

# ОПРЕДЕЛЕНИЕ ЗОН ВЕРОЯТНОГО ВОЗНИКНОВЕНИЯ СЕЙСМИЧЕСКИХ СОБЫТИЙ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

В. Г. Бондур<sup>1\*</sup> , О. С. Воронова<sup>1</sup> 

<sup>1</sup>Научно-исследовательский институт аэрокосмического мониторинга «АЭРОКОСМОС», Москва, Россия

\* **Контакт:** Валерий Григорьевич Бондур, vgbondur@aerocosmos.info

Для выявления зон вероятного возникновения сильных сейсмических событий предложены модели машинного обучения, основанные на методах регрессионного анализа. Проведена оценка производительности девяти линейных и нелинейных моделей, на основании результатов которой была выбрана модель случайного леса. Произведено улучшение качества обучения модели случайного леса за счет настройки гиперпараметров, а также использования кластеризации и полярных координат. Это позволило улучшить качество обучения модели, повысив значение коэффициента детерминации до 0,86. Проведен анализ возможности применения двух нейросетей с глубоким обучением, таких как многослойный перцептрон (Multi-layer Perceptron, MLP) и долгая краткосрочная память (Long Short-Term Memory, LSTM) с использованием для обучения параметров, которые были выбраны для модели случайного леса. С применением такой модели и выбранных нейросетей глубокого обучения были предсказаны зоны вероятного возникновения сейсмических событий для территории всего земного шара, а также детально проанализированы предсказанные зоны для территории Российской Федерации. В результате проведенных исследований применение нейросетей с глубоким обучением позволило выявить большее (приблизительно на 40%) количество зон максимальной сейсмичности (с магнитудами  $M \geq 6$ ) по сравнению с улучшенной моделью случайного леса.

**Ключевые слова:** землетрясения, сейсмоопасные территории, мониторинг, модели машинного обучения, нейросети.

**Цитирование:** Бондур В. Г., Воронова О. С. Определение зон вероятного возникновения сейсмических событий с использованием машинного обучения // Russian Journal of Earth Sciences. — 2025. — Т. 25. — ES5003. — DOI: 10.2205/2025es001055 — EDN: IOGILW

## 1. Введение

Землетрясения являются одними из самых опасных природных явлений [Болт, 1981]. Их распределение по земному шару имеет вполне определенные закономерности, связанные с природой этих явлений [Гельфанд и др., 1972]. Очаги землетрясений расположены, как правило, вдоль протяженных и относительно узких зон активных разломных структур земной коры и литосферы, в которых происходят сейсмогеодинамические взаимодействия плит [Хайн и Халилов, 2009].

Наиболее активная из сейсмических зон проходит вдоль западного побережья Южной Америки. Эта зона простирается вдоль границ нескольких плит. В Центральной Америке она разделяется на две ветви, одна из которых следует вдоль островной дуги Вест-Индии, а другая продолжается на север, расширяясь в пределах США, до западных хребтов Скалистых гор. Далее она проходит через Алеутские острова до Камчатки, а затем через Японские острова, Филиппины, Новую Гвинею и острова юго-западной части Тихого океана к Новой Зеландии и Антарктике [Пущаровский, 1972]. Две трети крупнейших землетрясений в мире приходится на этот Тихоокеанский пояс [Шерман и Злогодухова, 2011]. Катастрофические землетрясения в этом пояссе наносят наибольший

<https://elibrary.ru/ilogilw>

Получено: 26 мая 2025 г.

Принято: 2 сентября 2025 г.

Опубликовано: 8 сентября 2025 г.



© 2025. Коллектив авторов.

ущерб Японии. Землетрясения продолжают уносить человеческие жизни и наносить экономический ущерб и в других районах, лежащих в пределах этого пояса: Перу, Калифорния, Никарагуа, Филиппины и др. [Викулин и др., 2009].

Еще одна сейсмическая зона простирается от Азорских островов на восток через Альпы и Турцию. Хотя общее число землетрясений в этой зоне меньше, чем в Тихоокеанской, но примерно 75% жертв землетрясений во всем мире приходится именно на данный пояс, что объясняется высокой плотностью населения [Викулин и др., 2009].

В центральной части Атлантического океана выделяется также зона меньшего размера, расположенная вдоль Срединно-Атлантического хребта [Пущаровский, 2009]. Кроме того, существует ряд районов, где землетрясения происходят также довольно часто. К ним относятся Восточная Африка, Индийский океан, а в Северной Америке – долина реки Святого Лаврентия и северо-восток территории США [Шерман и Злододухова, 2011].

На территории Российской Федерации наиболее высокой сейсмичностью характеризуются Курило-Камчатский регион и остров Сахалин, район озера Байкал и Забайкалье, Северный Кавказ, Алтай, Саяны [Уломов, 2008; Федотов, 2005]. Менее сейсмичны Верхояно-Кольмский регион, районы Приамурья, Приморья, Корякии и Чукотки, но и в этих районах могут возникать достаточно сильные землетрясения [Федотов, 2005]. Относительно невысокая сейсмичность наблюдается на равнинах Восточно-Европейской, Скифской, Западно-Сибирской и Восточно-Сибирской платформ [Уломов, 2008].

Анализ природных опасностей свидетельствует о том, что такие стихийные бедствия, как сейсмические события и связанные с ними негативные воздействия на отдельные территории, представляют растущую угрозу безопасности людей и экономике [Осипов и др., 2000]. В связи с этим очень важным является проведение систематического мониторинга сейсмоопасных территорий, в том числе космического, и исследование особенностей изменения различных прогностических параметров, проявляющихся в различных средах, с целью их последующего использования для выявления предвестниковых особенностей, возникающих перед землетрясениями [Бондур и др., 2023, 2022; Липеровский и др., 2008; Akhoondzadeh and Marchetti, 2023].

Исследование и анализ предвестниковых явлений осуществляется, как правило, для уже произошедших сейсмических событий, когда известно место и время их свершения [Mogi, 1988; Соболев и Пономарев, 2003; Федотов, 2005]. Исследования, проведенные в последние годы, позволили установить связи между аномалиями различных геофизических полей, регистрировавшихся, в том числе по спутниковым данным, а также динамикой литосферных, атмосферных и ионосферных процессов в периоды подготовки и протекания сильных землетрясений, состоявшихся в различных регионах мира. Такие исследования производились для землетрясений, происходивших на различных территориях, например: Турции [Бондур и др., 2023; Akhoondzadeh and Marchetti, 2023]; Италии [Piscini et al., 2017]; Китая [Jiao et al., 2025]; Японии [Saito et al., 2011]; США (Южная Калифорния) [Бондур и др., 2010, 2016; Гапонова и др., 2019]; Чили [Смирнов и др., 2018]; Суматры [Бондур и др., 2007] и в других районах мира. Подобные аномалии регистрировались также в различных регионах Российской Федерации, таких как территории: Курило-Камчатской гряды [Bogdanov et al., 2020]; Северного Кавказа [Бондур и Воронова, 2022]; Республики Тыва [Кашкин и др., 2012]; в районе озера Байкал [Бондур и др., 2022] и в других регионах страны.

Прогноз землетрясений и составление точных прогнозных карт сейсмической опасности – проблема, которая пока не нашла своего окончательного решения. По продолжительности периодов, на которые должны осуществляться прогнозы землетрясений, обычно выделяют: долгосрочные [Федотов, 2005], среднесрочные [Кейлис-Борок, 1989; Кособоков, 2005], краткосрочные [Соболев и Пономарев, 2003]. В последнее время наблюдается существенный прогресс в детализации и уточнении прогнозов сейсмической опасности территорий. Однако, при этом возникает ряд проблем, связанных с достоверностью прогнозов, прежде всего, краткосрочных и среднесрочных, а также оценкой сейсмических рисков [Гвишиани и др., 2020].

Целью данного исследования являлось выявление зон вероятного возникновения сейсмических событий по результатам анализа пространственного распределения сильных землетрясений ( $M \geq 6$ ), построенного на основании данных периода с 1949 по 2024 гг. путем применения различных моделей машинного обучения, а также нейросетей с глубоким обучением.

Актуальность исследований, проводимых в рамках настоящей работы, обусловлена большим ущербом, наносимым такими катастрофическими природными процессами, как землетрясения, а также необходимостью предоставления оперативной информации о них для оценки рисков возникновения сильных землетрясений ( $M \geq 6$ ) в интересах их предупреждения и снижения последствий с использованием современных методов машинного обучения.

Применение методов машинного обучения в области геологии и сейсмологии позволяет усовершенствовать традиционные методы, которые использовались в предыдущие годы, и предоставляет возможность более достоверной оценки сейсмических рисков [Agarwal *et al.*, 2023].

При реализации методов машинного обучения важным являются способы формирования наборов данных и определение оптимального количества признаков, которые модель машинного обучения должна проанализировать и предсказать [Mallowhy *et al.*, 2019]. При этом следует отметить, что не существует одного лучшего метода, который подходит для решения всех проблем, связанных с формированием наборов данных, необходимых для реализации методов машинного обучения. Это обусловлено тем, что производительность таких методов зависит от многих факторов. Некоторые методы лучше всего подходят для небольших объемов данных, в то время как другие лучше работают для выборок данных значительных объемов. Ряд методов требует категориальных входных данных, а другие – количественных.

В настоящей работе на основании использования наборов данных о произошедших землетрясениях были исследованы возможности применения девяти различных моделей машинного обучения, таких как: линейная регрессия [Jarah *et al.*, 2024]; случайного леса [Shukla *et al.*, 2024]; градиентный бустинг [Ajai *et al.*, 2024];  $k$ -ближайших соседей [Ranjan *et al.*, 2019]; дерево принятия решений [Ermiş and Cürebal, 2024]; лассо регрессии [Mignan *et al.*, 2024]; ридж-регрессия [Liu and Macedo, 2022]; градиентной бустинговой сборки [Wei and Gao, 2024] и опорных векторов [Asim *et al.*, 2018]. Кроме того, применялись нейросети глубокого обучения, такие как: многослойный перцептрон [Mahmoudi *et al.*, 2016] и долгая краткосрочная память [Vardaan *et al.*, 2019]. Применение оптимизации гиперпараметров, введение дополнительных параметров, таких как кластеризация, полярные координаты и использование глобальных данных позволило улучшить обучение выбранной модели случайного леса. Настройка архитектуры нейросетей для глубокого обучения с использованием дополнительных слоев и с активационными функциями позволила выделить большее количество зон с магнитудами более 6.

