

На правах рукописи



РУКАВИЦЫН ВАДИМ ВЯЧЕСЛАВОВИЧ

**ОПРЕДЕЛЕНИЕ УСТОЙЧИВОСТИ ГЕОЛОГИЧЕСКОЙ СРЕДЫ С
ПРИМЕНЕНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ (НА ПРИМЕРЕ
Г. МОСКВЫ)**

Специальность 25.00.08 – «Инженерная геология, мерзлотоведение и
грунтоведение»

Автореферат
диссертации на соискание ученой степени
кандидата геолого-минералогических наук

Москва 2018

Работа выполнена в Федеральном государственном бюджетом образовательном учреждении высшего образования «Российский государственный геологоразведочный университет имени Серго Орджоникидзе» МГРИ-РГГРУ на кафедре экологии и природопользования

Научный руководитель: **Экзарьян Владимир Нишанович** д.г-м.н., профессор, заведующий кафедрой Экологии и природопользования МГРИ-РГГРУ

Официальные оппоненты: **Лаврусевич Андрей Александрович** д.г-м.н., профессор, заведующий кафедрой инженерной геологии и геоэкологии МГСУ

Аверкина Татьяна Ивановна канд. г-м.н., доцент кафедры инженерной и экологической геологии геологического факультета МГУ им. М.В. Ломоносова

Ведущая организация: **Федеральное государственное бюджетное учреждение науки Институт геоэкологии им. Е.М. Сергеева Российской академии наук (ИГЭ РАН)**

Защита диссертации состоится 22 ноября 2018 г. в 15:00 час. На заседании диссертационного совета Д.212.121.01 при ФГБОУ ВО «Российский государственный геологоразведочный университет имени Серго Орджоникидзе» (МГРИ-РГГРУ) по адресу: 117997, г. Москва, ул. Миклухо-Маклая, д. 23, ауд. 4-73

С диссертацией можно ознакомиться в библиотеке Российского государственного геологоразведочного университета имени Серго Орджоникидзе.

Отзывы на автореферат в двух экземплярах, заверенные печатью организации, просьба отправить по адресу: 117997, г. Москва, ул. Миклухо-Маклая, д. 23 Российский государственный геологоразведочный университет имени Серго Орджоникидзе, ученому секретарю диссертационного совета Д.212.121.01.

Телефон: +7(495) 433-65-44.

Автореферат разослан _____ 2018 г.

Ученый секретарь диссертационного совета,
д.г-м.н., доцент



С.Д. Ганова

ОБЩАЯ ХАРАКТЕРИСТИКА

Актуальность работы

Вопрос определения устойчивости геологической среды является важным, актуальным и довольно сложным в решении. Моделирование устойчивости геологической среды чаще всего базируется на использовании различных модификаций экспертных методов, а применение математического моделирования крайне затруднено из-за сложности формализации исходных данных.

В России и за рубежом проводится широкий комплекс исследований по разработке и внедрению моделей машинного обучения для автоматизации систем принятия решений и других важных задач. Машинное обучение – это научная дисциплина, которая занимается проектированием и разработкой алгоритмов, позволяющих компьютерам изменить своё поведение на базе некой информации, например, на базе информационных сенсоров или базы данных [Mitchell T. 1997]. Одним из основных направлений исследований машинного обучения является способность учиться автоматически решать сложные проблемы и принимать разумные решения, основанные на содержательной формализации и обработке фактического материала

Для решения инженерно-геологических задач машинное обучение применяется не часто и, в основном, решаемые задачи довольно просты. Однако опыт использования данного метода в других отраслях показывает актуальность и перспективность этого направления для решения инженерно-геологических задач.

Заимствуя логику принятия решений, описанную в работах Бондарика Г.К., Осипова В.И., Трофимова В.Т., Пендина В.В., Шестопалова В.М., Азроянца Э.А, Зайонца И.О. и других исследователей, метод машинного обучения позволит формализовать информацию о геологической среде и оптимизировать определение оценки её устойчивости.

Цель работы

Разработка методики оптимизации решения задачи определения устойчивости геологической среды при помощи машинного обучения.

Задачи исследований

В соответствии с целевой установкой были поставлены следующие задачи научных исследований:

- Обоснование применения машинного обучения для оптимизации анализа устойчивости геологической среды
- Разработка методики определения устойчивости геологической среды при помощи машинного обучения;
- Выбор критериев, характеризующих устойчивость геологической среды;
- Создание представительной базы данных, характеризующей устойчивость геологической среды посредством выбранных критериев.

- Обработка исходных данных, для дальнейшей их обработки при помощи машинного обучения;
- Определение устойчивости геологической среды г.Москвы при помощи машинного обучения и сравнение результатов с данными многолетних исследований анализа;
- Интерпретация результатов и обоснование применимости машинного обучения для определения устойчивости геологической среды на предпроектной стадии строительства.

Фактический материал, методы исследования

Основным объектом исследования устойчивости геологической среды послужили результаты инженерных изысканий г. Москвы. В их основе лежали карты, которые содержали в себе как некую базовую информацию, так и результаты экспертной оценки территории.

В ходе выполнения работы автором был проведен анализ литературных источников, описывающих возможности применения машинного обучения для решения инженерно-геологических задач и способы определения устойчивости геологической среды. В результате этого анализа была разработана методика определения устойчивости геологической среды при помощи машинного обучения с целью оптимизации выполнения данного вида исследований для получения возможности его более широкого применения.

На основе разработанной методики были собраны исходные данные, которые впоследствии были обработаны, специальным образом проанализированы и адаптированы для дальнейшего использования в машинном обучении. Данные включали в себя как параметры геологической среды, так и результаты экспертного моделирования ее устойчивости.

С помощью программного продукта WEKA была построена компьютерная модель устойчивости геологической среды г. Москвы, копирующая результаты ранее проведенного экспертного анализа, что позволяет частично доверить проведение такой сложной комплексной оценки искусственному интеллекту.

Разработка методики, создание базы исходных данных и моделирование проводилось в Российском Государственном Геологоразведочном Университете им. Серго Орджоникидзе на кафедре экологии и природопользования и в Университете Менделя г. Брно.

Научная новизна работы

1. Разработана уникальная методика определения устойчивости геологической среды на основе методов машинного обучения;
2. Предложенная методика определения устойчивости геологической среды на основе методов машинного обучения была апробирована при решении реальной инженерно-геологической задачи;
3. На основе ряда экспериментов были определены наиболее значимые параметры, характеризующие устойчивость геологической среды;

4. В результате внедрения новой методики создана представительная исходная база данных, адаптированная для работы с машинным обучением и позволяющая решать задачи по определению устойчивости геологической среды в городах центральной России.

5. На основе представленных экспериментальных данных определен программный алгоритм, способный наиболее точно проводить оценку устойчивости геологической среды;

6. При помощи проведенной оценки показана возможность применения машинного обучения для определения устойчивости геологической среды, что позволяет унифицировать решение подобных задач в будущем.

Защищаемые научные положения

1. Определение устойчивости геологической среды, выполненное на предпроектной стадии, позволяет оптимизировать и унифицировать проведение инженерных изысканий.

2. Одним из перспективных методов математического моделирования и прогнозирования устойчивости геологической среды является машинное обучение, позволяющий обрабатывать большие объемы разнородной и сложной по составу инженерно-геологической информации.

3. Модель устойчивости геологической среды г. Москвы, разработанная в данной работе, следует рассматривать в качестве аналога при определении состояния геологической среды урбанизированных территорий, находящихся в аналогичных инженерно-геологических условиях.

Практическая значимость работы

1. Проведенный анализ различных способов применения искусственного интеллекта для решения инженерно-геологических задач и подходов к анализу геологической среды позволил сформировать подход к определению ее устойчивости на основе машинного обучения.

2. Усовершенствованы существующие практические подходы к определению устойчивости геологической среды для целей оценки территории строительства.

3. Созданная представительная исходная база данных, характеризующая устойчивость геологической среды г. Москвы позволяет проводить ее оценку на сходных по условиям территориях в кратчайшие сроки и при наличии минимального набора информации.

