хасанова Н.М. Парамагнетизм и природа рассеянного органического вещества в пермских отложениях Татарстана // Георесурсы. 2006. № 2(19). С. 40–45.
10. Новоселов А.А., Константинов А.О. Карбонатные коры на фасадах зданий и сооружений города Тюмени: разнообразие и факторы формирования // Известия Томского политех. ун-та. Инжиниринг георесурсов. 2019. Т. 330. № 3. С. 40–49. <https://doi.org/10.18799/24131830/2019/3/163>
11. Русанов А.И., Есипова Н.Е., Соболев В.Д. Сильная зависимость краевого угла от давления // Доклады Академии наук. 2019. Т. 487. № 2. С. 169–173. <https://doi.org/10.31857/S0869-56524872169-173>
12. Семиколенных А.А., Таргульян В.О. Почвоподобные тела автохемолитотрофных экосистем пещер хребта Кугитантау (Восточный Туркменистан) // Почвоведение. 2010. № 6. С. 658–672. <https://doi.org/10.1134/S0032180X19010040>
13. Сидоренко С.А. Органическое вещество и биогенетические процессы в докембрии. М.: Наука, 1991. 104 с.
14. Софинская О.А., Костерин А.В., Галеев А.А. Неоднородность смачивания поверхности гидрофобизированных почв и почвообразующих пород // Почвоведение. 2022. № 3. С. 326–336. <https://doi.org/10.31857/S0032180X22030133>
15. Флоровская В.Н. Люминесцентно-битуминологический метод в нефтяной геологии. М.: Изд-во Моск. ун-та, 1957. 293 с.
16. Холодов В.А., Ярославцева Н.В., Яшин М.А., Фрид А.С., Лазарев В.И., Тюгай З.Н., Милановский Е.Ю. Контактные углы смачивания и водоустойчивость почвенной структуры // Почвоведение. 2015. № 6. С. 693–701. <https://doi.org/10.7868/S0032180X15060064>
17. Шапиро Т.Н., Дольникова Г.А., Немцева Н.В., Санджиева Д.А., Лобакова Е.С. Идентификация и физиологическая характеристика консорциума углеводородокисляющих бактерий нефти и нефтепродуктов // Журн. микробиол. 2018. № 4. С. 107–113. <https://doi.org/10.36233/0372-9311-2018-4-107-113>
18. Шеин Е.В., Верховцева Н.В., Быкова Г.С., Пашкевич Е.Б. Агрегатообразование в каолинитовой суспензии при микробиологической модификации поверхности глины // Почвоведение. 2020. № 3. С. 351–357. <https://doi.org/10.31857/S0032180X20030077>
19. Achtenhagen J., Goebel M.-O., Miltner A., Kaestner M. Bacterial impact on the wetting properties soil minerals // Biogeochemistry. 2015. V. 122(2-3). <https://doi.org/10.1007/s10533-014-0040-9>
20. Ahmed N., Siow K.S., Wee M.F.M.R. et al. A study to examine the ageing behaviour of cold plasma-treated agricultural seeds // Sci. Rep. 2023. V. 13. P. 1675. <https://doi.org/10.1038/s41598-023-28811-w>
21. Alhammadi A.M., AlRatrou A., Singh K. et al. In situ characterization of mixed-wettability in a reservoir rock at subsurface conditions // Sci. Rep. 2017. V. 7. P. 10753. <https://doi.org/10.1038/s41598-017-10992-w>
22. AlRatrou A., Blunt M. J., Bijeljic B. Spatial correlation of contact angle and curvature in pore-space images // Water Res. Res. 2018. V. 54. P. 6133–6152. <https://doi.org/10.1029/2017WR022124>
23. AlRatrou A., Blunt M. J., Bijeljic B. Wettability in complex porous materials, the mixed-wet state, and its relationship to surface roughness // Proc. Natl. Acad. Sci. USA. 2018. Sep. 4. V. 115(36). P. 8901–8906. <https://doi.org/10.1073/pnas.1803734115>
24. AlRatrou A., Raeini Q.A., Bijeljic B., Blunt M. Automatic measurement of contact angle in pore-space images // Adv. Water Res. 2017. P. 109. <https://doi.org/10.1016/j.advwatres.2017.07.018>
25. Andryukov B.G., Romashko R. V., Efimov T.A. et al. Mechanisms of adhesive-cohesive interaction of bacteria in the formation of a biofilm // Molecular Genetics, Microbiology and Virology. 2020. V. 35(4). P. 195–201. <https://doi.org/10.17116/molgen202038041155>