## 2. Исходные данные и особенности методики применения машинного обучения

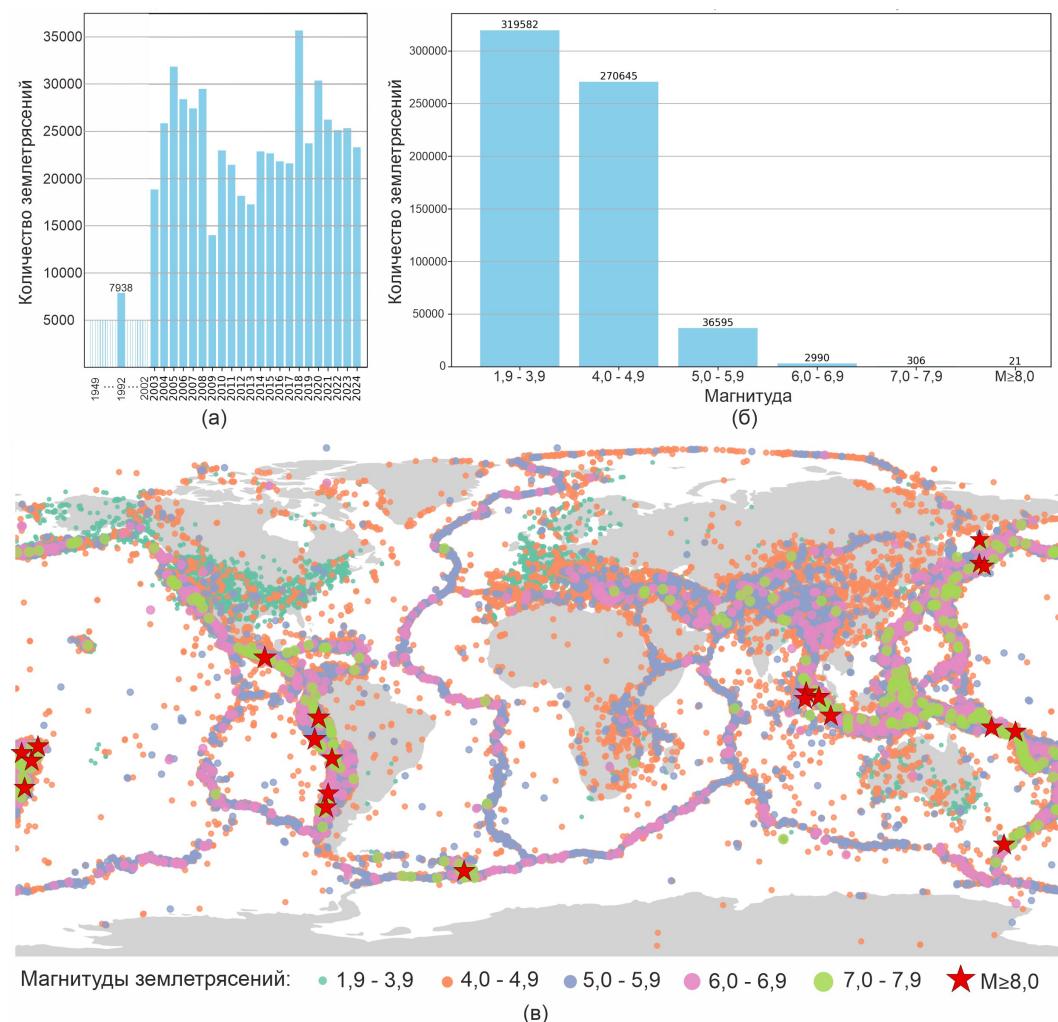
### 2.1. Формирование массива данных и предварительная обработка

Информация о произошедших землетрясениях была получена из архива Служба срочных донесений Геофизической службы РАН (СЕМЕ) [<http://www.ceme.gsras.ru>], данные которого представляются на территорию Российской Федерации. Также использовались данные из архива Геологической службы США (USGS – United States Geological Survey) [<http://earthquake.usgs.gov>] для всей территории земного шара. Для создания качественного каталога сейсмичности устранилось дублирование данных из двух архивов. В процессе поиска дубликатов были заданы критерии уникальности и уровень округления данных. Дополнительно из архива СЕМЕ была извлечена информация о количестве сейсмостанций, зарегистрировавших сейсмическое событие. Сравнение информации из двух архивов осуществлялось по столбцам даты, координат и магнитуды каждого сейсмического события. Сформированный массив данных был использован для анализа сейсмических событий, произошедших по всему Миру, начиная с 1949 г.

Для проведения исследований была осуществлена предварительная обработка исходных данных, включающая определение кодировки файла о произошедших землетрясениях, загрузку и фильтрацию данных по заданным критериям. Фильтрация осуществлялась с учетом необходимости устранения строк, в которых присутствовали ошибки, а также пропусков в столбцах.

В массивах данных, полученных из архива USGS, удалялись строки, если в столбце тип сейсмического события отсутствовало упоминание «earthquake». Это означало, что данное событие не принадлежало к классу землетрясений. Для повышения достоверности данных был использован параметр `nst`, содержащий информацию о количестве сейсмостанций, зарегистрировавших данное сейсмическое событие. Чем больше сейсмостанций зарегистрировало конкретное землетрясение, тем более точные координаты эпицентров сейсмических событий были получены.

На [рис. 1а](#) представлена информация о количестве землетрясений (данные из архивов СЕМЕ и USGS), происходивших на всем земном шаре ежегодно с 1949 по 2024 гг. Анализ [рис. 1а](#) свидетельствует о том, что ежегодное количество землетрясений, происходивших с 1949 по 2002 гг., не превышало 5000. Исключением являлся 1992 г., когда было зарегистрировано 7938 сейсмических событий (см. [рис. 1а](#)). В последующие годы (начиная с 2002 г.) были разработаны и использованы новые приборы и ме-



**Рис. 1.** Землетрясения, произошедшие на земном шаре с 1949 по 2024 гг. [<http://www.ceme-gsras.ru>; <http://earthquake.usgs.gov>]: а) ежегодное количество землетрясений; б) количество землетрясений, происходивших с различными магнитудами; в) пространственное распределение очагов землетрясений с различными магнитудами, происходившими за последние 75 лет.

тоды, усовершенствованы сейсмометры, что позволяло регистрировать существенно большее число землетрясений [Frankel *et al.*, 2002]. Сейсмические события различных периодов времени имели разную плотность распределения, что может влиять на получение итогового результата проводимых исследований. В связи с этим было применено масштабирование информационных параметров, которое помогает нормализовать данные и позволяет модели машинного обучения (МО) лучше понимать их структуру и выявлять закономерности.

На [рис. 1б](#) приведено распределение по годам количества землетрясений с различными магнитудами, происходивших с 1949 по 2024 гг. На [рис. 1в](#) представлено пространственное распределение очагов землетрясений с различными магнитудами, происходившими на земном шаре за последние 75 лет.

Для проведения дальнейших исследований после осуществления предварительной обработки сформированный массив данных был разделен на обучающую и тестовую выборки. Для обучающей выборки использовался набор данных, сформированных с 1949 до 2020 г., а в качестве тестовой выборки использовались данные о землетрясениях, происходивших с 2021 по 2024 гг.

Весь процесс обработки данных и применение методов машинного обучения осуществлялись в интерактивной облачной среде Google Colab [<https://colab.research.google.com>] с использованием программного кода, написанного на языке программирования Python и общедоступных библиотек pandas, numpy, sklearn, seaborn.

## 2.2. Выбор модели машинного обучения

Для проведения исследований нами были выбраны линейные и нелинейные модели машинного обучения, алгоритмы реализации которых основаны на регрессионном анализе. Регрессионный анализ позволяет учитывать дисбаланс плотности распределения значений сейсмических событий и представлять более глубокую интерпретацию данных. В качестве линейных моделей были выбраны такие, как: линейная регрессия (Linear Regression); лассо-регрессия (Lasso Regression); ридж-регрессия (Ridge Regression). В качестве нелинейных моделей машинного обучения были использованы следующие модели: случайный лес (Random Forest); градиентный бустинг (Gradient Boosting);  $k$ -ближайших соседей ( $k$ -Nearest Neighbors); дерево решений (Decision Tree); градиентная бустинговая сборка (Light Gradient Boosting Machine, LightGBM); метод опорных векторов (Support Vector Machine, SVM).

Для обучения моделей использовались такие параметры, как: даты произошедших землетрясений (день, месяц и год); координаты их эпицентров (широта, долгота); количество сейсмостанций, зарегистрировавших каждое сейсмическое событие; глубины очагов сейсмических событий. В качестве целевой переменной при этом была выбрана величина магнитуды свершившегося землетрясения. При этом набор данных для обучения всех выбранных 9-ти моделей машинного обучения был разделен на обучающую и тестовую выборки в процентном соотношении 80% на 20% соответственно. Для получения сравнительного анализа производительности каждой из девяти моделей их обучение происходило в примерно одинаковых условиях. Оценка производительности моделей осуществлялась с использованием различных метрик [Хайкин, 2006]: средней ошибки при кросс-валидации (CV MSE); средней ошибки модели на тестовом наборе (mean squared error, Test MSE); средней абсолютной ошибки (mean absolute error, MAE); а также коэффициента детерминации ( $R$ -squared,  $R^2$ ).

Средняя ошибка при кросс-валидации (CV MSE) позволяла оценивать, насколько хорошо модель предсказывает значения магнитуд на основе обучающих данных, чем меньше значение ошибки, тем лучше.

Средняя ошибка модели на тестовом наборе (Test MSE) показывала, насколько хорошо модель будет работать на новых данных. Чем меньше значение ошибки, тем лучше производительность модели.

Средняя абсолютная ошибка (MAE) позволяла оценивать, насколько предсказания отличаются от реальных значений в среднем. Чем меньше значение ошибки, тем предсказания более точные.