4. Разработанная методика использования машинного обучения доказала свою состоятельность при решении реальных задач и результаты этих решений коррелировались с данными многолетних изысканий, что имеет важное значение для прогноза негативных инженерно-геологических процессов при строительстве.

Апробация работы

Основные результаты работы и её отдельные положения докладывались на научных конференциях:

Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2018», Москва 2018; Новые идеи в науках о Земле, Москва 2017; Науки о Земле на современном этапе, Москва 2014; 9th Summer School of Applied Informatics, Бедрихов, Чешская Республика, 2012; Развитие экологической экономики и образования в мире и в Украине, Киев, Украина, 2011; MECON 2011, Острава, Чешская Республика, 2010; ISDMCI 2010, Евпатория, Украина, 2010; MendelNet 2009, Брно, Чешская Республика, 2009. В 2010 году автор получил грант IGA Университета Менделя.

Публикации

По теме диссертационной работы опубликовано 16 печатных работ, в том числе 2 в рецензируемых монографиях и 4 в реферируемых журналах ВАК.

Объем и структура работы

Диссертация состоит из введения, 6 глав, заключения и списка использованных источников из 114 наименований. Работа изложена на 187 страницах, содержит 29 рисунков, 13 таблиц.

Благодарности

Автор выражает искреннюю и глубокую благодарность своему научному руководителю – доктору геолого-минералогических наук, профессору Экзарьяну Владимиру Нишановичу.

Автор признателен всем сотрудникам кафедры экологии и природопользования за помощь и ценные советы при написании настоящей работы.

Автор пользуется случаем поблагодарить всех сотрудников Университета Менделя г. Брно за внимание, помощь и ценные советы в работе. Автор считает своим долгом выразить благодарность доценту Яну Жижка за содействие в изучении машинного обучения и методов его прикладного применения в инженерной геологии.

СОДЕРЖАНИЕ РАБОТЫ

В данной работе рассмотрен способ определения устойчивости геологической среды при помощи машинного обучения.

Под устойчивостью геологической среды, согласно формулировкам А.И. Ляпунова, У.Р. Эшби, Л.А. Чульского, Фроловой В. И., Татарина В.Н. понимается:

- способность сохранять некоторые характеристики;
- способность к изменениям, но в пределах некоторых толерантных пределов;
- способность к возвращению в базовое состояние, несмотря на воздействия.

В целом методологию решения поставленной задачи можно разделить на две части: обучение и тестирование. Обучение необходимо для создания базовой модели устойчивости геологической среды. Тестирование производится для проверки корректности модели, её корректировки и оценки возможности решения реальных задач. Таким образом, была получена рабочая система оптимизации процесса определения уровня устойчивости геологической среды. Анализ полученных результатов позволяет оптимизировать и унифицировать проведение инженерных изысканий предпроектной стадии.

Для определения точности моделирования была подобрана информация по городу Москве. Для оценки точности прогнозирования была использованы результаты многолетних ученых исследований, обобщенные в монографии «Москва: Геология и город» под редакцией Осипова В.И., Медведевой О.П.. В результате сравнивалась предсказанная устойчивость геологической среды с данными, полученными группой специалистов, на основе чего и определялась точность моделирования.

Оценка показала достаточно хорошую корреляционную связь между выбранными критериями и уровнем устойчивости геологической среды.

ЗАЩИЩАЕМЫЕ ПОЛОЖЕНИЯ И ИХ ОБОСНОВАНИЯ

Определение устойчивости геологической среды, выполненное на предпроектной стадии, позволяет оптимизировать и унифицировать проведение инженерных изысканий

При выборе участка для строительства нужно учитывать не только планировочные качества, не только удаленность или близость к уже имеющимся населенным пунктам, коммуникациям и дорогам, но и возможные последствия вмешательства в природную среду на конкретном участке. Природная, в том числе и геологическая, среда по-разному реагирует на техногенное воздействие. Эта реакция зависит как от интенсивности воздействия, так и от свойств самой среды.

Термин устойчивость широко используется в различных областях знаний, но до настоящего времени он является термином общего пользования. Так, например, Татаринов В.Н. определял её, как способность не изменяться под действием внешней нагрузки [Татаринов В.Н. 2000].

То же относится и к устойчивости геологической среды. Однако, в инженерной геологии понятие устойчивости начало употребляться в отношении оценки оползневых процессов, и этот фактор являлся основополагающим при ее определении. Однако при появлении понятия «геологическая среда» термин устойчивость приобрел более общее значение и стал применяться при прогнозировании и других процессов.

Так Бондарик Г.К. [Бондарик Г.К. 1984] употребляет термин «устойчивость» в контексте способности системы подавлять внешние импульсы и длительное время сохранять свои функции.

Такие импульсы воздействия или факторы, определяющие устойчивость геологической среды, разные авторы определяли по своему. Так, например, Инамов А.Н. [Инамов А.Н. 1989] в качестве ведущих факторов устойчивости инженерно-геологических условий выделял: геологическое строение, характер и степень расчлененности рельефа, тип фильтрационного разреза, интенсивность инженерно-хозяйственного воздействия, дренированность территории, уровень минерализации грунтовых вод, просадочность и засоленность грунтов. Это он использовал для составления карты устойчивости геологической среды.

Им же с соавтором была составлена Карта изменения геологической среды под воздействием ирригационно-мелиоративного строительства в центральной части Каршинской степи (Рис.1). Оценка изменений проводилась путем наложения однотипных карт. Эти аналитические карты позволили оценить произошедшие за 20 лет изменения глубин залегания грунтовых вод, их минерализации, засоленности грунтов зоны аэрации, региональные изменения инженерно-геологических свойств грунтов. Это дало возможность качественно и количественно отобразить изменения гидрогеологических и инженерно-геологических компонентов геологической среды, а также оценить интенсивность и направленность ее изменений в целом.