26. Arvind K.J. Microbiological processes in improving the behavior of soils for civil engineering applications: a critical appraisal // Journal of hazardous, toxic, and radioactive waste. 2022. V. 26(2). [https://doi.org/10.1061/\(ASCE\)HZ.2153-5515.0000686](https://doi.org/10.1061/(ASCE)HZ.2153-5515.0000686)
27. Aslam T., Deurer M., Müller K. et al. Does an increase in soil organic carbon improve the filtering capacity of aggregated soil for organic pesticides? A case study // Geoderma. 2009. V. 152. P. 187–193. <https://doi.org/10.1016/j.geoderma.2009.06.015>
28. Assadi-Langroudi A., O'Kelly B.C., Barreto D. et al. Recent advances in nature-inspired solutions for ground engineering (NiSE) // Int. J. Geosynthetics Ground Engineer. 2022. V. 8. <https://doi.org/10.1007/s40891-021-00349-9>
29. Atherton S., Polak D., Hamlett C. et al. Drop impact behaviour on alternately hydrophobic and hydrophilic layered bead packs // Chem. Engineering Res. Design. 2016. V. 110. P. 200–208. <https://doi.org/10.1016/j.cherd.2016.02.011>
30. Bachmann J., Goebel M.O. Soil water repellency / in Encyclopedia of Soils in the Environment. Elsevier, 2023. V. 5. P. 203–215. <https://doi.org/10.1016/B978-0-12-822974-3.00116-6>

31. *Bachmann J., McHale G.* Superhydrophobic surfaces: a model approach to predict contact angle and surface energy of soil particles // *Eur. J. Soil Sci.* 2009. V. 60(3). P. 420–430.
<https://doi.org/10.1111/j.1365-2389.2008.01118.x>
32. *Bachmann J., Woche S. K., Goebel M.-O. et al.* Extended methodology for determining wetting properties of porous media // *Water Res. Res.* 2003. V. 39.
<https://doi.org/10.1029/2003WR002143>
33. *Banks E.D., Taylor N.M., Gulley J. et al.* Bacterial calcium carbonate precipitation in cave environments: a function of calcium homeostasis // *Geomicrobiology J.* 2020. V. 27(5). P. 444–454.
<https://doi.org/10.1080/01490450903485136>
34. *Borah M.P., Kalit B.B., Jose S. et al.* Fabrication of hydrophobic surface on Eri silk/wool fabric using nano silica extracted from rice husk // *Silicon.* 2023. V. 15. P. 7039–7046.
<https://doi.org/10.1007/s12633-023-02568-3>
35. *Chao Z., Zhen L., Peng D.* Contact angle of soil minerals: A molecular dynamics study // *Computers and Geotechnics.* 2016. V. 75. P. 48–56.
<https://doi.org/10.1016/j.compgeo.2016.01.012>
36. *Chen J.H.M., Liu Y.* A facile and straightforward immersion approach to enhance the hydrophobicity of melamine sponge for efficient cleanup of crude oils and organic solvents // *J. Porous Mater.* 2024. V. 31. P. 587–596.
<https://doi.org/10.1007/s10934-023-01540-1>
37. *Chen M., Wu D., Chen D. et al.* Experimental investigation on the movement of triple-phase contact line during a droplet impacting on horizontal and inclined surface // *Chemical Engineering Science.* 2020. V. 226. P. 115864.
<https://doi.org/10.1016/j.ces.2020.115864>
38. *Dejong J.T., Kavazanjian, E.* Bio-mediated and Bio-inspired Geotechnics // *Geotechnical fundamentals for addressing new world challenges. Series in Geomechanics and Geoengineering.* Springer, 2019. P. 193–207.
https://doi.org/10.1007/978-3-030-06249-1_7
39. *Deshpande R.A., Navne J., Adelmark M.V. et al.* Understanding the light induced hydrophilicity of metal-oxide thin films // *Nat. Commun.* 2024. V. 15. P. 124.
<https://doi.org/10.1038/s41467-023-44603-2>
40. *Diehl D.* Soil water repellency: Dynamics of heterogeneous surfaces // *Colloids and Surfaces A: Physico-chem. Eng. Aspects.* 2013. V. 432. P. 8–18.