Коэффициента детерминации  $R^2$  – это статистическая мера, которая представляет собой долю дисперсии для зависимой переменной, которая объясняется независимой переменной или переменными в регрессионной модели. Метрика качества обучения модели  $R^2$  указывала насколько хорошо модель описывает данные. Её значения варьируются от 0 до 1. При этом 1 – это наилучший результат.

В результате проведения сравнительного анализа оценки производительности каждой из 9-ти моделей, обучение которых осуществлялось с использованием написанного программного кода, были получены метрики качества обучения каждой модели (табл. 1).

**Таблица 1.** Результаты обучения моделей машинного обучения

Название модели	CV MSE	Test MSE	MAE	$R^2$
Линейная регрессия	0,379	0,382	0,491	0,497
Случайный лес	0,124	0,122	0,256	0,839
Метод опорных векторов	0,132	0,193	0,230	0,601
Градиентный бустинг	0,172	0,173	0,311	0,772
$k$ -ближайших соседей	0,149	0,148	0,283	0,704
Дерево принятия решений	0,231	0,224	0,339	0,705
Лассо-регрессия	0,756	0,761	0,777	0,001
Ридж-регрессия	0,379	0,382	0,491	0,497
Градиентная бустинговая сборка	0,131	0,126	0,262	0,794

*Примечание к таблице:* CV MSE – средняя ошибка при кросс-валидации, Test MSE – средняя ошибка модели на тестовом наборе, MAE – среднее абсолютное значение ошибок предсказания.

Анализ полученных результатов, приведенных в табл. 1, показал, что наиболее эффективной оказалась модель случайного леса (Random Forest), благодаря своей способности учитывать взаимодействия между признаками и уменьшать переобучение. Наименьшие значения метрик качества модели случайного леса были получены для средней ошибки при кросс-валидации (0,124), средней ошибки модели на тестовом наборе (0,122), а также среднего абсолютного значения ошибок предсказания (0,256). Полученные значения коэффициента детерминации  $R^2$  (0,839) позволяли судить о том, что независимые переменные могут достаточно хорошо предсказать зависимую переменную.

### 2.3. Улучшение модели машинного обучения случайного леса (Random Forest)

На первом этапе проведения исследований улучшалось обучение выбранной модели случайного леса путем оптимизации гиперпараметров, влияющих на её производительность. Выбор гиперпараметров осуществлялся с использованием метода случайного поиска [Jarah *et al.*, 2024; Shukla *et al.*, 2024], по заданному диапазону их изменений и применением результатов кросс-валидации модели с целью выявления оптимальных гиперпараметров. Случайный поиск (RandomizedSearchCV) – это метод, который выбирает случайные комбинации параметров из заданного диапазона их изменений и оценивает их, чтобы найти наиболее оптимальные.

На втором этапе исследований к модели случайного леса с уже настроенными гиперпараметрами с использованием метода случайного поиска для обучения был введен дополнительный параметр – кластеризация на основе координат (широты, долготы) и глубин очагов землетрясений. При этом применялся алгоритм кластеризации  $k$ -means [Novianti *et al.*, 2017]. Суть кластеризации  $k$ -means заключается в инициировании среднего значения по каждой из переменных (центроидов) для каждого кластера объектов,

сформированных на основе среднего арифметического результатов наблюдений, а также в их переопределении путём минимизации суммарного квадратичного отклонения от центроидов кластеров.

Одним из самых сложных этапов применения алгоритма  $k$ -means является выбор оптимального числа кластеров  $k$ . Выбор таких значений  $k$  осуществлялся нами путем построения и анализа графика зависимости суммарной внутрикластерной дисперсии от количества кластеров. Внутрикластерная дисперсия (или сумма квадратов расстояний между объектами и их центроидом) показывала, насколько компактными являются кластеры. Чем меньше внутрикластерная дисперсия, тем более упорядочены и однородны кластеры.

На рис. 2а представлен построенный график зависимости суммарной внутрикластерной дисперсии от количества кластеров. Анализ рис. 2а позволил выявить точку ( $k = 3$ ), после которой дальнейшее увеличение числа кластеров не приводило к значительному снижению внутрикластерной дисперсии, а значит не улучшало качество кластеризации.

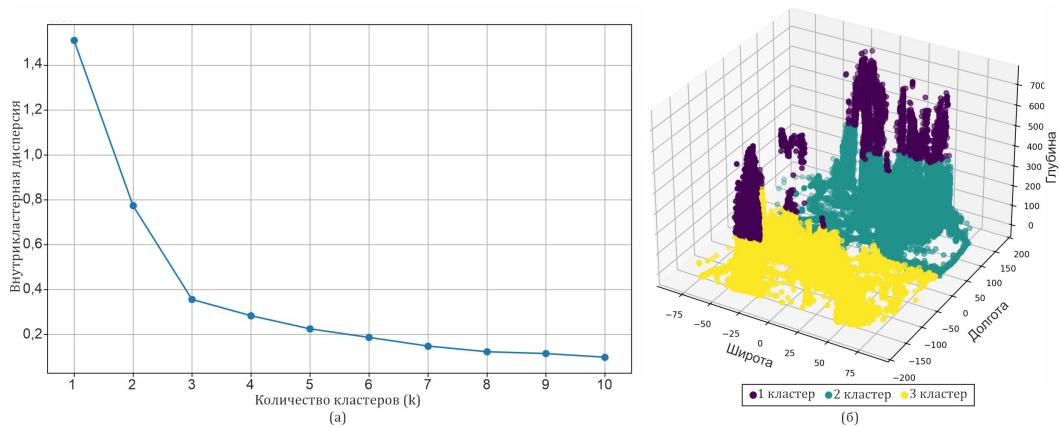


Рис. 2. График зависимости суммарной внутрикластерной дисперсии от количества кластеров (а); результат кластеризации землетрясений по координатам и глубине очага (б).

Выявленное количество кластеров ( $k = 3$ ) было использовано для кластеризации с применением алгоритма  $k$ -means. Результат такой кластеризации представлен на рис. 2б. На этом рисунке каждая точка в трехмерном пространстве окрашена в зависимости от принадлежности к определенному кластеру. Анализ рис. 2б, показал, что кластеры четко разделяются друг от друга без значительного наложения. Это свидетельствовало об эффективности используемой модели, позволяющей выделять группы данных со схожими признаками и улучшать прогнозирование.

На третьем этапе проведения исследований дополнительно был введен в обучение модели случайного леса еще один параметр (полярные координаты: радиус и угол) к уже настроенным гиперпараметрам с использованием метода случайного поиска и кластеризаций. Преобразование координат в полярные и использование их в качестве дополнительного признака позволяло модели улучшать качество обучения.

В результате проведения исследований после каждого из этапов улучшения обучения модели случайного леса были получены метрики качества, представленные в табл. 2.

Реализованные процедуры позволили улучшить значения метрик качества модели случайного леса (табл. 2), в том числе: уменьшить среднюю ошибку MSE (с 0,124 до 0,106), уменьшить среднюю абсолютную ошибку MAE (с 0,257 до 0,238) и повысить коэффициент детерминации  $R^2$  (с 0,839 до 0,859).

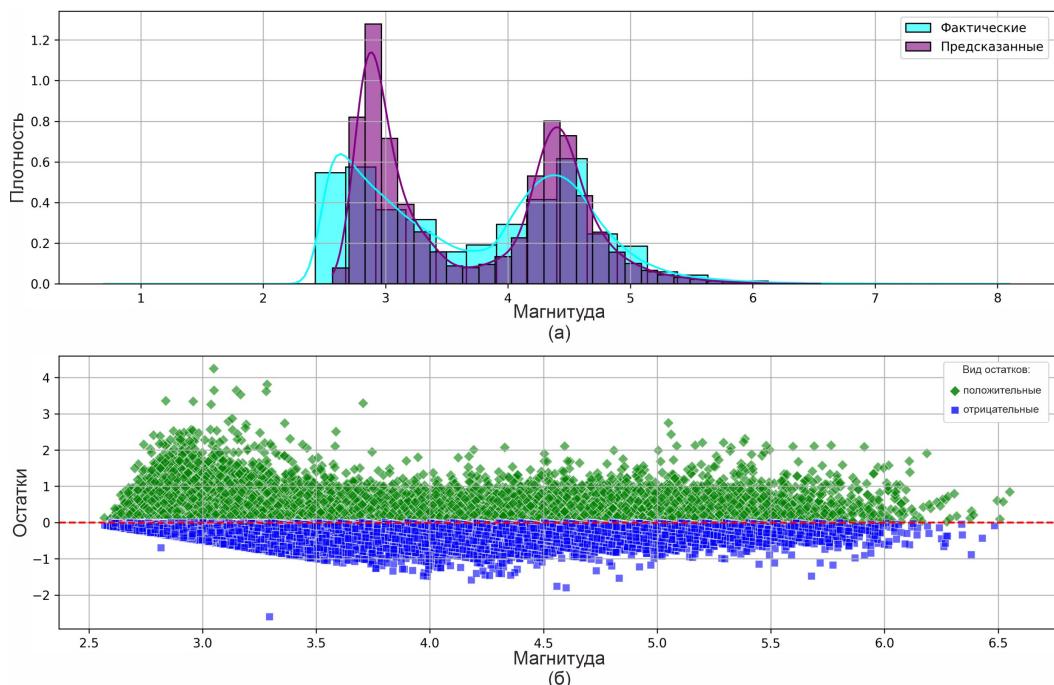
На следующем этапе исследований была осуществлена проверка качества обучения модели случайного леса на тестовых данных, содержащих информацию о землетрясениях, происходивших с 2021 по 2024 гг. Были также рассчитаны вероятности попадания тестовых магнитуд землетрясений в предсказанные зоны сейсмичности. Расчет вероятно-

**Таблица 2.** Результаты улучшения обучения модели случайного леса

Этапы проведения исследований	MSE	MAE	$R^2$
Модель случайного леса (RF)	0,124 269	0,256 715	0,839 505
1. Модель случайного леса + метод случайного поиска гиперпараметров	0,118 873	0,253 63	0,843 291
2. Модель случайного леса + метод случайного поиска гиперпараметров + кластеризация	0,108 391	0,242 259	0,857 207
3. Модель случайного леса + метод случайного поиска гиперпараметров + кластеризация + полярные координаты	0,106 210	0,238 596	0,859 416

стей, выраженных в процентах, осуществлялся как отношение количества фактических магнитуд в сейсмической зоне к общему количеству предсказанных значений магнитуд в этой зоне. Эти вероятности составляли: для зоны сейсмичности с магнитудами  $M = (1,9-3,9)$  – 86,42%; для зоны сейсмичности с магнитудами  $M = (4,0-4,9)$  – 82,14%; для зоны сейсмичности с магнитудами  $M = (5,0-5,9)$  – 59,40%; для зоны сейсмичности с магнитудами  $M = 6$  и более – 26,95%. Анализ результатов проверки качества обучения на тестовых данных показал, что модель случайного леса достаточно хорошо выявляет зоны сейсмичности до магнитуды  $M = 5,9$ . В связи с меньшей плотностью распределения сильных землетрясений с  $M = 6$  используемую модель случайного леса необходимо дорабатывать. Например, вводить дополнительные параметры обучения, а также улучшать алгоритмы самой модели.