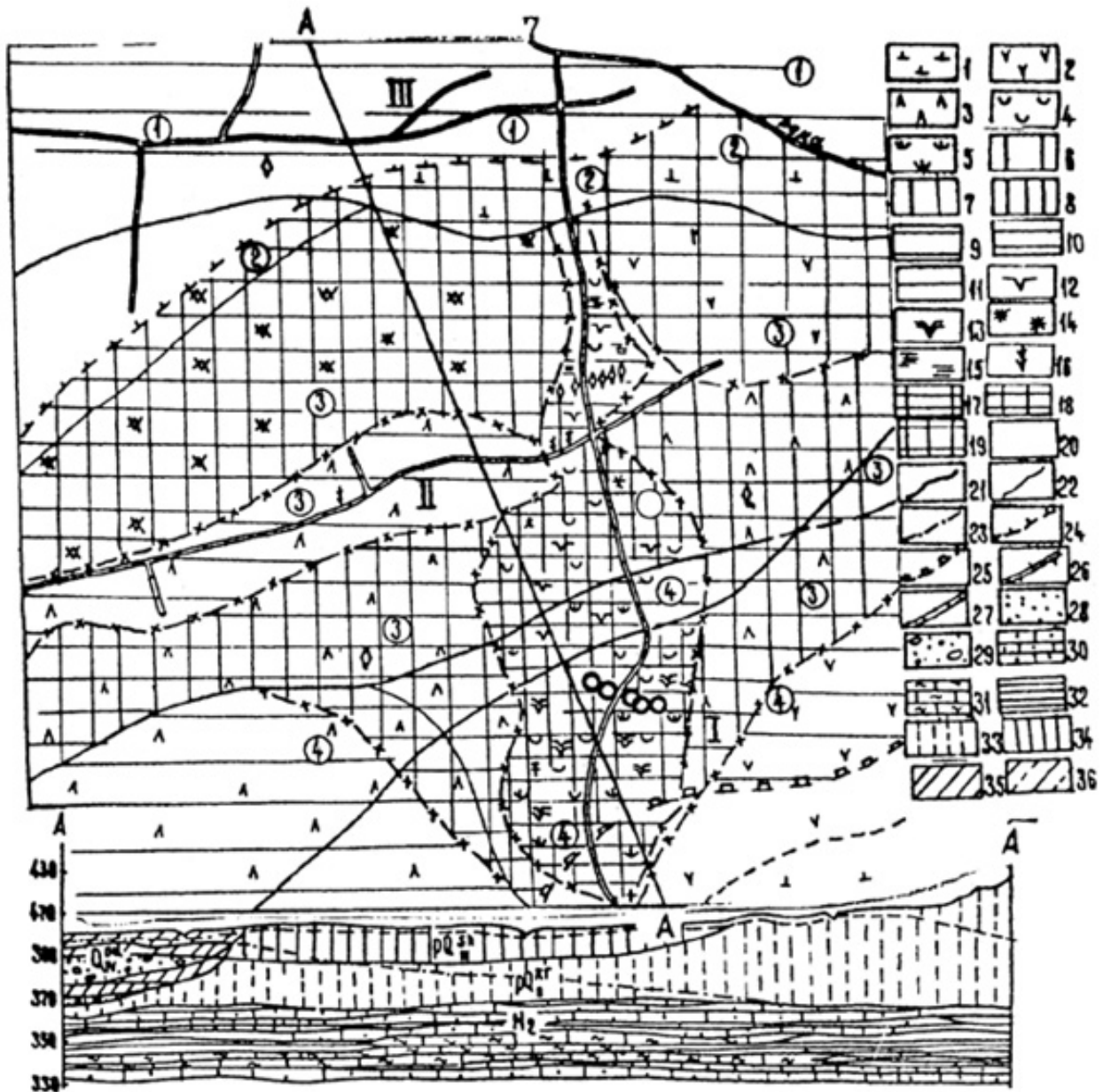


Рисунок 1. Макет Карты изменения геологической среды под воздействием ирригационно-мелиоративного строительства в центральной части Каршинской степи. Геолого-геоморфологическое строение и гидрогеологические условия. Равнины: I — пролювиальная волнистая сильнорасчлененная рШкр (лесс, 20–40 м); II — пролювиальная плоская, рШш (лессовидные суглинки и супеси, 5–15 м); III — аллювиальная плоская аVад (переслаивающиеся суглинки, супеси и пески, 20–30 м). Подъем уровня грунтовых вод, м: 1 — 2–5; 2 — 5–10; 3 — 10–20; 4 — 20–30; 5 — 30–40. Увеличение содержания легкорастворимых солей в грунтах, %: 6 — 0,1–0,3; 7 — 0,3–0,6; 8 — более 0,6; естественной влажности, %: 9 — 1,5; 10 — 5–10; 11 — более 10. Инженерно-геологические процессы и явления, просадка: 12 — до 15 см, 13 — более 15 см; 14 — засоление; 15 — подтопление и заболачивание; 16 — размыв и обрушение берегов каналов, коллекторов и дрен. Степень измененности геологической среды: 17 — сильная; 18 — средняя; 19 — слабая; 20 — неизменная. Границы: 21 — морфогенетических типов рельефа; участков с различной: 22 — глубиной залегания УГВ, 23 —

степенью измененности геологической среды; 24 — староорошаемой зоны; 25 — новоорошаемой; 26 — каналы; 27 — коллекторы; 28 — песок; 29 — гравий, галька; 30 — песчаник; 31 — алевролит; 32 — глина; 33 — лессовидные суглинки и супеси; 34 — лесс; 35 — суглинки плотные; 36 — супеси плотные. [Э. В. Мавлянов, А. Н. Инамов, 1986]

Методы оценки устойчивости геологической среды отличались у разных авторов. Так Котловым В.Ф. и Коффом Г.Л. [Котлов В.Ф. 1997] была составлена серия аналитических оценочно-прогнозных детальнх карт масштаба 1:10 000 опасности воздействия геологических процессов для разных стадий проектирования городской застройки, наземной и подземной инфраструктуры города, реконструкции и капитального ремонта. Опасность воздействия геологических процессов на здания и сооружения оценивалась с точки зрения устойчивости самих зданий. Модель взаимодействия геологической среды и сооружений для г. Москвы выражена авторами на рисунке 2.



Рисунок 2. Граф трехмерной модели геологической среды. Базовые элементы модели, отражающие развитие и протекание геологических процессов: 1 — рельеф; 2 — состояние грунтовой толщи; 3 — уровень подземных вод; 4 — высотное положение кровли каменноугольных отложений; 5 — ее рельефность; 6 — литологический тип пород, залегающих в кровле каменноугольных толщ; 7 — наличие или отсутствие регионального юрского водоупора; 8 — наличие или отсутствие разуплотненных зон [Кофф Г. Л. 1999]

Устойчивость зданий оценивалась с позиций рассмотрения протекания различных процессов: 1) на границе раздела «сооружение — геологическая среда»; 2) в области взаимодействия сооружения с толщей грунтов; 3) в пределах геологической среды, окружающей данную область. Соответственно процессы были подразделены на прямые, воздействующие непосредственно на сооружение (позиции 1, 2, рис. 2), и косвенные, действующие опосредованно через область взаимодействия (позиция 3) [Кофф Г. Л. 1999].

В этой модели в качестве критериев, определяющих состояние геологической среды, были использованы рельеф, толщи пород и подземные воды. По степени воздействия эти критерии были разделены на два уровня:

- на первом уровне на здания воздействуют прямые процессы, которые определяются: 1) рельефом, 2) состоянием грунтовых толщ в области взаимодействия с сооружением и 3) близостью к дневной поверхности (а значит и к фундаменту) первого от поверхности водоносного горизонта;

- на втором уровне действуют косвенные процессы, которые определяются, применительно к рассматриваемой территории: 1) характером поверхности кровли каменноугольных отложений; 2) распространением регионального юрского водоупора, отделяющего эти отложения от вышележащих четвертичных, относящихся к первому уровню геологической среды; 3) наличием зон разуплотнения, контролирующими тектонические нарушения, а также линий метрополитена [Кофф Г. Л. 1999].

Оценка опасности на 1-м модельном уровне, при учете особенностей рельефа, типа грунтовой толщи и уровня грунтовых вод, показывает, по мнению авторов карт, опасность воздействия геологических процессов, которую необходимо учитывать на стадиях проекта детальной планировки и проекта застройки.

Петренко А.С. совместно с этими авторами на основе указанных выше выводов были составлены карты геологического риска г. Москвы и опасности воздействия геологических процессов на объекты строительства [Петренко А.С. 2001]. Он также при оценке риска и опасности воздействия оценивались не только геологическая устойчивость территории, но и типы зданий, их фундаменты и другие конструктивные особенности объектов. Пример таких карт представлен на рисунках 3-4.