<http://dx.doi.org/10.1016/j.colsurfa.2013.05.011>
41. *Dorobantu L., Bhattacharjee S., Foght J.M., Gray M.R.* Atomic force microscopy measurement of heterogeneity in bacterial surface hydrophobicity // *Langmuir.* 2008. V. 24(9). P. 4944–4951.
<https://doi.org/10.1021/la7035295>
42. *Eberlein C., Baumgarten T., Starke S., Heipieper H. J.* Immediate response mechanisms of Gram-negative solvent-tolerant bacteria to cope with environmental stress: cis-trans isomerization of unsaturated fatty acids and outer membrane vesicle secretion // *Appl. Microbiol. Biotechnol.* 2018. V. 102. P. 2583–2593.
<https://doi.org/10.1007/s00253-018-8832-9>
43. *Fahland M., Mishra R.* Computational model and simulations of contact angle and geometry effects on centrifugal microfluidic step-emulsification // *Microfluid Nanofluid.* 2023. V. 27. P. 59.
<https://doi.org/10.1007/s10404-023-02666-z>
44. *Farber L., Al-Aaraj, D., Smith R., Gentzler M.* Formation and internal microstructure of granules from wetting and non-wetting efavirenz/ ron oxide blends // *Chem. Engineering Sci.* 2020. V. 227. P. 115909.
<https://doi.org/10.1016/j.ces.2020.115909>
45. *Fér M., Leue M., Kodešová R., Gerke H., Ellerbrock R.* Droplet infiltration dynamics and soil wettability related to soil organic matter of soil aggregate coatings and interiors // *J. Hydrology and Hydromechanics.* 2016. V. 64(2). P.
<https://doi.org/10.1515/johh-2016-0021>
46. *Francone A., Merino S., Retolaza A. et al.* Impact of surface topography on the bacterial attachment to micro- and nano-patterned polymer films // *Surfaces and Interfaces.* 2021. V. 27. P. 101494.
<https://doi.org/10.1016/j.surfin.2021.101494>
47. *Galeev A.A., Vinokurov V.M., Mouraviev F.A., Osin Y. N.* EPR and SEM study of organo-mineral associations in Lower Permian evaporite dolomites // *Appl. Magnetic Resonance.* 2019. V. 35. P. 473–479.
<https://doi.org/10.1007/s00723-009-0178-0>
48. *Gao D., Wang F., Lyu B. et al.* Multifunctional cotton fabric with durable antibacterial, superhydrophobicity, and UV resistance based on Ag@TiO₂ Janus nanoparticles // *Cellulose.* 2024. V. 31. P. 2617–2633.
<https://doi.org/10.1007/s10570-023-05727-2>
49. *Ghodrati M., Mousavi-Kamazani M., Bahrami Z.* Synthesis of superhydrophobic coatings based on silica nanostructure modified with organosilane compounds by sol-gel method for glass surfaces // *Sci. Rep.* 2023. V. 13. P. 548.
<https://doi.org/10.1038/s41598-023-27811-0>
50. *Gordon C.* Contact angle distribution of particles at fluid interfaces // *Langmuir: the ACS J. Surfaces Colloids.* 2014. V. 31. P. 891–897.
<https://doi.org/10.1021/la5040195>
51. *Gray C.J., Engel A.S.* Microbial impact on aquifer carbonate geochemistry // *The ISME J. Int. Soc. Microbial. Ecology.* 2013. V. 7. P. 325–337.
<https://doi.org/10.1038/ismej.2012.105>
52. *Guo R., Dalton L., Fan M. et al.* The role of the spatial heterogeneity and correlation length of surface wettability on two-phase flow in a CO₂-Water-Rock System // *Adv. Water Res.* 2020. V. 146. P. 103763.