Визуализация результатов проверки качества обучения модели случайного леса на тестовой выборке представлена на [рис. 3](#).



**Рис. 3.** Результаты проверки качества обучения модели случайного леса на тестовой выборке: а) гистограмма фактических и предсказанных значений магнитуд землетрясений; б) график остатков (ромбы-остатки от 0 до 10; квадраты-остатки от -10 до 0).

Анализ [рис. 3](#) позволил выявить наличие перекрытия между гистограммами фактических и предсказанных значений магнитуд землетрясений. Это указывает на эффективность модели случайного леса для предсказания сейсмических событий. Из анализа [рис. 3](#) следует, что, например, для магнитуд землетрясений  $M = 2,4-2,6$  их предсказанные значения магнитуд занижены относительно фактических, а для магнитуд землетрясений  $M = 2,7-3,1$  их предсказанные значения, наоборот, завышены.

ны относительно фактических значений (см. [рис. 3а](#)). Для магнитуд землетрясений  $M = 3,2\text{--}4,1$  фактические и предсказанные значения практически совпадают. Небольшое завышение предсказанных значений магнитуд землетрясений выявлено для диапазона магнитуд  $M = 4,2\text{--}4,5$ . Наилучшее совпадение фактических и предсказанных значений магнитуд выявлено для землетрясений с магнитудами  $M \geq 4,5$ .

На [рис. 3б](#) представлена разница между фактическими и предсказанными магнитудами (график остатков). Положительные значения остатков (ромбы на [рис. 3б](#)) указывают на то, что используемая модель недооценила значения магнитуд землетрясений, а отрицательные (квадраты на [рис. 3б](#)) наоборот переоценила. Наличие горизонтальной линии на уровне 0 помогает визуально оценить, как остатки распределены относительно нуля. Идеальная модель должна иметь остатки, сосредоточенные вокруг этой линии. Распределение остатков носит, в основном, равномерный характер. Однако, в интервале магнитуд от  $M = 2,6$  до  $M = 3,9$  отрицательных остатков меньше (квадраты, [рис. 3б](#)), это говорит о том, что модель реже переоценивала предсказанные значения магнитуд относительно фактических.

Для оценки качества обучения модели случайного леса был проведен анализ предсказанных зон, которые не подтвердились на тестовых данных. Модель вывела относительно низкий процент предсказанных, но не подтвержденных случаев (до 4%), что говорит о её способности правильно идентифицировать большинство магнитуд. Для зон с магнитудами 6 и более наблюдалось крайне низкое количество предсказаний и подтвержденных случаев, что указывает на недостаточную чувствительность модели к сильным сейсмическим событиям. Процент пропущенных магнитуд в диапазоне от 0,14% до 3,67% указывает на то, что модель иногда не распознает землетрясения, которые должны попасть в определенные зоны.

Проведенный анализ результатов обучения модели на тестовой выборке позволил сделать вывод, что улучшенная модель случайного леса хорошо предсказывает магнитуды землетрясений, однако требует доработки для улучшения точности в зонах сильных магнитуд.

#### 2.4. Процесс обучения нейросетей

Для проведения исследований были выбраны нейросети глубокого обучения, такие как Многослойный перцептрон (Multi-layer Perceptron, MLP) и нейросеть с архитектурой Долгая краткосрочная память (Long Short-Term Memory, LSTM) [[Mahmoudi et al., 2016](#); [Vardaan et al., 2019](#)].

Процесс обучения этих нейросетей проводился на том же массиве данных (с 1949 по 2024 гг.), что и улучшенная модель случайного леса. Импортирование всех необходимых библиотек и функций для выбранных нейросетей осуществлялось также с использованием программного кода, написанного на языке программирования Python в программной среде Google Colab [<https://colab.research.google.com>].

Архитектура полносвязной нейросети (MLP), которая использовалась в нашем исследовании, состояла из входного слоя (соответствовал количеству признаков в обучавшем наборе), пяти скрытых слоев и выходного слоя, генерировавшего предсказанное значение. Первый скрытый слой позволял модели учиться на нелинейных зависимостях и состоял из 512 нейронов с функцией активации ReLU (Rectified Linear Unit). Второй слой с 256 нейронами, третий с 128, четвертый с 64, а пятый с 32-я нейронами. Эти слои находятся между входным и выходным слоями и позволяли фиксировать сложные взаимосвязи в данных.

При использовании нейросети MLP также был применен метод обучения обратного распространения ошибки и оптимизатор (Adam) [[Хайкин, 2006](#)], который автоматически подстраивал скорость обучения для каждого параметра.

Применение нейросети с архитектурой LSTM осуществлялось с использованием двух слоёв (с 64 и с 32 нейронами), а также выходного слоя с одним нейроном. В нейросеть LSTM была включена функция активации ReLU (Rectified Linear Unit), помогающая более эффективно справляться с проблемами затухающего градиента

[Хайкин, 2006]. Для предотвращения переобучения в нейросеть LSTM входил в том числе слой (Dropout), который обнулял 20% нейронов во время обучения.

Качество обучения нейросетей было проанализировано по метрикам, приведенным в табл. 3.

**Таблица 3.** Результаты обучения нейросетей

Нейросети	MSE	MAE	$R^2$
Многослойный перцептрон (MLP)	0,137	0,276	0,819
Долгая краткосрочная память (LSTM)	0,153	0,293	0,797

Анализ полученных метрик позволил обнаружить их незначительное ухудшение по сравнению с улучшенной моделью случайного леса (см. табл. 2): MSE на 0,03–0,05; MAE на 0,04–0,06 и  $R^2$  на 0,04–0,06. Ввиду несущественных отличий полученных метрик можно предположить, что используемые нейросети могут быть применены для предсказания зон максимальной сейсмичности.

### 3. Результаты исследований

Улучшенная модель случайного леса (Random Forest, RF) с настроенными гиперпараметрами, кластеризацией  $k$ -means и полярными координатами в качестве дополнительных параметров, была обучена на полном массиве данных о землетрясениях, происходивших с 1949 по 2024 гг. В результате этого обучения модели были получены значения коэффициента детерминации  $R^2 = 0,863$ , значения метрик MSE = 0,101 и MAE = 0,230.

По результатам обучения улучшенной модели случайного леса, а также нейронных сетей (MLP, LSTM) на 75-летнем массиве данных о землетрясениях были предсказаны зоны вероятного возникновения сейсмических событий для территории всего земного шара. На рис. 4 представлены распределения этих зон на примере территории Российской Федерации, полученных с использованием модели случайного леса (RF) – рис. 4а, а также нейросетей Многослойный перцептрон (MLP) – рис. 4б и Долгая краткосрочная память (LSTM) – рис. 4в.

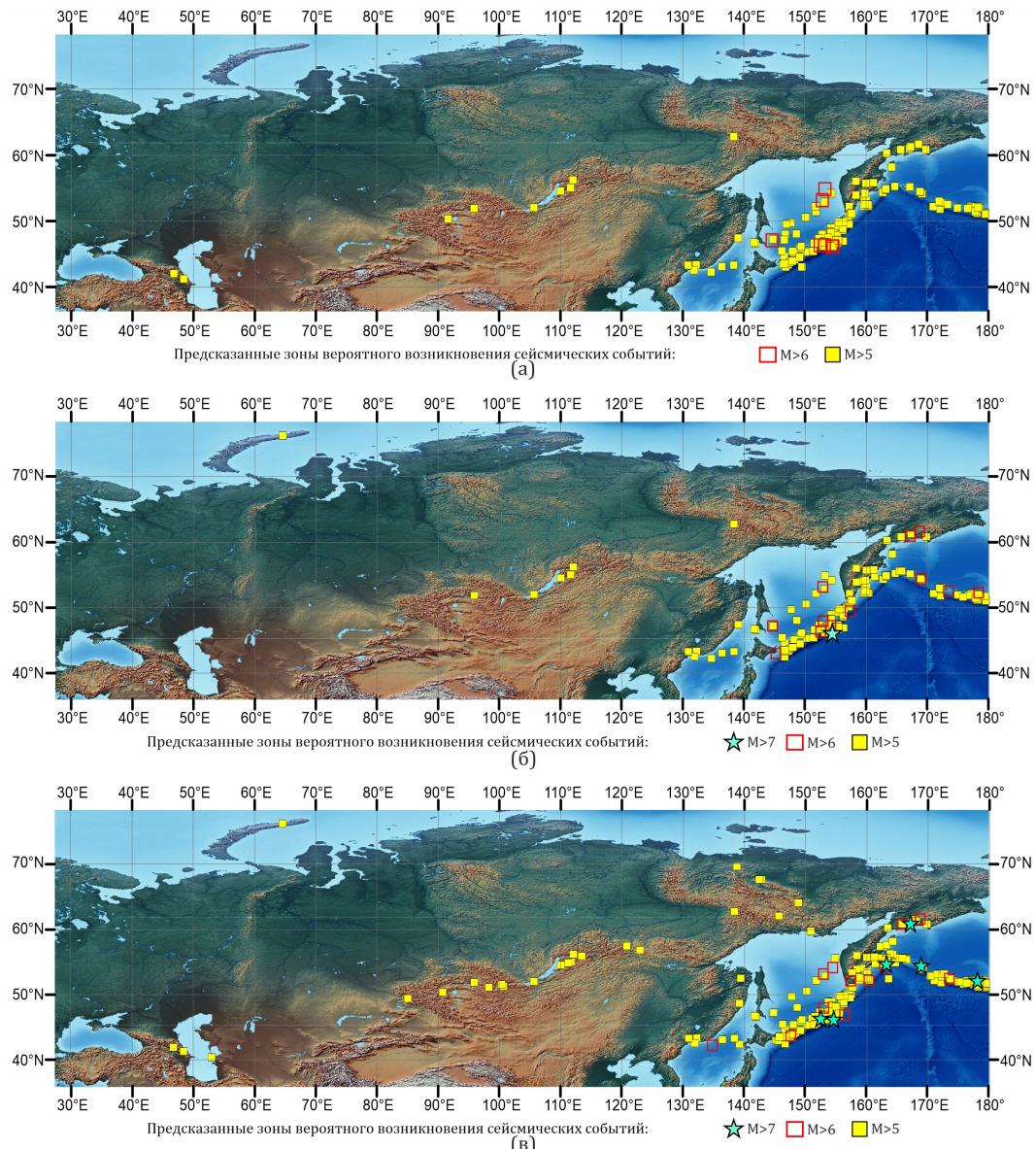
Анализ результатов, представленных на рис. 4а, показал, что применение улучшенной модели RF не позволило спрогнозировать зоны сейсмичности с магнитудами  $M \geq 7$ . Вероятно, это связано с малым количеством обучающих данных с большой магнитудой на территории Российской Федерации. Однако, точность и качества производительности модели RF имели наилучшие показатели метрик в отличие от нейросетей глубокого обучения MLP и LSTM.