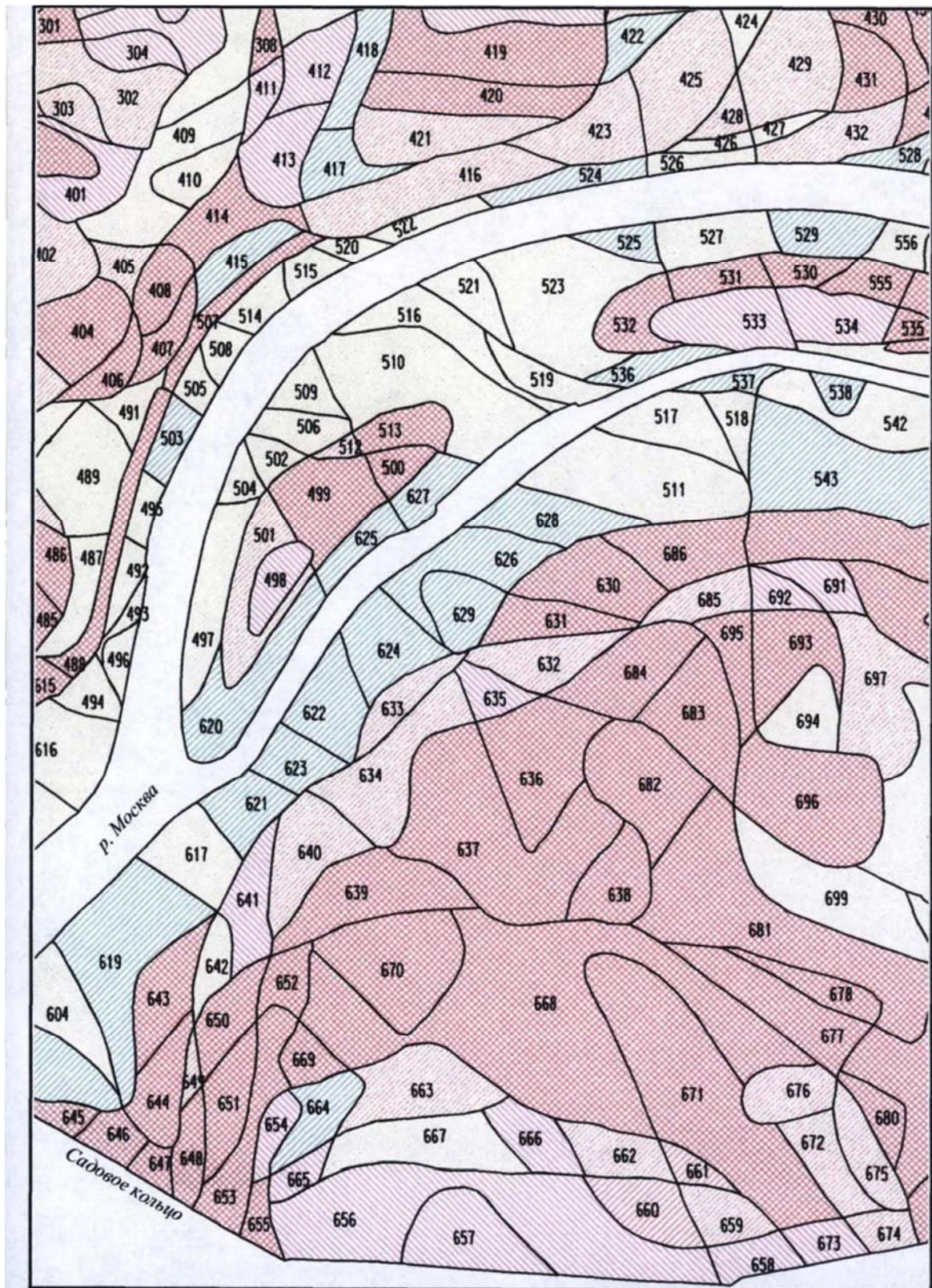


Рисунок 3. Фрагмент карты опасности воздействия геологических процессов для проектирования наземной и подземной инфраструктуры города (г. Москва, центральная часть) [Петренко А.С. 2001]



Рисунок 4. Условные обозначения к картам опасности воздействия геологических процессов на территории г. Москвы [Петренко А.С. 2001]

Академиком Осиповым В.И. для оценки устойчивости геологической среды города Москвы применялся учет трех важнейших опасных геологических процессов, развивающихся на территории города: карстово-суффозионного, подтопления и образования оползней. В других своих работах академик Осипов В.И. предлагает учитывать повышенную сжимаемость техногенных грунтов, загрязнение геологической среды промышленными и бытовыми отходами, техногенные физические поля, такие как вибрационное, электрическое и тепловое [Осипов В.И. 1994].

Всё это показывает, что проблема определения устойчивости геологической среды стояла и стоит довольно остро, а получаемая при этом информация очень востребована. Однако для этого зачастую необходимо провести огромный объем изысканий и потратить на исследования не один год.

Инженерные изыскания подразделяются на три стадии:

1. Предпроектная (Технико экономическое обоснование)
2. Проектная;
3. Рабочая документация.

На начальной стадии используются сравнительно-геологические методы прогноза, дополняемые отдельными поверочными расчётами, а также

вероятностно-статистические методы. На последующих стадиях, когда выполняется детальное обоснование схем размещения отдельных сооружений и возрастает необходимость в количественных моделях, применяются в большом объеме методы моделирования [Трофимов В.Т. 2015].

Определение уровня устойчивости геологической среды зачастую проводится на стадии проекта и рабочей документации из-за сложности и долговременности исследований для получения необходимой информации. Однако наибольшую полезность эта информация имеет именно на предпроектной стадии, поскольку, зная уровень устойчивости геологической среды, можно более достоверно обосновать место расположения объекта строительства, оптимизировать объем изысканий, сконцентрировав основные усилия в местах повышенного риска, получить необходимую информацию для принятия проектных решений. В идеальном варианте устойчивость геологической среды должна определяться ещё на предпроектной стадии и использоваться при составлении программы инженерных изысканий. Для решения этой задачи было применено машинное обучение.

Одним из перспективных методов математического моделирования и прогнозирования устойчивости геологической среды является машинное обучение, позволяющий обрабатывать большие объемы разнородной и сложной по составу инженерно-геологической информации.

Определение устойчивости геологической среды во многом сводится к моделированию и прогнозированию, в данном случае произведенным при помощи методов машинного обучения. Г.К. Бондарик понимал пол инженерно-геологическим прогнозированием предсказание инженерно-геологических условий [Бондарик Г.К. 1986]. Классификация инженерно-геологических прогнозов представлена на рисунке 5

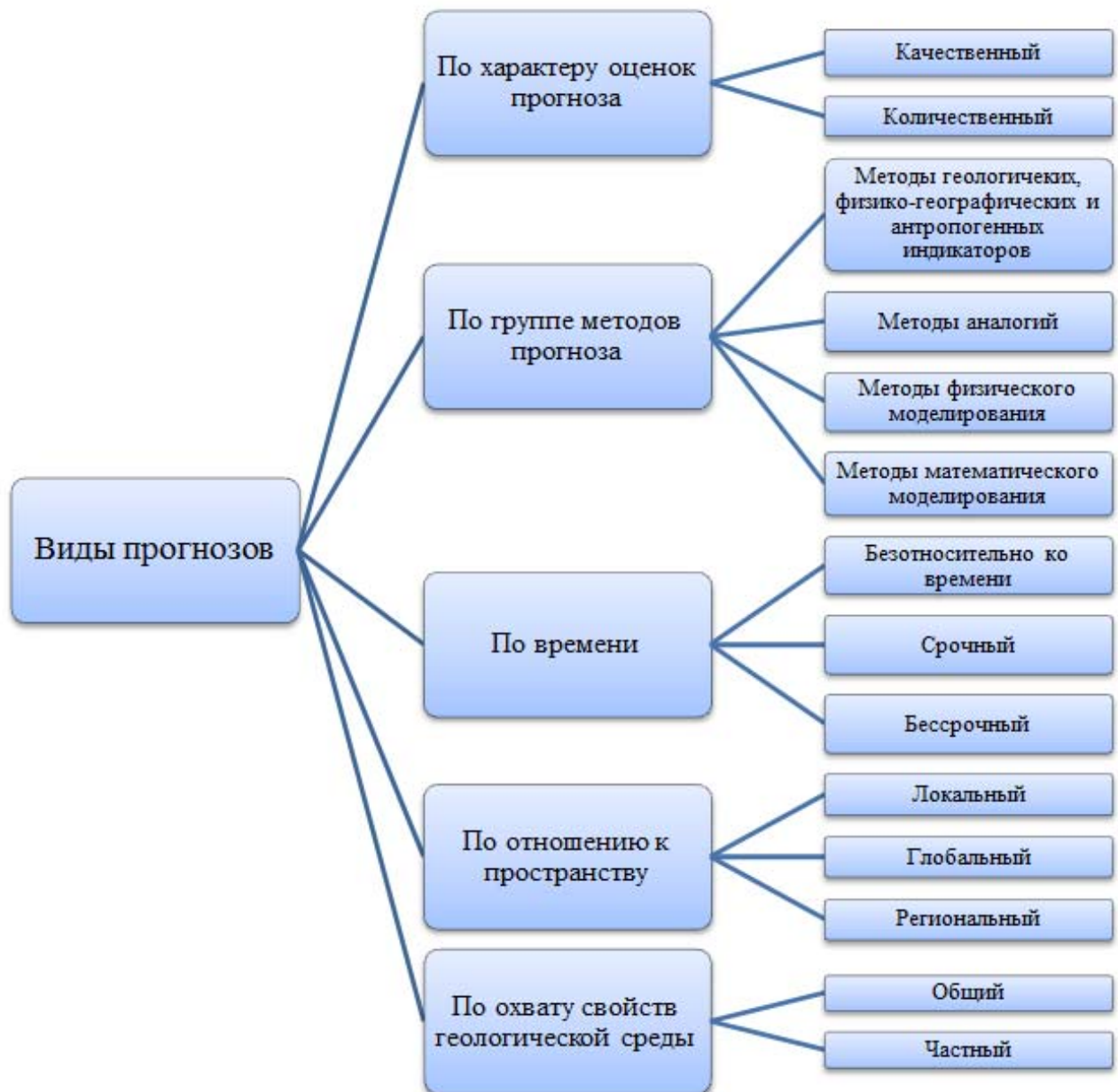


Рисунок 5. Классификация инженерно-геологических прогнозов [Бондарик Г.К. 1986].