<https://doi.org/10.1016/j.advwatres.2020.103763>
53. *Guvenesen N.C., Demir S., Ozdemir G.* Effects of magnesium and calcium cations on biofilm formation by *Sphingomonas paucimobilis* from an industrial environment // *Current*

- Opinion in Biotechnology. 2013. V. 24(1). P. S68. <https://doi.org/10.1016/j.copbio.2013.05.185>
54. Haider S.A., Raj A. Liquid drops on compliant and non-compliant substrates: an ellipsoid-based fitting for approximating drop shape and volume // Microfluid Nanofluid. 2023. 27. P. 49. <https://doi.org/10.1007/s10404-023-02659-y>
55. Hamlett C., Atherton S., Shirtcliffe N. et al. Transitions of water-drop impact behaviour on hydrophobic and hydrophilic particles // Eur. J. Soil Sci. 2013. V. 64. P. 324–333. <https://doi.org/10.1111/ejss.12003>
56. Hao X., Yao H., Zhang P. et al. Multifunctional solar water harvester with high transport selectivity and fouling rejection capacity // Nat. Water. 2023. V. 1. P. 982–991. <https://doi.org/10.1038/s44221-023-00152-y>
57. Hark R., Harmon R.S. Geochemical Fingerprinting Using LIBS // Springer Series in Optical Sciences. 2014. V. 182. P. 309–344. <https://doi.org/10.1007/978-3-642-45085-3-12>
58. Hassanloofard Z., Gharekhani M., Zandi M. et al. Fabrication and characterization of cellulose acetate film containing *Falcaria vulgaris* extract // Cellulose. 2023. V. 30. P. 6833–6853. <https://doi.org/10.1007/s10570-023-05337-y>
59. Hata T., Tsukamoto M., Mori H., Kuwano R., Gourc J.P. Evaluation of multiple soil improvement techniques based on microbial functions // Proc. GeoFrontiers Adv. Geotechnical Engineering. Dallas. 2011. V. 211. P. 3945–3955. [https://doi.org/10.1061/41165\(397\)403](https://doi.org/10.1061/41165(397)403)
60. Hiremani V., Goudar N., Gasti T. et al. Exploration of multifunctional properties of piper betel leaves extract incorporated polyvinyl alcohol-oxidized maize starch blend films for active packaging application // J. Polymers Env. 2022. V. 30. P. 1314–1329. <https://doi.org/10.1007/s10924-021-02277-1>
61. Hoefs J. Geochemical fingerprints: a critical appraisal // Eur. J. Mineral. 2010. V. 22. P. 3–15. <https://doi.org/10.1127/0935-1221/2010/0022-1997>
62. Huhtamäki T., Tian X., Korhonen J., Ras R. Surface-wetting characterization using contact-angle measurements // Nature Protocols. 2018. V. 13. P. 1521–1538. <https://doi.org/10.1038/s41596-018-0003-z>
63. Ibrahim A., Elkhatatny S. Data-driven models to predict shale wettability for CO₂ sequestration applications // Sci. Rep. 2023. 13. <https://doi.org/10.1038/s41598-023-37327-2>
64. Jańczuk B., Białopiotrowicz T. Components of surface free energy of some clay minerals // Clays and Clay Minerals. 1988. V. 36. P. 243–248.
65. Jung H., Kim K., Ko J.-H. Effect of a marine bacterial biofilm on adhesion and retention of pseudo barnacle to silicone coating surface // Korean J. Chem. Engineering. 2014. V. 31. P. 262–267. <https://doi.org/10.1007/s11814-013-0218-1>
66. Kandukuri P., Deshmukh S., Katiresan S. Influence of the static contact angle on the liquid film coverage for falling-film systems // Flow Turbulence Combust. 2023. V. 111. P. 1253–1277. <https://doi.org/10.1007/s10494-023-00484-5>
67. Kirichenko E., Gatapova E. Studying of the contact angle hysteresis on various surfaces // MATEC Web of Conf. 2016. V. 72. P. 01045. <https://doi.org/10.1051/matecconf/20167201045>
68. Kirk S., Strobel M., Christopher S. L., Stuart J. A statistical comparison of contact angle measurement methods // J. Adhesion Sci. Technol. 2019. V. 33(16). P. 1758–1769. <https://doi.org/10.1080/01694243.2019.1611400>
69. Kocijan A., Conradi M., Hočevr M. The influence of surface wettability and topography on the bioactivity of TiO₂/Epoxy Coatings on AISI 316L // Stainless Steel Materials. 2019. V. 12. P. 1877. <https://doi.org/10.3390/ma12111877>