Применение многослойной нейросети глубокого обучения MLP позволило спрогнозировать зоны максимальной сейсмичности с магнитудами  $M \geq 7$  (см. рис. 4б). Особая форма нейросети LSTM, способная анализировать долгосрочные зависимости в данных, позволила предсказать наибольшее количество зон высокой сейсмичности с магнитудами  $M \geq 7$  на территории Российской Федерации. Это иллюстрируется на рис. 4в.

На рис. 5 представлены результаты предсказания зон вероятного возникновения сильных сейсмических событий на всем земном шаре, полученные с использованием улучшенной модели RF, а также нейросетей глубокого обучения MLP и LSTM.

Для анализа полученных результатов производилось построение круговых диаграмм, использование которых позволило выявить сходство и различия в количестве предсказанных зон сейсмической активности с магнитудами  $M = 5,0$ – $5,9$  и  $M > 6$  для модели RF, а также нейросетей MLP и LSTM (рис. 5а). Анализ рис. 5а показал, что наибольшее количество уникальных зон сейсмической области с магнитудами  $M = 5,0$ – $5,9$  (1478) было предсказано нейросетью LSTM. Уникальными считались зоны, которые не были предсказаны другой нейросетью или моделью машинного обучения.

Сопоставление результатов предсказания, полученных моделью RF и нейросетями MLP и LSTM в сейсмической области с магнитудами 5,0–5,9 (см. левую часть рис. 5а), позволило обнаружить общее количество совпадающих зон модели RF и ней-

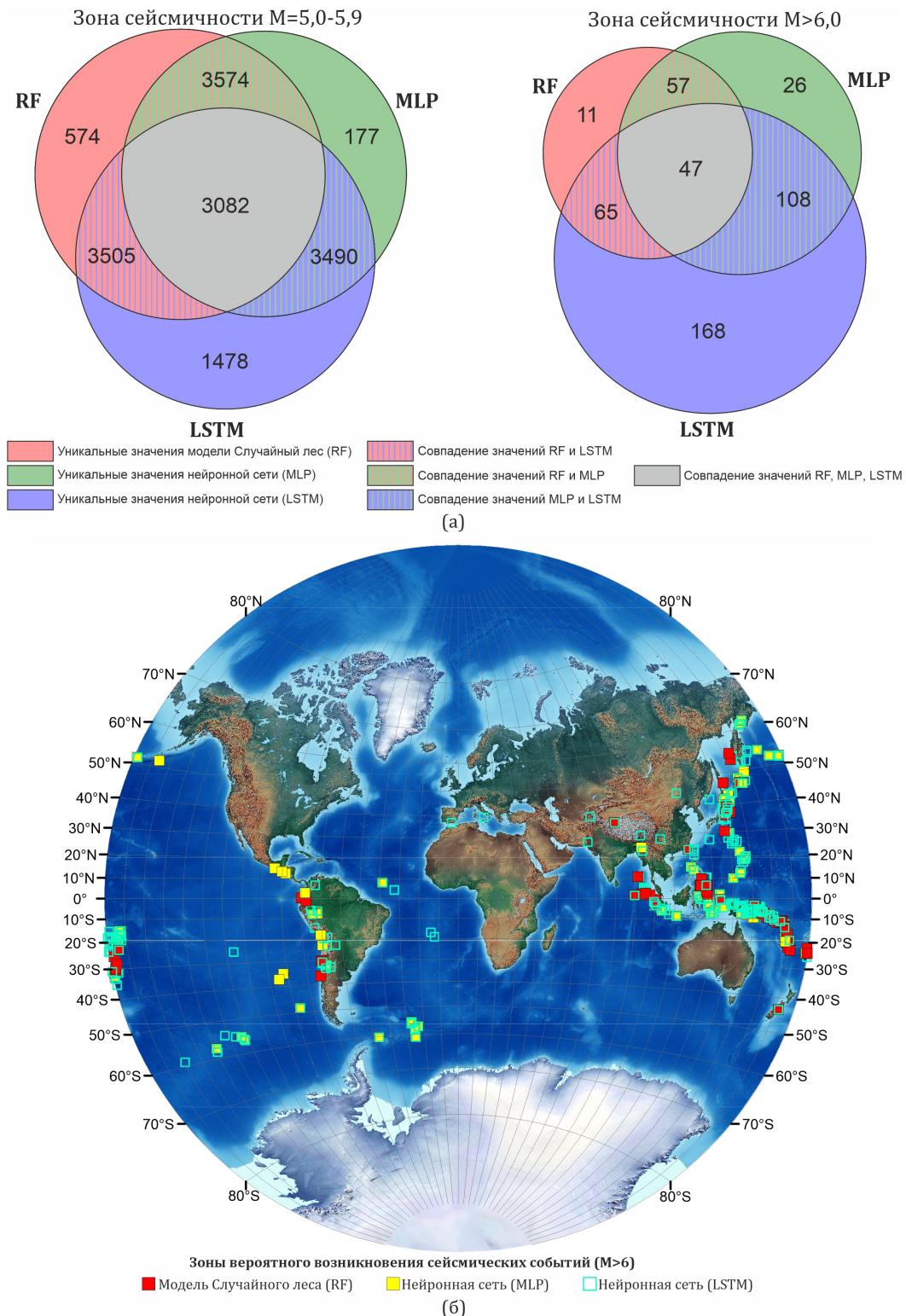


**Рис. 4.** Предсказанные зоны вероятного возникновения сейсмических событий на примере территории Российской Федерации: а) моделью случного леса (RF); б) нейросетью Многослойный перцептрон (MLP); в) нейросетью Долгая краткосрочная память (LSTM).

росетей MLP, и LSTM равное 3082. А также было выявлено количество совпадающих предсказанных зон RF с MLP (3574), MLP с LSTM (3490) и LSTM с RF (3505).

Анализ правой части [рис. 5а](#) показал, что в области сейсмичности с магнитудами  $M \geq 6$  количество совпадающих зон, предсказанных моделью RF и нейросетями MLP, LSTM составило 47. Максимальное количество совпадающих предсказанных зон было обнаружено с помощью нейросетей MLP и LSTM (108). Самое большое количество уникальных зон в области сейсмичности с магнитудами  $M \geq 6$  было предсказано нейросетью LSTM (168).

На [рис. 5б](#) представлено полученное в настоящей работе пространственное распределение зон максимальной сейсмичности ( $M \geq 6$ ), предсказанных улучшенной моделью RF и нейросетями MLP, LSTM на всем земном шаре. В результате анализа полученных данных можно сделать вывод о том, что и улучшенная модель RF, а также выбранные нейросети глубокого обучения MLP и LSTM хорошо предсказывают зоны сейсмической активности. Наибольшее количество зон максимальной сейсмичности ( $M \geq 6$ ) выявлено с использованием LSTM (см. [рис. 5б](#)).



**Рис. 5.** Результаты предсказания зон вероятного возникновения сильных сейсмических событий в мире: а) Сопоставление значений магнитуд землетрясений, предсказанных моделью RF, а также нейросетями MLP и LSTM; б) Пространственное распределение предсказанных зон максимальной сейсмичности ( $M \ge 6$ ).

Благодаря своей архитектуре многослойности и наличию активационных функций и способности изучать долгосрочные зависимости в данных, результаты, полученные с использованием нейросетей глубокого обучения MLP и LSTM, позволили выделить большее (приблизительно на 40%) количество зон сейсмичности с магнитудами  $M \ge 6$ .

Для дальнейшего повышения качества моделей МО и нейронных сетей целесообразно осуществлять совместный анализ данных, полученных из различных источников, осуществлять оценку информативности признаков, а также вводить дополнительные параметры для обучения, такие, например, как расположение очагов землетрясений относительно активных тектонических разломов и др.

#### 4. Заключение

В настоящей работе с использованием моделей машинного обучения и нейросетей глубокого обучения проведен анализ плотностей распределения землетрясений в сейсмоопасных регионах всего земного шара за 75-летний период (с 1949 по 2024 гг.), по результатам которого были выявлены зоны вероятного возникновения сильных сейсмических событий.

С использованием архивов Службы срочных донесений Геофизической службы РАН (СЕМЕ) и Геофизической Службы США (USGS) был сформирован массив исходных данных о свершившихся сейсмических событиях и проведена их предварительная обработка, включающая определение кодировки файлов о произошедших землетрясениях, загрузка и фильтрация данных по заданным критериям.