Выбор методов инженерно-геологического прогнозирования связан со стадийностью инженерно-геологических исследований.

В.Т. Трофимов в своей работе писал, что на начальной стадии в большей степени используются методы аналогий и индикаторов. Тогда как на последующих стадиях, когда выполняется детальное обоснование схем размещения отдельных сооружений и возрастает необходимость в количественных моделях, применяются в большом объеме методы математического и физического моделирования [Трофимов В.Т. 2015].

Так как для определения устойчивости геологической среды необходимо обрабатывать большие объемы разнородной и сложной по составу инженерно-геологической информации, а в соответствии с задачей исследования это необходимо делать на предпроектной стадии, то необходимо совместить методы индикаторов и аналогий с методами математического моделирования.

Здесь на помощь приходит интеллектуальный анализ данных.

Интеллектуальный анализ данных — собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности [Паклин Н.Б. 2009].

Под данными при этом понимается фактическая информация в формализованном виде, которую можно передавать и обрабатывать в некотором информационном процессе, т.е. в процессе создания, поиска, получения, сбора, обработки, хранения, накопления, распространения и использования информации.

Информация же при этом представляет собой сведения о каком-либо явлении или процессе. Информация является базовой единицей измерения при анализе данных и присутствует как в самих данных, так и после их обработки для получения знаний, т.е. проверенной информации, позволяющей решать конкретную задачу.

Интеллектуальный анализ данных объединяет различные методы типизации, моделирования, основанные на применении различных способах получения знаний из набора данных.

Также при интеллектуальном анализе важно наглядно представлять результаты вычислений, чтобы даже не специалисты могли пользоваться результатами его обработки.

Изначально интеллектуальный анализ используется при наличии достаточно большой базы данных, содержащей сложно структурированную, разноплановую и часто «скрытую» информацию. Извлечение подобной информации из большого объема данных является сложной задачей. Поэтому необходимо найти или разработать методы получения знаний, которые скрыты в большом количестве исходных данных.

Под скрытыми знаниями понимаются знания:

- неизвестные ранее — т.е. знания, являющимися новыми, а не подтверждающими ранее полученную информацию;

- нетривиальные – т.е. не очевидные при просматривании или при проведении простой статистической обработки;

- полезные – т.е. представляющие собой ценность для ученого или потребителя;

- легко интерпретируемые – т.е. знания, которые легко привести в наглядную для пользователя форму и которые можно легко объяснить в рамках исследуемой задачи.

Все эти требования удовлетворяют задаче определения устойчивости геологической среды.

На данный момент собрано огромное количество данных инженерно-геологических исследований в различных областях, что идеально подходит для компьютерного интеллектуального анализа. При этом машина (компьютер) может помочь специалистам разобраться в постановке задачи и в выборе методов ее решения. Одним из способов применения компьютерного интеллектуального анализа данных является машинное обучение.

Машинное обучение в своей основе представляет собой анализ влияния группы факторов на некий признак. В нашем случае это анализ влияния взаимосвязанных между собой природных и техногенных факторов на устойчивость геологической среды. По сути это смесь многофакторного анализа, метода аналогии и статистики. В инженерной геологии многофакторный анализ был описан Бондариком Г.К. в его теории геологического поля, Пендиным В.В. в комплексном количественном анализе информации в инженерной геологии и многими другими исследователями.

Так как в качестве примера применения машинного обучения для определения уровня устойчивости геологической среды был выбран город Москва, исходные данные были взяты именно по этому городу. Для этого были использованы результаты исследований группы ученых, опубликованных под редакцией Осипова В.И., Медведевой О.П. [Осипов В.И. 1997]. Как я уже упоминал, эти авторы в качестве факторов, влияющих на устойчивость геологической среды, выделяли три важнейших опасных геологических процесса, развивающихся на территории города: карстово-суффозионный, подтопление, образование оползней. Однако, для получения информации по этим процессам исследователями предварительно были обработаны данные по геологическому строению, гидрогеологии, тектонике, гидрологии, ландшафтно-геоморфологическим условиям и др. Т.е. следует учитывать, что данные по опасным геологическим процессам сами по себе уже являлись результатом обработки огромного числа исходных параметров, описывающих инженерно-геологические условия, т.е. были результатом промежуточного моделирования. Такие данные не могут быть получены на первоначальном этапе. А, как указывалось выше, с точки зрения полезности оценки устойчивости геологической среды, намного лучше было бы ее получить в самом начале исследований.

Поэтому для оценки при помощи машинного обучения в качестве этих факторов использовались следующие:

1. Рельеф;
2. Разломы;
3. Линиаментные зоны;
4. Уровень грунтовых вод;
5. Гидрологическая сеть;
6. Мощность техногенных отложений.

Именно эти параметры непосредственно связаны с устойчивостью, позволяют описать каждый из проанализированных специалистами опасных геологических процессов и становятся доступны на первом этапе изысканий. Часто они уже имеются в качестве архивной информации ещё на предпроектной стадии, что позволит оценить устойчивость геологической среды в самом начале работ. Среди указанных факторов геология и гидрогеология представлена довольно слабо (техногенные отложения и УГВ). Это связано с тем, что более сложная геология и гидрогеология сильно отличается в различных местах. В случае же создания модели при помощи машинного обучения для решения задач определения устойчивости геологической среды на предпроектной стадии необходимо, чтобы базовые факторы были максимально универсальны, и возможно было применять данную модель в максимально большом количестве ситуации с минимальными изменениями ее параметров. Поэтому компьютеру была поставлена задача определения устойчивости на базе вышеперечисленных факторов, как наиболее универсальных, полно отражающих геологическую картину и получаемых на самых ранних стадиях работ.

В данной работе для этого анализа использовалось так называемое «обучение с учителем». Оно заключается в использовании результатов проведенных специалистами исследований (в данном случае группой во главе с Осиповым В.И.) так, чтобы на их основе «научить» машину, копировать логику принятия решений при определении уровня устойчивости геологической среды г. Москвы. После чего, когда возникнет задача сделать подобный анализ на новой неизвестной, но со сходными инженерно-геологическими условиями территории, машина, отталкиваясь от полученных знаний, могла бы провести подобный анализ уже самостоятельно без участия специалистов.

Схема машинного обучения и место в тестирования в ней изображено на рисунке 6



Рисунок 6. Схема машинного обучения при определении устойчивости геологической среды

Исходные данные по определенным факторам (рельеф, разломы, линиаменты, УГВ, гидрографическая сеть, мощность техногенных отложений), представленные, главным образом, в виде карт по г. Москва оцифровываются и представляются в виде таблиц.

В таблице, описывающей каждый фактор, содержалась информация о координатах каждой условной точки на оцифрованной карте и значении параметра в этой точке. В случае рельефа, например, в таблице приводились данные по абсолютной высотной отметке в каждой точке, а в случае разломов указывалось просто его наличие или отсутствие в указанной точке. В таких условиях компьютер мог более менее спокойно обработать эту информацию.