70. Krylach I.V., Fokina M.I., Kudryashov S.I. et al. Microfluidic water flow on laser-patterned MicroCoat®-coated steel surface // Appl. Surface Sci. 2021. V. 581. P. 152258. <https://doi.org/10.1016/j.apsusc.2021.152258>
71. Kulshreshtha Y., Vardon P.J., Du Y. et al. Biological stabilisers in earthen construction: a mechanistic Understanding of their response to water-ingress // The 4th International Conf. on Bio-Based Building Materials, June 16th–18th, 2021, Barcelona, Spain. Conference Paper. June, 2021. <https://doi.org/10.4028/www.scientific.net/CTA.1.529>
72. Lai H., Wu S., Cui M., Chu J. Recent development in biogeotechnology and its engineering applications // Front. Struct. Civ. Eng. 2021. V. 15(5). P. 1073–1096. <https://doi.org/10.1007/s11709-021-0758-0>
73. Law K.-Y., Zhao H. Surface wetting: Characterization, contact angle, and fundamentals. Switzerland: Springer Cham, 2016. 162 p. <https://doi.org/10.1007/978-3-319-25214-8>
74. Leelamanie D.A.L., Karube J., Yoshida A. Characterizing water repellency indices: Contact angle and water drop penetration time of hydrophobized sand // Soil Sci. Plant Nutrition. 2008. V. 54. P. 179–187. <https://doi.org/10.1111/j.1747-0765.2007.00232.x>
75. Leelamanie D.A.L., Karube J. Effects of hydrophobic and hydrophilic organic matter on the water repellency of model sandy soils // Soil Sci. Plant Nutrition. 2009. V. 55. P. 462–467. <https://doi.org/10.1111/j.1747-0765.2009.00388.x>
76. Lefebvre G., Galet L., Chamayou A. Dry coating of talc particles with fumed silica: Influence of the silica concentration on the wettability and dispersibility of the composite particles // Powder Technology. 2011. V. 208(2). P. 372–377. <https://doi.org/10.1016/j.powtec.2010.08.031>

77. *Li Y., Yu D., Wang X. et al.* Lauric arginate/cellulose nanocrystal nanorods-stabilized alkenyl succinic anhydride pickering emulsion: enhancement of stabilization and paper sizing performance // *Cellulose*. 2022. V. 29. P. 1–17.
<https://doi.org/10.1007/s10570-022-04502-z>
78. *McHale G., Newton M., Shirtcliffe N.* Water-repellent soil and its relationship to granularity, surface roughness and hydrophobicity: A materials science view // *Eur. J. Soil Sci.* 2005. V. 56. P. 445–452.
<https://doi.org/10.1111/j.1365-2389.2004.00683.x>
79. *McHale G., Newton M.* Liquid marbles: Principles and applications // *Soft Matter*. 2011. 7. P. 5473–5481.
<https://doi.org/10.1039/C1SM05066D>
80. *Melim L.A., Northup D.E., Boston P.J., Spilde M.N.* Preservation of fossil microbes and biofilm in cave pool carbonates and comparison to other microbial carbonate environments // *Palaeos*. 2016. V. 31. P. 177–189.
<http://dx.doi.org/10.2110/palo.2015.033>
81. *Meng J., Yang G., Liu L. et al.* Cell adhesive spectra along surface wettability gradient from superhydrophilicity to superhydrophobicity // *Sci. China Chem.* 2017. V. 60. P.
<https://doi.org/10.1007/s11426-016-9031-8>
82. *Mitik-Dineva N.* Bacterial attachment to micro- and nanostructured surfaces. Thesis ... Doctor of Philosophy. 2009. Swinburne University of Technology.
83. *Mundozah A.L., Tridon C.C., Cartwright J.J., Salman A D., Hounslow M.J.* Wetting of binary powder mixtures // *Int. J. Pharmaceutics*. 2019. V. 572. P. 118770.
<https://doi.org/10.1016/j.ijpharm.2019.118770>
84. *Nagy N.* Capillary Bridges on Hydrophobic Surfaces: Analytical contact angle determination // *Langmuir*. 2022. V. 38.
<https://doi.org/10.1021/acs.langmuir.2c00674>
85. *Nembrini S., König I. R., Wright M. N.* The revival of the Gini importance? // *Bioinformatics*. 2018. V. 34(21). P. 3711–3718.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6198850/>
86. *Nguyen H., Ybarra A., Başağaoğlu H. et al.* Biofilm viscoelasticity and nutrient source location control biofilm growth rate, migration rate, and morphology in shear flow // *Sci. Rep.* 2021. V. 11.