Для определения зон вероятного возникновения сейсмических событий был проведен отбор из девяти моделей машинного обучения, основанных на различных методах регрессионного анализа. Для выбора лучшей модели при решении этой задачи были исследованы линейные модели, такие как линейная регрессия (Linear Regression), лассо-регрессия (Lasso Regression) и ридж-регрессия (Ridge Regression). Кроме того, были изучены нелинейные модели: случайный лес (Random Forest), градиентный бустинг (Gradient Boosting),  $k$ -ближайших соседей ( $k$ -Nearest Neighbors), дерево решений (Decision Tree), градиентная бустинговая сборка (Light Gradient Boosting Machine, LightGBM), а также метод опорных векторов (Support Vector Machine, SVR). По результатам оценки производительности была выбрана модель случайного леса, показавшая наилучшие результаты.

Проведено улучшение обучения выбранной модели случайного леса, путем оптимизации гиперпараметров с использованием метода случайного поиска, а также с введением дополнительных параметров, таких как кластеризация и полярные координаты. В процессе улучшения модели случайного леса были предсказаны зоны максимальной сейсмичности с использованием данных за период с 1949 по 2020 гг. Предсказанные зоны были протестираны на магнитудах землетрясений, свершившихся с 2021 по 2024 гг. Рассчитаны вероятности попадания тестовых магнитуд в предсказанную сейсмическую зону, которые составили: для зоны с магнитудами  $M = 1,9\text{--}3,9$  – 86,42%; для зоны с магнитудами  $M = 4,0\text{--}4,9$  – 82,14%; для зоны с магнитудами  $M = 5,0\text{--}5,9$  – 59,40%; для зоны с магнитудами  $M \geq 6$  – 26,95%. Применение улучшенной модели случайного леса, обученной на наборе данных о землетрясениях, происходивших с 1949 по 2024 гг., позволило повысить значения метрик качества обучения до значений  $MSE = 0,104$ ;  $MAE = 0,235$ ;  $R^2 = 0,863$ , а также предсказать зоны вероятного возникновения сейсмических событий для территории земного шара. Таким образом, анализ проведенных исследований на тестовой выборке позволил сделать вывод, что улучшенная модель случайного леса хорошо предсказывает магнитуды землетрясений.

Для предсказания зон вероятного возникновения сейсмических событий кроме модели случайного леса (RF) были использованы также нейросети глубокого обучения многослойный перцептрон (MLP) и долгая краткосрочная память (LSTM). Полученные метрики качества обучения этих нейросетей ( $R^2$  составил 0,819 и 0,797) были незначительно ниже по сравнению с улучшенной моделью случайного леса ( $R^2 = 0,863$ ). Однако, благодаря многослойности и способности изучать долгосрочные зависимости в данных, результаты, полученные с использованием нейросетей MLP и LSTM, позволили выделить большее (приблизительно на 40%) количество зон сейсмичности с магнитудами  $M \geq 6$  по сравнению с улучшенной моделью RF.

Результаты проведенных исследований продемонстрировали эффективность применения моделей машинного обучения и нейросетей глубокого обучения для выявления зон максимального риска возникновения значительных сейсмических событий в сейсмоопасных регионах земного шара.

В процессе подготовки настоящей работы у берегов Камчатского п-ова 30 июля 2025 г. произошло землетрясение с магнитудой  $M = 8,8$ . Эпицентр этого события находился на расстоянии около 260 км от предсказанной нейросетью LSTM зоны вероятного возникновения сейсмических событий с магнитудой более 7.

Полученные результаты могут использоваться Министерством Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий и его подведомственными организациями, а также Федеральным государственным бюджетным учреждением науки Федеральный исследовательский центр «Единая геофизическая служба Российской академии наук» (ФИЦ ЕГС РАН) и его региональными Сейсмологическими центрами, а также другими ведомствами и организациями, занимающимися мониторингом сейсмоопасных территорий.

**Благодарности.** Работа выполнена в НИИ «АЭРОКОСМОС» в рамках Государственного задания.

### Список литературы

- Болт Б. Землетрясения. Общедоступный очерк. — М. : Мир, 1981. — 256 с.
- Бондур В. Г. и Воронова О. С. Регистрация из космоса аномальных вариаций тепловых полей при сейсмических событиях на территории Северного Кавказа с 2017 по 2022 гг. // Исследование Земли из космоса. — 2022. — № 6. — С. 13–26. — <https://doi.org/10.31857/S0205961422060021>.
- Бондур В. Г., Гарагаш И. А., Гохберг М. Б. и др. Геомеханические модели и ионосферные вариации для крупнейших землетрясений при слабом воздействии градиентов атмосферного давления // Доклады академии наук. — 2007. — Т. 414, № 4. — С. 540–543.
- Бондур В. Г., Гарагаш И. А., Гохберг М. Б. и др. Связь между вариациями напряженно-деформированного состояния земной коры и сейсмической активностью на примере Южной Калифорнии // Доклады академии наук. — 2010. — Т. 430, № 3. — С. 400–404.
- Бондур В. Г., Гарагаш И. А., Гохберг М. Б. и др. Эволюция напряженного состояния Южной Калифорнии на основе геомеханической модели и текущей сейсмичности // Физика Земли. — 2016. — № 1. — С. 120–132. — <https://doi.org/10.7868/S000233371601004X>.
- Бондур В. Г., Цидилина М. Н., Гапонова Е. В. и др. Регистрация из космоса аномалий различных геофизических полей при подготовке разрушительных землетрясений в Турции в феврале 2023 г. // Исследование Земли из космоса. — 2023. — № 4. — С. 3–25. — <https://doi.org/10.31857/S0205961423340018>.
- Бондур В. Г., Цидилина М. Н., Гапонова Е. В. и др. Совместный анализ аномальных вариаций различных геофизических полей по космическим данным при подготовке землетрясения в районе оз. Байкал 22 сентября 2020 г. ( $M=5.6$ ) // Исследование Земли из космоса. — 2022. — № 5. — С. 3–19. — <https://doi.org/10.31857/S0205961422050049>.
- Викулин А. В., Акманова Д. Р., Осипова Н. А. и др. Повторяемость сильных землетрясений и миграции их очагов вдоль сейсмического пояса // Вестник Камчатского Государственного Технического Университета. — 2009. — № 10. — С. 17–25.
- Гапонова Е. В., Зверев А. Т. и Цидилина М. Н. Выявление аномалий линеаментных систем по космическим изображениям во время сильных землетрясений в Калифорнии с магнитудами 6.4 и 7.1 // Исследование Земли из космоса. — 2019. — № 6. — С. 36–47. — <https://doi.org/10.31857/S0205-96142019636-47>.
- Гвишиани А. Д., Соловьев А. А. и Дзебоев Б. А. Проблема распознавания мест возможного возникновения сильных землетрясений: актуальный обзор // Физика Земли. — 2020. — № 1. — С. 5–29. — <https://doi.org/10.31857/S0002333720010044>.
- Гельфанд И. М., Губерман Ш. А., Извекова М. Л. и др. О критериях высокой сейсмичности // Доклады Академии наук СССР. — 1972. — Т. 202, № 6. — С. 1317–1320.
- Кашкин В. Б., Романов А. А., Григорьев А. С. и др. Тропосферные эффекты землетрясений в Туве, наблюдаемые с искусственных спутников Земли // Журнал СФУ. Техника и технология. — 2012. — Т. 5, № 2. — С. 220–228.
- Кейлис-Борок В. И. Динамика литосферы и прогноз землетрясений // Природа. — 1989. — № 12. — С. 10–18.