Затем было необходимо эти данные обработать. При обработке указанные таблички сливались в одну таким образом, чтобы в координаты 1ой точки первой таблички совпадали с координатами 1ой точки второй таблички и т.д. При этом получалась единая матрица значений, содержащая сведения сразу обо всех значимых факторах в каждой условной точке. Далее эта матрица была

проверена на наличие шумов, отсутствующих значений и других дефектов данных для более быстрой и качественной обработки

После этого идентичным образом оцифровывалась карта устойчивости геологической среды г. Москвы, являющаяся результатом многолетней работы группы специалистов.

В итоге получилась такая таблица, содержащая координаты и значения категории устойчивости геологической среды в каждой условной точке, совпадающей с точками уже описанной матрицы. После чего эти данные так же добавлялись в матрицу, завершая ее. Матрица описывала 40 000 точек.

Благодаря этим действиям была получена так называемая «обучающая выборка», т.е. выборка, содержащая в себе и значения факторов, характеризующих устойчивость геологической среды, и сами значения устойчивости в каждой условной точке исследуемой территории. В случае, если бы выборка содержала только значения факторов, а устойчивость геологической среды была бы нам неизвестна, то такая выборка была бы тестовой. Стоит отметить, что для обучающей выборки были использованы данные по территориям, находящимся исключительно внутри МКАД.

Далее проводилось уже непосредственно машинное обучение. Для этого полученная обучающая выборка была загружена в специализированный программный продукт WEKA, содержащий в себе сотни алгоритмов машинного обучения. Задача этих алгоритмов – тем или иным способом найти взаимосвязь между значениями факторов и устойчивости геологической среды в каждой точке выборки. Так, идя от точки к точке, алгоритм ищет новые взаимосвязи между ними, самостоятельно усложняется, корректирует себя и идёт дальше. В итоге, чем больше значений в обучающей выборке, тем лучше алгоритм сумеет приспособиться к решению поставленной задачи. Для большей эффективности обучения и проверки его точности использовался так называемый метод 10-тикратной перекрестной проверки.

При этом, данные разбиваются на 10 частей. После чего на 9 из них производится описанное выше обучение, а оставшаяся часть данных используется для тестирования (т.е. значения устойчивости геологической среды как бы закрываются от компьютера, и компьютер должен уже самостоятельно вслепую их определить, после этого предсказанные компьютером значения сравниваются с реальными).

Процедура повторяется 10 раз; в итоге каждая из 10 частей данных поочередно используется для тестирования. В результате получается оценка эффективности выбранной модели с наиболее равномерным использованием имеющихся данных. Поиск наиболее подходящего для решения поставленной задачи алгоритма и его параметров осуществляется экспериментально.

Когда наилучший алгоритм найден, проводится корректировка модели, заключающаяся в возможном изменении количества исходных параметров по причине того, что машина, например, не может увязать один из них с устойчивостью геологической среды, что усложняет ей расчет, или дополнительной обработке данных, после чего обучение начинается заново.

Так происходит до тех пор, пока не будут найдены наилучшие параметры выборки и алгоритма ее обработки.

В итоге всех описанных манипуляций была получена база данных, при помощи которой компьютер обучился определять уровень устойчивости геологической среды, копируя логику группы специалистов, с точностью более 75%, что соответствует принятому Александром Борисовичем Лисенковым [Лисенков А.Б. 2003] в своей работе допустимому порогу ошибки при пространственном прогнозировании.

Таким образом, машинное обучение позволяет обрабатывать большие объемы разнородной и сложной по составу инженерно-геологической информации, после чего проводить математическое моделирование и прогнозирование устойчивости геологической среды, что дает право использовать его и в дальнейшем для решения подобных задач.

Модель устойчивости геологической среды г. Москвы, разработанная в данной работе, следует рассматривать в качестве аналога при определении состояния геологической среды урбанизированных территорий, находящихся в аналогичных инженерно-геологических условиях.

Описанная методика применялась для определения уровня устойчивости геологической среды г. Москвы и сравнения его результатов с результатами проводимой ранее экспертной оценки.

Для этого в качестве обучающей выборки при создании модели были использованы данные по г. Москве, за исключением двух районов: Куркино Можаниновский.

Для дополнительной проверки точности модели была создана отдельная тестовая выборка, содержащая в себе значения выбранных факторов по территориям районов Куркино и Можаниновскому, расположенных на северо-западе Москвы за МКАД, в то время как компьютер обучался определять устойчивость геологической среды исключительно на точках, находящихся внутри МКАД.

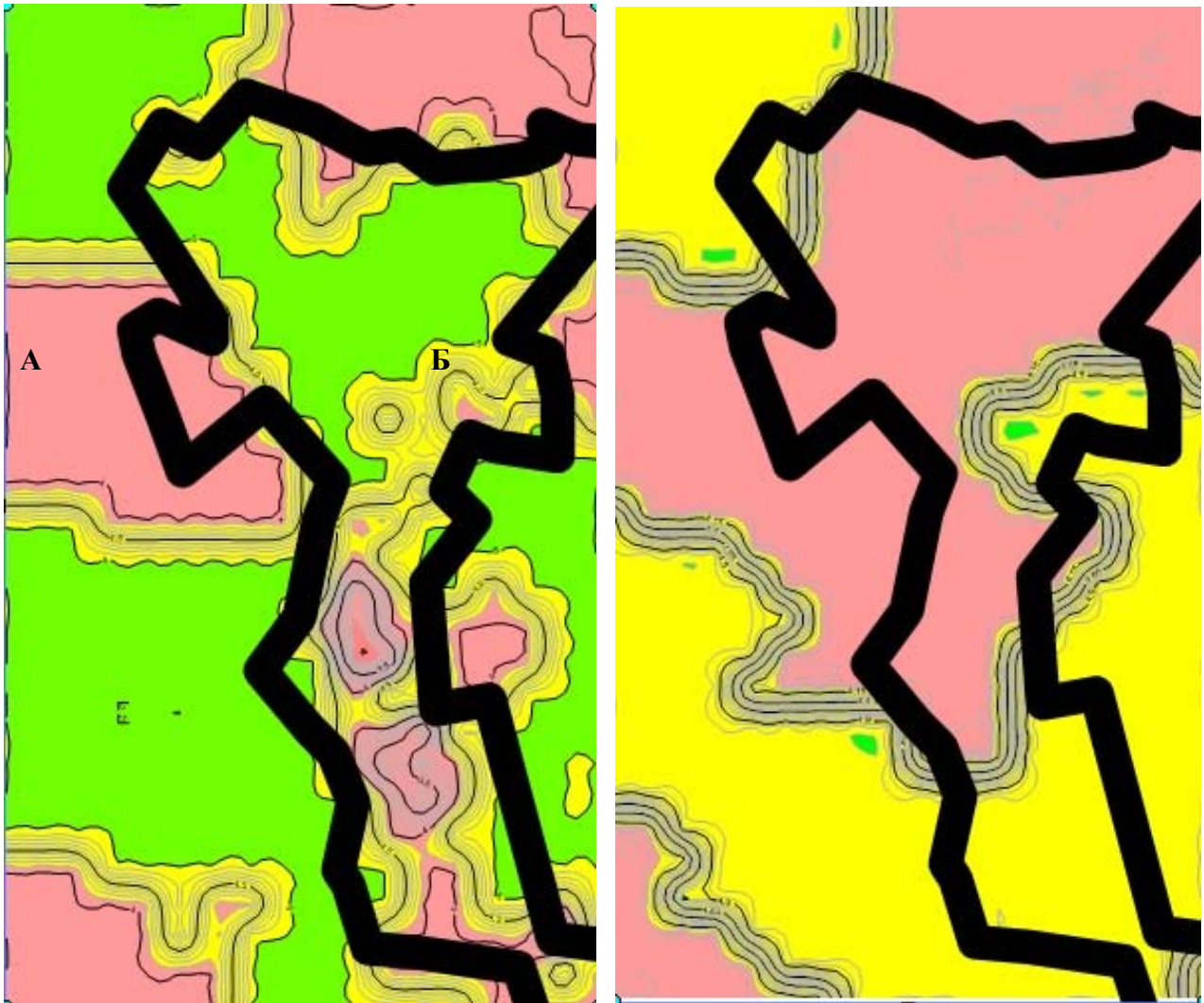
Однако стоит заметить, что специалистами во главе с Осиповым В.И. оценка данных районов проведена была. Поэтому, в отличии от компьютера, мы можем сравнить предсказанные им значения устойчивости с реальной ситуацией.