<https://doi.org/10.1038/s41598-021-95542-1>
87. *Ojeda G., Gi J.M., Mattana S. et al.* Biochar ageing effects on soil respiration, biochar wettability and gaseous CO₂ adsorption // *Mitig. Adapt. Strateg. Glob. Change* 2024. V. 29. P. 11.
<https://doi.org/10.1007/s11027-024-10107-7>
88. *Pommé L.E., Bourqui R., Giot R., Auber D.* Relative Confusion Matrix: An efficient visualization for the comparison of classification models // *Artificial Intelligence and visualization: advancing visual knowledge discovery. Studies in Computational Intelligence*. V. 1126. Springer, 2024. P. 223–243.
https://doi.org/10.1007/978-3-031-46549-9_7
89. *Pronk G.J., Heister K., Vogel C. et al.* Interaction of minerals, organic matter, and microorganisms during biogeochemical interface formation as shown by a series of artificial soil experiments // *Biol. Fertil. Soils*. 2017.
<https://doi.org/10.1007/s00374-016-1161-1>
90. *Regaieg M., Nono F., Faisal T.F. et al.* Large-Pore network simulations coupled with innovative wettability anchoring experiment to predict relative permeability of a mixed-wet rock // *Transp. Porous Med.* 2023. V. 147. P. 495–517.
<https://doi.org/10.1007/s11242-023-01921-9>
91. *Rohit S., Haider A., Raj A.* ANN-aided stiffness characterization of thin membranes using droplet motion // *Acta Mechanica*. 2023. V. 235. P. 1–18.
<https://doi.org/10.1007/s00707-023-03755-4>
92. *Ruiz-Cabello F.J., Rodríguez-Valverde M.A., Marmur A., Cabrerizo-Vilchez M.* Comparison of sessile drop and captive bubble methods on rough homogeneous surfaces: a numerical study // *Langmuir: ACS J. Surfaces Colloids*. 2011. V. 27. P. 9638–9643.
<https://doi.org/10.1021/la201248z>
93. *Semprebon C., McHale G., Kusumaatmaja H.* Apparent contact angle and contact angle hysteresis on liquid infused surfaces // *Soft Matter*. 2016. V. 13(1). P. 101–110.
<https://doi.org/10.1039/C6SM00920D>
94. *Shang J., Flury M., Harsh J., Zollars R.* Comparison of different methods to measure contact angles of soil colloids // *J. Colloid Interface Sci.* 2008. V. 328. P. 299–307.
<https://doi.org/10.1016/j.jcis.2008.09.039>
95. *Shang X., Luo Z., Gatapova E., Kabov O., Bai B.* GN-BC-based front-tracking method for the three-dimensional simulation of droplet motion on a solid surface // *Computers Fluids*. 2018. V. 172. P. 181–195.
<https://doi.org/10.1016/j.compfluid.2018.06.021>
96. *Shein E.V., Verkhovtseva N.V., Milanovsky E.Yu., Romancheva A.A.* Microbiological modification of kaolinite and montmorillonite surface: changes in physical and chemical parameters (model experiment) // *Biogeosystem. Technique*. 2016. V. 3(9). P. 229–234.
<https://doi.org/10.13187/bgt.2016.9.229>
97. *Sofinskaya O.A., Andrushkevich O.Y., Galiullin B.M. et al.* Surface Properties of Carbonate Speleothems in Karst Caves Changing Under Biofilms // *Biogenic–Abiogenic Interactions in Natural and Anthropogenic Systems 2022. Proceedings in Earth and Environmental Sciences*. Springer, 2023. P. 495–511.
https://doi.org/10.1007/978-3-031-40470-2_29
98. *Sofinskaya O.A., Mannapova L.M., Usmanov R.M. et al.* Biogeochemical interface development in a model carbonate-clayey soil // *Environ. Earth Sci.* 2024. V. 83. P. 6.
<https://link.springer.com/article/10.1007/s12665-023-11312-4>
99. *Spilde M.N., Boston P.J., Northup D.E., Odenbach K.J.* Rock coatings: potential biogenic indicators // *Ground Truth From Mars*. 2008. V. 1. P. 4045.