- Кособоков В. Г. Прогноз землетрясений и геодинамические процессы. Часть 1. Прогноз землетрясений: основы, реализация, перспективы. (Вычислительная сейсмология; Вып. 36). — М. : ГЕОС, 2005. — 179 с.
- Липеровский В. А., Похотов О. А., Мейстер К. В. и др. Физические модели связей в системе литосфера-атмосфера-ионосфера перед землетрясениями // Геомагнетизм и аэрономия. — 2008. — Т. 48, № 6. — С. 831–843.
- Моги К. Предсказание землетрясений. — М. : Мир, 1988. — 382 с.
- Осипов В. И., Шойгу С. К. и Соболев Г. А. Природные опасности России. Сейсмические опасности. — М. : КРУК, 2000. — 296 с.
- Пущаровский Ю. М. Введение в тектонику Тихоокеанского сегмента Земли. Труды ГИН, вып. 234. — М. : Наука, 1972. — 228 с.
- Пущаровский Ю. М. Тектонические провинции Атлантического океана // Геотектоника. — 2009. — № 3. — С. 3–13.
- Смирнов В. М., Смирнова Е. В., Цидилина М. Н. и др. Сейсмоионосферные вариации во время сильных землетрясений на примере землетрясения 2010 г. в Чили // Космические исследования. — 2018. — Т. 56, № 4. — С. 283–292. — <https://doi.org/10.31857/S002342060000347-9>.
- Соболев Г. А. и Пономарев А. В. Физика землетрясений и предвестники. — М. : Наука, 2003. — 270 с.
- Уломов В. И. Сейсмичность территории России // Изменение окружающей среды и климата. Природные и связанные с ними техногенные катастрофы. Т. 1. — М. : ИФЗ РАН, 2008. — С. 13–18.
- Федотов С. А. Долгосрочный сейсмический прогноз для Курило-Камчатской дуги. — М. : Наука, 2005. — 302 с.
- Хайн В. Е. и Халилов Э. Н. Цикличность геодинамических процессов: ее возможная природа. — М. : Научный мир, 2009. — 520 с.
- Хайкин С. Нейронные сети: полный курс. Второе издание. — 2-е изд. — М. : Вильямс, 2006. — 1104 с.
- Шерман С. И. и Злогодухова О. Г. Сейсмические пояса и зоны Земли: формализация понятий, положение в литосфере и структурный контроль // Геодинамика и тектонофизика. — 2011. — Т. 2, № 1. — С. 1–34. — <https://doi.org/10.5800/GT-2011-2-1-0031>.
- Agarwal N., Arora I., Saini H., et al. A Novel Approach for Earthquake Prediction Using Random Forest and Neural Networks // EAI Endorsed Transactions on Energy Web. — 2023. — Vol. 10. — P. 1–6. — <https://doi.org/10.4108/ew.4329>.
- Ajai V., Gandhimathi U. S., Suntosh B. D. S., et al. Machine Learning-Based Seismic Activity Prediction // Utilizing AI and Machine Learning for Natural Disaster Management. — IGI Global, 2024. — P. 293–306. — <https://doi.org/10.4018/979-8-3693-3362-4.ch017>.
- Akhoondzadeh M. and Marchetti D. Study of the Preparation Phase of Turkey's Powerful Earthquake (6 February 2023) by a Geophysical Multi-Parametric Fuzzy Inference System // Remote Sensing. — 2023. — Vol. 15, no. 9. — P. 2224. — <https://doi.org/10.3390/rs15092224>.
- Asim K. M., Idris A., Iqbal T., et al. Earthquake prediction model using support vector regressor and hybrid neural networks // PLOS ONE. — 2018. — Vol. 13, no. 7. — e0199004. — <https://doi.org/10.1371/journal.pone.0199004>.
- Bogdanov V., Gavrilov V., Pulinets S., et al. Responses to the preparation of strong Kamchatka earthquakes in the lithosphere-atmosphere-ionosphere system, based on new data from integrated ground and ionospheric monitoring // E3S Web of Conferences. — 2020. — Vol. 196. — P. 14. — <https://doi.org/10.1051/e3sconf/202019603005>.
- Ermiş İ. and Cürebal İ. Earthquake Probability Prediction with Decision Tree Algorithm: The Example of Izmir, Türkiye // Journal of Artificial Intelligence and Data Science. — 2024. — Vol. 4, no. 2. — P. 59–67.
- Frankel A. D., Petersen M. D., Mueller C. S., et al. Documentation for the 2002 Update of the National Seismic Hazard Maps. Open-File Report 02-420. — USGS, 2002. — 33 p.
- Jarah N. B., Alasadi A. H. H. and Hashim K. M. A New Algorithm for Earthquake Prediction Using Machine Learning Methods // Journal of Computer Science. — 2024. — Vol. 20, no. 2. — P. 150–156. — <https://doi.org/10.3844/jcssp.2024.150.156>.
- Jiao Q., Liu Q., Lin C., et al. Spatiotemporal Analysis of Atmospheric Chemical Potential Anomalies Associated with Major Seismic Events ( $M_s \geq 7$ ) in Western China: A Multi-Case Study // Remote Sensing. — 2025. — Vol. 17, no. 2. — P. 311. — <https://doi.org/10.3390/rs17020311>.
- Liu C. and Macedo J. Machine learning-based models for estimating seismically-induced slope displacements in subduction earthquake zones // Soil Dynamics and Earthquake Engineering. — 2022. — Vol. 160. — P. 107323. — <https://doi.org/10.1016/j.soildyn.2022.107323>.
- Mahmoudi J., Arjomand M. A., Rezaei M., et al. Predicting the Earthquake Magnitude Using the Multilayer Perceptron Neural Network with Two Hidden Layers // Civil Engineering Journal. — 2016. — Vol. 2, no. 1. — P. 1–12.

- Mallouhy R., Jaoude C. A., Guyeux C., et al. Major earthquake event prediction using various machine learning algorithms // 2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM). — IEEE, 2019. — P. 1–7. — <https://doi.org/10.1109/ict-dm47966.2019.9032983>.
- Mignan A., Rinaldi A. P., Lanza F., et al. A Multi-LASSO model to forecast induced seismicity at enhanced geothermal systems // *Geoenergy Science and Engineering*. — 2024. — Vol. 236. — P. 212746. — <https://doi.org/10.1016/j.geoen.2024.212746>.
- Novianti P., Setyorini D. and Rafflesia U. K-Means cluster analysis in earthquake epicenter clustering // *International Journal of Advances in Intelligent Informatics*. — 2017. — Vol. 3, no. 2. — P. 81–89. — <https://doi.org/10.26555/ijain.v3i2.100>.
- Piscini A., Santis A. De, Marchetti D., et al. A Multi-parametric Climatological Approach to Study the 2016 Amatrice-Norcia (Central Italy) Earthquake Preparatory Phase // *Pure and Applied Geophysics*. — 2017. — Vol. 174, no. 10. — P. 3673–3688. — <https://doi.org/10.1007/s00024-017-1597-8>.
- Ranjan G. S. K., Verma A. K. and Radhika S. K-Nearest Neighbors and Grid Search CV Based Real Time Fault Monitoring System for Industries // 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). — IEEE, 2019. — P. 1–5. — <https://doi.org/10.1109/i2ct45611.2019.9033691>.
- Saito A., Tsugawa T., Otsuka Y., et al. Acoustic resonance and plasma depletion detected by GPS total electron content observation after the 2011 off the Pacific coast of Tohoku Earthquake // *Earth, Planets and Space*. — 2011. — Vol. 63, no. 7. — P. 863–867. — <https://doi.org/10.5047/eps.2011.06.034>.
- Shukla S. S., Dhanya J., Kumar P., et al. An Ensemble Random Forest Model for Seismic Energy Forecast // *Natural Hazards and Earth System Sciences Discussion [preprint]*. — 2024. — P. 40. — <https://doi.org/10.5194/nhess-2024-129>.
- Vardaan V. K., Bhandarkar T., Satish N., et al. Earthquake trend prediction using long short-term memory RNN // *International Journal of Electrical and Computer Engineering (IJECE)*. — 2019. — Vol. 9, no. 2. — P. 1304–1312. — <https://doi.org/10.11591/ijece.v9i2.pp1304-1312>.
- Wei M. and Gao K. Machine Learning Predicts the Slip Duration and Friction Drop of Laboratory Earthquakes in Sheared Granular Fault // *Journal of Geophysical Research: Machine Learning and Computation*. — 2024. — Vol. 1, no. 4. — <https://doi.org/10.1029/2024JH000398>.

# IDENTIFICATION OF AREAS OF PROBABLE SEISMIC EVENTS USING MACHINE LEARNING

V. G. Bondur<sup>1\*\*</sup>  and O. S. Voronova<sup>1</sup> 

<sup>1</sup>Institute for Scientific Research of Aerospace Monitoring “AEROCOSMOS”, Moscow,

\*\*Correspondence to: Valery G. Bondur, office@aerocosmos.info

Machine learning models based on regression analysis methods have been selected to identify areas of maximum risk for major seismic events. The performance of nine linear and nonlinear models was evaluated, resulting in the selection of the Random Forest model. The quality of training for the Random Forest model was improved through hyperparameter tuning as well as the use of clustering and polar coordinates. It allowed the improvement of quality of model training, increasing the value of the coefficient of determination to 0.86. An analysis was conducted on the applicability of two neural networks with deep learning: Multi-layer Perceptron (MLP) and Long Short-Term Memory (LSTM), using training parameters that were selected for the Random Forest model. Using this model and selected neural networks with deep learning, areas of maximum risk for seismic events were predicted for the entire globe, with a detailed analysis of predicted areas for the territory of the Russian Federation. As a result of the conducted research, the use of neural networks with deep learning made it possible to identify a greater (~40%) number of zones of maximum seismicity (with  $M > 6$ ) compared to the improved Random Forest model.

Received: May 26, 2025

Accepted: September 2, 2025

Published: September 8, 2025



© 2025. The Authors.

**Citation:** Bondur V. G., Voronova O. S. (2025), Identification of Areas of Probable Seismic Events Using Machine Learning, *Russian Journal of Earth Sciences*, 25, ES5003, <https://doi.org/10.2205/2025ES001055>, EDN: IOGILW

## References

- Agarwal N., Arora I., Saini H., et al. A Novel Approach for Earthquake Prediction Using Random Forest and Neural Networks // EAI Endorsed Transactions on Energy Web. — 2023. — Vol. 10. — P. 1–6. — <https://doi.org/10.4108/ew.4329>.
- Ajai V., Gandhimathi U. S., Suntosh B. D. S., et al. Machine Learning-Based Seismic Activity Prediction // Utilizing AI and Machine Learning for Natural Disaster Management. — IGI Global, 2024. — P. 293–306. — <https://doi.org/10.4018/979-8-3693-3362-4.ch017>.
- Akhoondzadeh M. and Marchetti D. Study of the Preparation Phase of Turkey’s Powerful Earthquake (6 February 2023) by a Geophysical Multi-Parametric Fuzzy Inference System // Remote Sensing. — 2023. — Vol. 15, no. 9. — P. 2224. — <https://doi.org/10.3390/rs15092224>.
- Asim K. M., Idris A., Iqbal T., et al. Earthquake prediction model using support vector regressor and hybrid neural networks // PLOS ONE. — 2018. — Vol. 13, no. 7. — e0199004. — <https://doi.org/10.1371/journal.pone.0199004>.
- Bogdanov V., Gavrilov V., Pulinets S., et al. Responses to the preparation of strong Kamchatka earthquakes in the lithosphere-atmosphere-ionosphere system, based on new data from integrated ground and ionospheric monitoring // E3S Web of Conferences. — 2020. — Vol. 196. — P. 14. — <https://doi.org/10.1051/e3sconf/202019603005>.
- Bolt B. Earthquakes. A Popular Outline. — M. : Mir, 1981. — 256 p. — (In Russian).
- Bondur V. G., Garagash I. A., Gokhberg M. B., et al. Geomechanical models and ionospheric variations related to strongest earthquakes and weak influence of atmospheric pressure gradients // Doklady Earth Sciences. — 2007. — Vol. 414, no. 1. — P. 666–669. — <https://doi.org/10.1134/S1028334X07040381>.
- Bondur V. G., Garagash I. A., Gokhberg M. B., et al. Connection between variations of the stress-strain state of the Earth’s crust and seismic activity: The example of Southern California // Doklady Earth Sciences. — 2010. — Vol. 430, no. 1. — P. 147–150. — <https://doi.org/10.1134/S1028334X10010320>.