Данная тестовая выборка также содержала в себе данные по всем указанным выше критериям, характеризующих устойчивость геологической среды, в конкретных точках территории, однако сама категория устойчивости оставалась неизвестной.

Таким образом, построение компьютером модели устойчивости геологической среды по территориям этих двух районов наглядно показало возможности данной системы определять устойчивость геологической среды для любых территорий с похожими инженерно-геологическими условиями при наличии исключительно архивной информации и без проведения дополнительных изысканий

Эта тестовая выборка содержала в себе уже лишь 540 значений. Т.е. информации по этой территории было откровенно мало, но, за счёт уже накопленных при обучении «знаний», компьютер смог провести моделирование территории этих двух районов г. Москвы.

Для тестирования были использованы 2 алгоритма: Random forest, так как он показал наивысшую точность моделирования и Bagging, так как он показал наилучший результат из алгоритмов, не относящихся к группе деревьев решений. Результатом компьютерного моделирования стали схемы, указанные на рисунках 7 и 8.



Категории устойчивости геологической среды:

	Весьма опасная
	Опасная
	Малоопасная
	Неопасная

Рисунок 7. Результаты анализа модели устойчивости геологической среды по тестовой выборке при использовании алгоритмов Random Forest (А) и Bagging (Б).

Результаты данных расчетов были сравнены с результатами многолетних исследований группы специалистов (Рисунок 8).

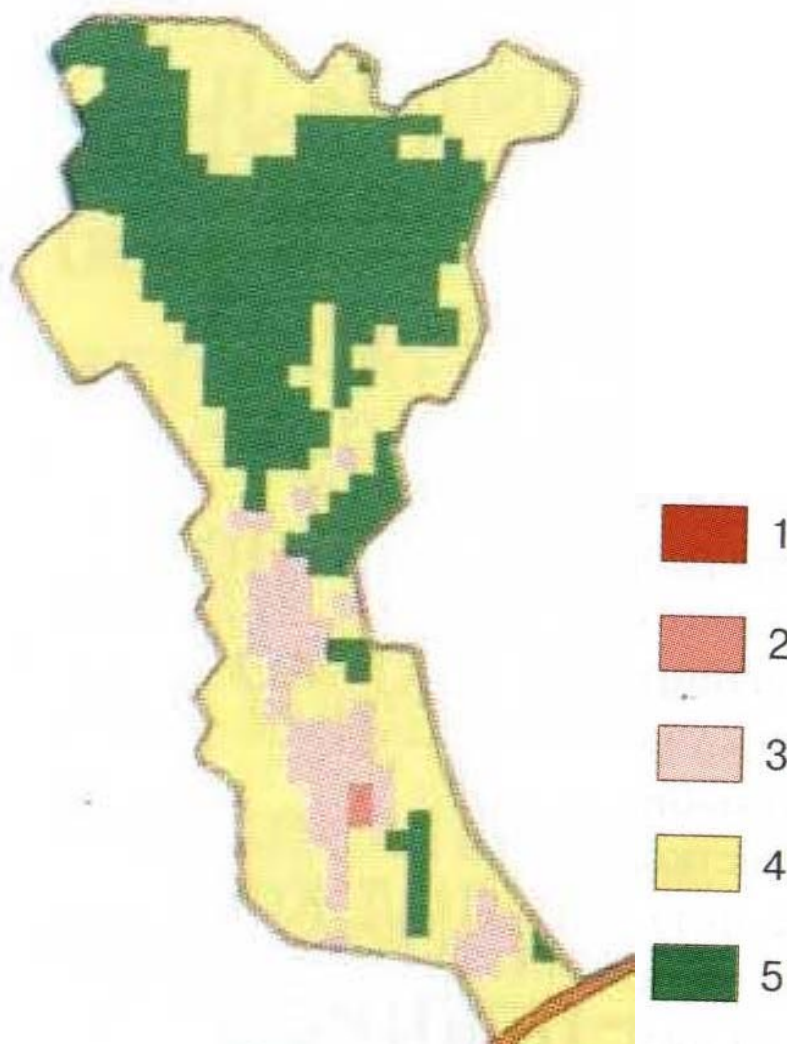


Рисунок 8. Фрагмент карты устойчивости геологической среды г. Москвы. Категории устойчивости геологической среды: 1 – чрезвычайно опасная, 2 – весьма опасная, 3 – опасная, 4 – малоопасная, 5 – неопасная [Осипов В.И. 1997].

Контуры изолиний модели построенной компьютером и построенной специалистами практически совпадают. А учитывая то, что эта оценка является скорее качественным прогнозом, нежели количественным, и точные реально существующие границы областей с различными категориями устойчивости геологической среды могут незначительно отличаться от предложенных специалистами, результаты моделирования можно считать ещё более успешными.

Для количественного определения сходства моделей был проведен анализ процентного соотношения территорий каждой категории в первой и второй модели. В результате получилось следующее соотношение категорий устойчивости в модели, построенной компьютером:

Весьма опасная – 0,04%

Опасная – 17,57%

Малоопасная – 48,82%

Неопасная – 33,57%

В то время как в модели, построенной специалистами, соотношение было такое:

Весьма опасная – 0,28%

Опасная – 8,94%

Малоопасная – 52,97%

Неопасная – 37,81%

Суммарная ошибка компьютерной модели составляет при этом 17,26%, что даже меньше, чем при обучении. Согласно проведенному тестированию модели можно полагать, что компьютерная модель хорошо себя показывает при выполнении прогнозов и, несмотря на трудную задачу, тестирование было успешным.

Следовательно, при появлении задачи анализа устойчивости геологической среды на похожих территориях можно использовать уже разработанную и представленную в данной работе модель в качестве аналога. При этом особенностью этой модели является то, что она является перманентно развивающейся. Т.е. при появлении новых проанализированных данных используемые алгоритмы машинного обучения могут скорректировать свою работу с учетом новой информации, что увеличит точность последующего моделирования. И с каждым новым исследованием точность будет всё ближе приближаться к 100%, в случае достаточного разнообразия исследуемых территории и правильной корректировки результатов определения устойчивости геологической среды.

Если учитывать результаты перекрестной проверки, анализа точности обучающей выборки, информацию, полученную из сопоставления данных результатов моделирования устойчивости геологической среды и выбранных критериев оценки в Москве, то с уверенностью можно сказать, что модель дает достоверные количественные результаты и может выполнять свою задачу. Также из анализа следует, что алгоритм Random Forest показал себя с лучшей стороны и модель, построенная на его основе, является более точной.

Согласно проведенному тестированию модели можно полагать, что компьютерная модель хорошо себя показывает при выполнении прогнозов. Несмотря на трудную задачу, тестирование было успешным. В итоге была получена модель, позволяющая оценить устойчивость геологической среды без непосредственного анализа опасных процессов. Параметры процессов были определены компьютером за счет имеющейся информации, описывающей рельеф, гидрогеологические условия и тектонику изучаемого района. Благодаря этому проведение оценки устойчивости геологической среды становится намного быстрее, что позволяет специалисту в большей степени сконцентрироваться на анализе всех возможных опасностей для строительства и лишь скорректировать построенную компьютером модель.

В работе было показано, что такое моделирование может быть осуществлено компьютером. Соответственно при увеличении обучающей выборки, введении корректур в структуру модели и проведении работ по созданию сбалансированной полноценной базы данных исходных параметров моделирования, данные для которой будут взяты из зон с различными

инженерно-геологическими условиями, возможно построение универсальной модели, способной выдавать прогнозы о наличии геологического риска в любых масштабах и на любых территориях с достаточно большой точностью.