100. *Spilde M.N., Kooser A., Boston P.J., Northup D.E.* Speleosol: A Subterranean Soil // ICS Proceedings. Mineralogy. 2009. P. 338–344.
101. *Tarabal V.S., Abud Y.K.D., da Silva F.G. et al.* Effect of DMPEI coating against biofilm formation on PVC catheter surface // World J. Microbiol. Biotechnol. 2024. V. 40. P. 6. <https://doi.org/10.1007/s11274-023-03799-7>
102. *Tan P.N.* Receiver Operating Characteristic // Encyclopedia of Database Systems. Springer, 2009. https://doi.org/10.1007/978-0-387-39940-9_569
103. *Tariq Z., Ali M., Hassanpouryouzband A. et al.* Predicting wettability of mineral/CO₂/brine systems via data-driven machine learning modeling: Implications for carbon geo-sequestration // Chemosphere. 2023. V. 345. P. 140469. <https://doi.org/10.1016/j.chemosphere.2023.140469>
104. *Unkovich M., McBeath T., Llewellyn R. et al.* Challenges and opportunities for grain farming on sandy soils of semi-arid south and south-eastern Australia // Soil Res. 2020. V. 58(4). P. 323–334. <https://doi.org/10.1071/SR19161>
105. *Wagner D., Milodowski A. E., West J. M., Wragga J., Yoshikawa H.* Mineralogical comparisons of experimental results investigating the biological impacts on rock transport processes // Environ. Sci.: Processes Impacts. 2013. V. 15. P. 1501. <https://doi.org/10.1039/C3em00188a>
106. *Wang H., Orejon D., Song D. et al.* Non-wetting of condensation-induced droplets on smooth monolayer suspended graphene with contact angle approaching 180 degrees // Commun. Mater. 2022. V. 3. P. 75. <https://doi.org/10.1038/s43246-022-00294-8>
107. *Wang L., van Paassen L., Pham V., Mahabadi N., He J., Gao Y.* A (simplified) biogeochemical numerical model to predict saturation, porosity and permeability during Microbially Induced Desaturation and Precipitation // Water Res. Res. 2023. V. 59. P. <https://doi.org/10.1029/2022WR032907>
108. *Wang Z., Yang, Y., Xiang, W. et al.* Performance and mechanisms of greywater treatment in a bio-enhanced granular-activated carbon dynamic biofilm reactor // NPJ Clean Water. 2022. V. 5. P. 56. <https://doi.org/10.1038/s41545-022-00198-7>
109. *Weisbrod N., McGinnis T., Rockhold M.L., Niemet M.R., Selker J.S.* Effective Darcy-scale contact angles in porous media imbibing solutions of various surface tensions // Water Resour. Res. 2009. V. 45(4). P. <https://doi.org/10.1029/2008WR006957>
110. *Werb M., Falcón G.C., Bach N.C. et al.* Surface topology affects wetting behavior of *Bacillus subtilis* biofilms // NPJ Biofilms Microbiomes. 2017. V. 3. P. 11. <https://doi.org/10.1038/s41522-017-0018-1>
111. *Williams D., Kuhn A., Amann M. et al.* Computerized Measurement of Contact Angles 1 // Galvanotechnik. 2010. V. 101. P. 2502–2512.
112. *Woche S., Goebel M.-O., Kirkham M. et al.* Contact angle of soils as affected by depth, texture, and land management // Eur. J. Soil Sci. 2005. V. 56. P. 239–251. <https://doi.org/10.1111/j.1365-2389.2004.00664.x>
113. *Wróblewski P., Kachel S.* The concept of the contact angle in the process of oil film formation in internal combustion piston engines // Sci. Rep. 2023. V. 13. P. 20715. <https://doi.org/10.1038/s41598-023-47763-9>
114. *Wu J., Zhang M., Wang X., Li S., Wen W.* A Simple approach for local contact angle determination on a heterogeneous surface // Langmuir: ACS J. Surfaces Colloids. 2011. V. 27. P. 5705–5708. <https://doi.org/10.1021/la200697k>.