- Bondur V. G., Garagash I. A., Gokhberg M. B., et al. The evolution of the stress state in Southern California based on the geomechanical model and current seismicity // *Izvestiya, Physics of the Solid Earth*. — 2016. — Vol. 52, no. 1. — P. 117–128. — <https://doi.org/10.1134/S1069351316010043>.
- Bondur V. G., Tsidilina M. N., Gaponova E. V., et al. Satellite Registration of Anomalies of Various Geophysical Fields during the Preparation of Destructive Earthquakes in Turkey in February 2023 // *Izvestiya, Atmospheric and Oceanic Physics*. — 2023. — Vol. 59, no. 9. — P. 1009–1027. — <https://doi.org/10.1134/s0001433823090049>.
- Bondur V. G., Tsidilina M. N., Gaponova E. V., et al. Combined Analysis of Anomalous Variations in Various Geophysical Fields during Preparation of the M5.6 Earthquake near Lake Baikal on September 22, 2020, Based on Satellite Data // *Izvestiya, Atmospheric and Oceanic Physics*. — 2022. — Vol. 58, no. 12. — P. 1532–1545. — <https://doi.org/10.1134/S0001433822120052>.
- Bondur V. G. and Voronova O. S. Detection from Space of Anomalous Variations in Thermal Fields during Seismic Events in the Northern Caucasus in 2017–2022 // *Izvestiya, Atmospheric and Oceanic Physics*. — 2022. — Vol. 58, no. 12. — P. 1546–1556. — <https://doi.org/10.1134/S0001433822120064>.
- Ermiş İ. and Cürebal İ. Earthquake Probability Prediction with Decision Tree Algorithm: The Example of Izmir, Türkiye // *Journal of Artificial Intelligence and Data Science*. — 2024. — Vol. 4, no. 2. — P. 59–67.
- Fedotov S. A. Long-Term Seismic Forecast for the Kuril-Kamchatka Arc. — M. : Nauka, 2005. — 302 p. — (In Russian).
- Frankel A. D., Petersen M. D., Mueller C. S., et al. Documentation for the 2002 Update of the National Seismic Hazard Maps. Open-File Report 02-420. — USGS, 2002. — 33 p.
- Gaponova E. V., Zverev A. T. and Tsidilina M. N. Detecting Lineament System Anomalies during Strong 6.4 and 7.1 Earthquakes in California from Satellite Imagery // *Izvestiya, Atmospheric and Oceanic Physics*. — 2020. — Vol. 56, no. 9. — P. 1062–1071. — <https://doi.org/10.1134/S000143382009011X>.
- Gelfand I. M., Guberman Sh. A., Izvekova M. L., et al. On Criteria of High Seismicity // *Doklady Akademii Nauk SSSR*. — 1972. — Vol. 202, no. 6. — P. 1317–1320. — (In Russian).
- Gvishiani A. D., Soloviev A. A. and Dzeboev B. A. Problem of Recognition of Strong-Earthquake-Prone Areas: a State-of-the-Art Review // *Izvestiya, Physics of the Solid Earth*. — 2020. — Vol. 56, no. 1. — P. 1–23. — <https://doi.org/10.1134/s1069351320010048>.
- Haykin S. *Neural Networks: A Comprehensive Foundation*. Second Edition. — 2nd ed. — New Jersey : Prentice Hall, 2006. — 1104 p.
- Jarah N. B., Alasadi A. H. H. and Hashim K. M. A New Algorithm for Earthquake Prediction Using Machine Learning Methods // *Journal of Computer Science*. — 2024. — Vol. 20, no. 2. — P. 150–156. — <https://doi.org/10.3844/jcssp.2024.150.156>.
- Jiao Q., Liu Q., Lin C., et al. Spatiotemporal Analysis of Atmospheric Chemical Potential Anomalies Associated with Major Seismic Events ( $M_s \geq 7$ ) in Western China: A Multi-Case Study // *Remote Sensing*. — 2025. — Vol. 17, no. 2. — P. 311. — <https://doi.org/10.3390/rs17020311>.
- Kashkin V. B., Romanov A. A., Grigoriev A. S., et al. Troposphere Effects of Tuva Earthquakes Detected with Spase Technology // *Journal of Siberian Federal University. Engineering & Technologies*. — 2012. — Vol. 5, no. 2. — P. 220–228. — (In Russian).
- Keilis-Borok V. I. *Lithosphere Dynamics and Earthquake Prediction* // *Priroda*. — 1989. — No. 12. — P. 10–18. — (In Russian).
- Khain V. E. and Khalilov E. N. *Cycles in Geodynamic Processes: Their Possible Nature*. — M. : Nauchnyi Mir, 2009. — 520 p. — (In Russian).
- Kosobokov V. G. *Earthquake Prediction and Geodynamic Processes. Part 1. Earthquake Prediction: Fundamentals, Implementation, Prospects. (Computational Seismology)*; Vol. 36). — M. : GEOS, 2005. — 179 p. — (In Russian).
- Liperovsky V. A., Pokhotelov O. A., Meister K. V., et al. Physical models of coupling in the lithosphere-atmosphere-ionosphere system before earthquakes // *Geomagnetism and Aeronomy*. — 2008. — Vol. 48, no. 6. — P. 795–806. — <https://doi.org/10.1134/s0016793208060133>. — (In Russian).
- Liu C. and Macedo J. Machine learning-based models for estimating seismically-induced slope displacements in subduction earthquake zones // *Soil Dynamics and Earthquake Engineering*. — 2022. — Vol. 160. — P. 107323. — <https://doi.org/10.1016/j.soildyn.2022.107323>.
- Mahmoudi J., Arjomand M. A., Rezaei M., et al. Predicting the Earthquake Magnitude Using the Multilayer Perceptron Neural Network with Two Hidden Layers // *Civil Engineering Journal*. — 2016. — Vol. 2, no. 1. — P. 1–12.
- Mallouhy R., Jaoude C. A., Guyeux C., et al. Major earthquake event prediction using various machine learning algorithms // *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*. — IEEE, 2019. — P. 1–7. — <https://doi.org/10.1109/ict-dm47966.2019.9032983>.

- Mignan A., Rinaldi A. P., Lanza F., et al. A Multi-LASSO model to forecast induced seismicity at enhanced geothermal systems // *Geoenergy Science and Engineering*. — 2024. — Vol. 236. — P. 212746. — <https://doi.org/10.1016/j.geoen.2024.212746>.
- Mogi K. *Earthquake Prediction*. — Tokyo : Academic Press, 1985. — 382 p.
- Novianti P., Setyorini D. and Rafflesia U. K-Means cluster analysis in earthquake epicenter clustering // *International Journal of Advances in Intelligent Informatics*. — 2017. — Vol. 3, no. 2. — P. 81–89. — <https://doi.org/10.26555/ijain.v3i2.100>.
- Osipov V. I., Shoigu S. K. and Sobolev G. A. *Natural Hazards of Russia. Seismic Hazards*. — M. : KRUK, 2000. — 296 p. — (In Russian).
- Piscini A., Santis A. De, Marchetti D., et al. A Multi-parametric Climatological Approach to Study the 2016 Amatrice-Norcia (Central Italy) Earthquake Preparatory Phase // *Pure and Applied Geophysics*. — 2017. — Vol. 174, no. 10. — P. 3673–3688. — <https://doi.org/10.1007/s00024-017-1597-8>.
- Pushcharovsky Yu. M. *Introduction into Tectonics of the Pacific Segment of the Earth. Transactions*, vol. 234. — M. : Nauka, 1972. — 228 p. — (In Russian).
- Pushcharovsky Yu. M. *Tectonic Provinces of the Atlantic Ocean* // *Geotectonics*. — 2009. — No. 3. — P. 3–13. — (In Russian).
- Ranjan G. S. K., Verma A. K. and Radhika S. K-Nearest Neighbors and Grid Search CV Based Real Time Fault Monitoring System for Industries // 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). — IEEE, 2019. — P. 1–5. — <https://doi.org/10.1109/i2ct45611.2019.9033691>.
- Saito A., Tsugawa T., Otsuka Y., et al. Acoustic resonance and plasma depletion detected by GPS total electron content observation after the 2011 off the Pacific coast of Tohoku Earthquake // *Earth, Planets and Space*. — 2011. — Vol. 63, no. 7. — P. 863–867. — <https://doi.org/10.5047/eps.2011.06.034>.
- Sherman S. I. and Zlogodukhova O. G. Seismic Belts and Zones of the Earth: Formalization of Notions, Positions in the Lithosphere, and Structural Control // *Geodynamics & Tectonophysics*. — 2011. — Vol. 2, no. 1. — P. 1–34. — <https://doi.org/10.5800/gt-2011-2-1-0031>. — (In Russian).
- Shukla S. S., Dhanya J., Kumar P., et al. An Ensemble Random Forest Model for Seismic Energy Forecast // *Natural Hazards and Earth System Sciences Discussion* [preprint]. — 2024. — P. 40. — <https://doi.org/10.5194/nhess-2024-129>.
- Smirnov V. M., Smirnova E. V., Tsidilina M. N., et al. Seismoinospheric variations during strong earthquakes on the example of the earthquake of 2010 in Chile // *Kosmicheskie issledovaniia*. — 2018. — Vol. 56, no. 4. — P. 283–292. — <https://doi.org/10.31857/s002342060000347-9>. — (In Russian).
- Sobolev G. A. and Ponomarev A. V. *Physics of Earthquakes and Precursors*. — M. : Nauka, 2003. — 270 p. — (In Russian).
- Ulomov V. I. *Seismicity of the Russian Territory* // *Environmental and Climate Change. Natural and Related Technogenic Catastrophes*. Vol. 1. — M. : IFZ RAN, 2008. — P. 13–18. — (In Russian).
- Vardaan V. K., Bhandarkar T., Satish N., et al. Earthquake trend prediction using long short-term memory RNN // *International Journal of Electrical and Computer Engineering (IJECE)*. — 2019. — Vol. 9, no. 2. — P. 1304–1312. — <https://doi.org/10.11591/ijece.v9i2.pp1304-1312>.
- Vikulin A. V., Akmanova D. R., Osipova N. A., et al. Recurrence of Strong Earthquakes and Migration of Their Sources along the Seismic Belt // *Vestnik of Kamchatka State Technical University*. — 2009. — No. 10. — P. 17–25. — (In Russian).
- Wei M. and Gao K. Machine Learning Predicts the Slip Duration and Friction Drop of Laboratory Earthquakes in Sheared Granular Fault // *Journal of Geophysical Research: Machine Learning and Computation*. — 2024. — Vol. 1, no. 4. — <https://doi.org/10.1029/2024JH000398>.