Далее в главе 6 описываются пути применения методики и развития исследований.

Так полученная модель может быть использована для определения качественных рисков для проживания населения и строительства жилых кварталов, а также вследствие того, что устойчивости геологической среды влияет и на общее проявление природных процессов, построенная модель очень полезна при выборе мест расположения крупных промышленных объектов, таких как атомные электростанции.

В будущем это исследование может быть продолжено в разных направлениях. Дело в том, что уровень устойчивости геологической среды имеет связи со многими природными параметрами, которые могут быть рассчитаны и спрогнозированы. Также модель помогает оценить риски хозяйственного использования территории, качество территории для жизни. При этом машинное обучение является хорошей частью интеллектуального анализа модели и системы поддержки принятия решений.

Все описанные способы исследования развития являются очень перспективным и полезным. Это становится особенно актуальным, если учесть, что все модели в потенциале универсальны. Это означает, что с одной обучающей выборкой можно анализировать любые территории в любом масштабе. С одним хорошо подготовленным набором данных и одной хорошо рассчитанной моделью можно решить множество разных задач без внесения каких-либо существенных изменений.

ЗАКЛЮЧЕНИЕ

Разработанная методика определения устойчивости геологической среды на основе методов машинного обучения действительно позволяет осуществлять данный анализ на предпроектной стадии с достаточной для оптимизации инженерных изысканий точностью.

Машинное обучение можно рассматривать, как один из основных методов математического моделирования и прогнозирования устойчивости геологической среды, который позволяет обрабатывать большие объемы разнородной и сложной по составу инженерно-геологической информации.

Построенная в данной работе модель позволяет проводить оценку устойчивости геологической среды с точностью более 75% с учетом ее дальнейшего развития и повышении точности моделирования

Построенная модель, позволяет определять устойчивость геологической среды на различных территориях, сходных по инженерно-геологическим условиям с г. Москва. Однако, при расширении обучающей выборки, география потенциального применения модели может быть также расширена.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в журналах, входящих в Перечень рецензируемых научных изданий ВАК:

- 1) Рукавицын В.В. Анализ состояния экосистем крупных городов методами машинного обучения [Текст] / Рукавицын В.В. // Научно-методический журнал. Геоэкология, инженерная геология, гидрогеология, геокриология. – 2018. – №1. – С. 79-88. (10 п.л. / 10 п.л.)
- 2) Экзарьян В.Н., Рукавицын В.В. Методика оценки уровня устойчивости экосистем при помощи машинного обучения [Текст] / Экзарьян В.Н., Рукавицын В.В. // Известия высших учебных заведений. Научно-методический журнал. Геология и разведка. – 2015. – №1. – С.38-42. (5 п.л. / 3,1 п.л.)
- 3) Экзарьян В.Н., Рукавицын В.В. Разработка критериев оценки сохранности окружающей среды при нефтедобыче [Текст] / Экзарьян В.Н., Рукавицын В.В. // Известия высших учебных заведений. Научно-методический журнал. Геология и разведка. – 2016. – №4. – С.70-74. (5 п.л. / 3,6 п.л.)
- 4) Экзарьян В.Н., Рукавицын В.В., Зюляева М.В. Автоматизация комплексного экологического мониторинга территории аэропорта с использованием метода машинного обучения [Текст] / Экзарьян В.Н., Рукавицын В.В., Зюляева М.В. // Известия высших учебных заведений. Геология и разведка. – 2017. – № 4. – С. 72-78. (6,4 п.л. / 4,1 п.л.)

Публикации в других изданиях и материалах конференций:

- 5) Безноздрева Е.А., Воробьев Д.С., Емельянова Л.Г., Землянова О.И., Каргашина М.А., Касимов Д.В., Кривонос Е.В., Михеева А.И., Пинаев В.Е., Протопопов Н.Ф., Резников И.С., Рукавицын В.В., Рыбкина Г.И., Скворцова Е.А., Сосновских С.Л., Сысоева И.А., Хуршудов А.Г., Шахин Д.А., Якунин С.А. Сборник инновационных решений по сохранению биоразнообразия для нефтедобывающего сектора. – М.:РА ИЛЬФ. – 2015. – 274 с. (274 п.л. / 30 п.л.)
- 6) Рукавицын В.В. Анализ состояния экосистем крупных городов методами машинного обучения [Текст] / Рукавицын В.В. // Материалы XIII международной научно-практической конференции «Новые идеи в науках о земле». – М.: МГРИ-РГГРУ. – 2017. – С. 132-133. (2 п.л. / 2 п.л.)
- 7) Рукавицын В.В. Определение уровня устойчивости геологической среды с применением методов машинного обучения. Рукавицын В.В. // Материалы Международного молодежного научного форума «ЛОМОНОСОВ-2018» / Отв. ред. И.А. Алешковский, А.В. Андриянов, Е.А. Антипов. [Электронный ресурс] — М.: МАКС Пресс, 2018. (1 п.л./1п.л.)
- 8) Рукавицын В.В. Машинное обучение в экологическом прогнозировании [Текст] / Рукавицын В.В. // Материалы XI Международной научно-практической конференции «Наука о земле на современном этапе». – М.: ООО "Издательство "Спутник+". – 2014. – С 62-64. (2 п.л. / 2 п.л.)

9) Рукавицын В.В. Применение искусственного интеллекта для оценки степени риска проживания населения [Текст] / Рукавицын В.В. // Материалы международной научно-практической конференции «Развитие экологической экономики и образования в Украине и в мире». – Киев.: КНУ – 2011. – С 54-56. (3 п.л. / 3 п.л.)

10) Рукавицын В.В., Экзарьян В.Н. Оценка устойчивости экосистем России основная экологическая характеристика здоровья населения [Текст] / Рукавицын В.В., Экзарьян В.Н. // Здоровье - основа человеческого потенциала: проблемы и пути их решения. – Санкт-Петербург: Федеральное государственное автономное образовательное учреждение высшего образования "Санкт-Петербургский политехнический университет Петра Великого – 2014. – Т. 9. – № 2. – С. 827-829. (2,3 п.л. / 1,5 п.л.)

11) Rukavitsyn V. Real Text Data Analysis Using Machine-Learning Approach [Text] / Rukavitsyn V. // Conference proceedings MEKON 2011. – Ostrava. – 2011. – p. 23. (0,6 п.л. / 0,6 п.л.)

12) Rukavitsyn V., Zizka J. Modern Data-Mining Methods Applied to Real-World Agricultural/Economic Data Sets [Text] /Rukavitsyn V., Zizka J.// Conference proceedings MendelNET PEF 2009. – Brno. – 2009. – p. 59. (0,8 п.л. / 0,5 п.л.)

13) Rukavitsyn V., Zizka J. Automatic Categorization of Reviews and Opinions of Internet E-Shoping Customers [Text] /Rukavitsyn V., Zizka J.// International Journal of Online Marketing. – 2011. – Vol.1. – No.2. – pp. 70-79. (10 п.л. / 7 п.л.)

14) Rukavitsyn V., Zizka J.; Inductive Formation of Ecosystem Condition Models [Text] /Rukavitsyn V., Zizka J.// Conference proceedings 9th Summer School of Applied Informatic. – Bedřichov. – 2012. – pp. 143-151. (9 п.л. / 6 п.л.)

15) Rukavitsyn V., Zizka J. Opinion Classification in Text Entries Using Machine-Learning Approach [Text] /Rukavitsyn V., Zizka J.// Conference proceedings ISDMCI 2010. – Evpatoria. – 2010. – pp. 283-288. (6 п.л. / 4 п.л.)

16) Zizka J., Rukavitsyn V. Automatic Categorization of Reviews and Opinions of Internet E-Shopping Customers [Text] /Rukavitsyn V., Zizka J.// Transdisciplinary Marketing Concepts and Emergent Methods for Virtual Environments. – Pennsylvania: IGI Global. – 2012. – pp. 154-163. (10 п.л. / 6 п.л.)