115. *Xu Z., Li Z., Liu Q.* Recent advances in studying colloidal interactions in mineral processing // Mining, Metallurgy Exploration. 2019. 36. P. 35–53. <https://doi.org/10.1007/s42461-018-0023-9>
116. *Yan J., Moreau A., Khodaparast S. et al.* Bacterial biofilm material properties enable removal and transfer by capillary peeling // Adv. Materials. 2018. V. 30. <https://doi.org/10.1002/adma.201804153>
117. *Yuan Y., Hays M., Hardwig P., Kim J.* Surface characteristics influencing bacterial adhesion to polymeric substrates // RSC Adv. 2017. V. 7. P. 14254–14261. <https://doi.org/10.1039/C7RA01571B>.
118. *Zeng Ch., Van Paassen L.A., Zheng J. et al.* Soil stabilization with microbially induced desaturation and precipitation (MIDP) by denitrification: a field study // Acta Geotechnica. 2022. V. 17. P. 5359–5374. <https://doi.org/10.1007/s11440-022-01721-3>
119. *Zhang B., Wang J., Liu Z. et al.* Beyond Cassie equation: Local structure of heterogeneous surfaces determines the contact angles of microdroplets // Sci. Rep. 2014. V. 4. P. 5822. <https://doi.org/10.1038/srep05822>
120. *Zhang L., Wang S., Wang T. et al.* Polishing mechanisms of various surfactants in chemical mechanical polishing relevant to cobalt interconnects // Int. J. Adv. Manuf. Technol. 2023. V. 128. P. 5425–5436. <https://doi.org/10.1007/s00170-023-12246-8>
121. *Zorina A.S., Maksimova Y.G., Demakov V.A.* Biofilm formation by monocultures and mixed cultures of *Alcaligenes faecalis* 2 and *Rhodococcus ruber* Gt 1 // Microbiology. 2019. V. 88(2). P. 164–171. <https://doi.org/10.1134/S0026261719020140>
122. *Zuo Y., Ding M., Bateni A., Hoorfar M., Neumann A.* Improvement of interfacial tension measurement using a captive bubble in conjunction with axisymmetric drop shape analysis (ADSA) // Aspects. 2004. V. 250. P. 233–246. <https://doi.org/10.1016/j.colsurfa.2004.04.081>

Application of Machine Learning Algorithms to Classify Soil Components with Different Hydrophilicity

O. A. Sofinskaya^{a, *}, F. A. Mouraviev^a, D. Rakonjac^{b, c}, and L. M. Mannapova^a

^a*Kazan (Volga region) Federal University, Kazan, 420008 Russia*

^b*Belarusian State University, Minsk, 220030 Belarus*

^c*Moscow Institute of Physics and Technology (National Research University),
Moscow Region, Dolgoprudny, 141701 Russia*

*e-mail: ushik2001@mail.ru

The capabilities of the pilot model classifier trained to recognize microbial activity traces on solid surfaces for studying soils and soil-like bodies have been preliminarily assessed. A database of 500 samples described by the authors and in open sources from 1988 to the present was collected for machine learning; among them, 59 samples represented soil horizons, 146 parent rocks and soil-like bodies, as well as rock-forming minerals, accompanying components of soil formation, xenobiotics common in technogenically transformed landscapes of the world. The samples were involved in the database as options of dispersion, coverage with biofilms and films of other nature, chemical and physical treatment. The array of sample features significant for machine learning included quantiles of the wetting contact angle distribution and generalizing categorical indicators of surface geometry, mineral composition, and state of organic matter. The classification target function was the presence of microbial activity stable traces on a solid surface. Missing data were reconstructed using Monte Carlo procedure and bootstrapping. As a result of numerical experiments on optimizing the machine learning a balanced training dataset containing 1233 pseudo-sample elements was obtained. Six classifier models with parameter variations were trained and evaluated. The most productive classifier, a five-layer neural network with randomly dropout neurons, demonstrated a prediction accuracy of 0.74 and an ROC AUC of 0.80 on the test sample, which is higher than that of simpler and faster classifiers (accuracy and ROC AUC of 0.70). Based on the disagreement between the classifications of a human expert and trained algorithm common feature of samples that are difficult for machine classification were established: with traces of life activity, carbonate, dispersed, which allows one to determine the direction of collecting information to improve the performance of the classifier. The development of an algorithm for recognizing traces of microbial activity is useful for clarifying the mechanisms of biogeochemical and biogeotechnological processes in soils of various origins, including soil formation and terraforming.

Keywords: biofilms, wetting contact angle, classification algorithms, neural network, selection of hyperparameters