

DOI: <https://doi.org/10.37162/2618-9631-2025-3-77-91>

УДК 551.509.3+004.032.26

Применение измерений напряженности электрического поля атмосферы и методов машинного обучения в прогнозировании суточных сумм осадков в летние месяцы в городе Нальчик

А.М. Беккиев, И.Х. Машуков, В.А. Шаповалов

*Высокогорный геофизический институт,
Кабардино-Балкарская Республика, г. Нальчик, Россия
vet555_83@mail.ru*

Рассматривается актуальная задача краткосрочного прогнозирования суточных атмосферных осадков с использованием метеорологической информации и данных, полученных на основе измерений напряженности электрического поля атмосферы, но без привлечения данных о прошлых значениях осадков. Проведенные исследования показали высокую эффективность применения искусственного интеллекта в решении поставленной задачи, в частности, таких методов машинного обучения, как модели градиентного бустинга, деревья решений, нейронные сети. Данные для исследования за период с 2020 по 2025 год получены с метеостанции Нальчик (Россия, WMO ID 37212) и измерителя напряженности электрического поля EFM-100 установленного на крыше здания ФГБУ «ВГИ».

Ключевые слова: прогноз атмосферных осадков, машинное обучение, градиентный бустинг, деревья решений, нейронные сети, временные ряды, метеорологические данные, спектральный анализ, вейвлет-анализ

Application of atmospheric electric field strength measurements and machine learning methods in predicting daily precipitation in the summer months: A case study for the city of Nalchik

A.M. Bekkiev, I.K. Mashukov, V.A. Shapovalov

*High-Mountain Geophysical Institute,
Kabardino-Balkarian Republic, Nalchik, Russia
vet555_83@mail.ru*

The paper considers a relevant problem of short-term forecasting of daily precipitation using meteorological information and data obtained from measurements of the electric field strength of the atmosphere, but without involving data on past precipitation values. The studies showed high efficiency of applying artificial intelligence in solving the problem, in particular, machine learning methods such as gradient boosting models, decision trees, and neural networks. The data for the study over the period from 2020 to 2025 were obtained from the Nalchik weather station (Russia, WMO ID 37212) and the EFM-100 electric field strength meter installed on the roof of the High-Mountain Geophysical Institute building.

Keywords: precipitation forecast, machine learning, gradient boosting, decision trees, neural networks, time series, meteorological data, spectral analysis, wavelet analysis

Введение

Прогнозирование атмосферных осадков критически важно для различных отраслей экономики, таких как авиация, сельское хозяйство, городская инфраструктура [9], а также для безопасности жизнедеятельности. Можно утверждать, что прогноз осадков является одной из ключевых задач метеорологии, имеющей важное значение для управления водными ресурсами [2, 7], например, при прогнозировании стока горных рек, предупреждении и защите от стихийных бедствий, вызванных гидрологическими опасными явлениями и их последствиями [4, 8]), а также для системы мониторинга и прогнозирования лесных пожаров [10]. Согласно последнему докладу Межправительственной группы экспертов по изменению климата (IPCC), наблюдается увеличение интенсивности и частоты экстремальных осадков, что связано с повышением температуры и увеличением содержания влаги в атмосфере. Это приводит к более частым и интенсивным ливням, наводнениям и другим опасным погодным явлениям, что требует более точной и своевременной информации о возможности их возникновения.

Традиционные методы прогнозирования осадков – это статистические модели и численные модели погоды, основанные на математических моделях атмосферы, радиолокационных и спутниковых данных, они требуют значительных вычислительных ресурсов и сталкиваются с трудностями в точном прогнозировании осадков из-за сложности атмосферных процессов и ограничений в разрешающей способности моделей [11, 18]. Кроме того, недостаточная плотность наблюдательной сети, особенно в развивающихся странах, таких как страны Африки, усугубляет проблему. Современные подходы, использующие методы машинного обучения, позволяют обрабатывать большие объемы данных и выявлять сложные зависимости между метеорологическими параметрами и осадками [15, 19]. Например, глубокие нейронные сети успешно применяются для краткосрочного и среднесрочного прогнозирования, обеспечивая более высокую точность по сравнению с традиционными методами [16]. Модель DGMR от компании DeepMind, при тех же требованиях к данным, что и модель среднесрочного метеорологического прогноза от Европейского центра среднесрочных прогнозов погоды (ECMWF), превосходит существующие методы в прогнозировании осадков на ближайшие часы.

При этом выполняемые в последнее время исследования демонстрируют широкий спектр методологических подходов – от классических статистических методов и Фурье-анализа до передовых алгоритмов машинного обучения и гибридных моделей, что свидетельствует о неослабевающем спросе на такие прогнозы по различным исходным данным. Так, в [6] показано, что комбинированное применение фрактального анализа и Фурье-преобразования позволяет эффективно выделять периодические компоненты метеорологических рядов для прогнозирования температуры и осадков, что даёт прочную математическую основу для дальнейших исследований. Классические модели временных рядов, такие как ARIMA,

в сочетании с анализом трендов также успешно используются для прогноза осадков, о чем свидетельствует работа [12] по региону Вади-Шуэйб в Иордании.

Среди методов машинного обучения значимое внимание уделяется многослойным нейронным сетям, градиентному бустингу и автоэнкодерам. Например, в [21] продемонстрировано, что нейросетевые модели, обученные на многомасштабных данных, существенно повышают точность прогноза суточных осадков в тропическом климате Шри-Ланки. В [13] использовался метод градиентного бустинга XGBoost для корректировки ошибок численных моделей ECMWF, тем самым было поучено улучшение краткосрочного прогноза осадков в различных регионах Китая. Предложенный в [17] сезонно-интегрированный автоэнкодер показал высокую устойчивость при краткосрочном прогнозе ежедневных осадков за счёт нелинейного сжатия признаков. Гибридные подходы, объединяющие данные радиолокационные и дистанционного зондирования с алгоритмами машинного обучения, были успешно апробированы в [14, 20].

Несмотря на эти достижения, в литературе остаётся недостаточно проработанным использование данных измерений напряжённости электрического поля атмосферы в сочетании со стандартными метеорологическими параметрами, как то приземные температура, давление, влажность. В контексте прогнозирования селевой опасности с учетом данных о атмосферных осадках авторами [1] акцентировано внимание на фоновых методах с учётом электрического поля, показана перспективность включения электрических параметров при вероятностном прогнозировании экстремальных осадков.

Региональная специфика Северного Кавказа и прилегающих территорий требует учёта локальных особенностей рельефа и микроклимата. Так, в [3] выявлены тенденции смещения максимумов ливневой и грозовой активности на Северном Кавказе, что обусловлено сложным взаимодействием атмосферных процессов и топографических факторов. В работе [5] продемонстрировано, что нейросетевые модели с учётом региональных данных позволяют улучшить точность прогноза в Крымском регионе, что свидетельствует о важности адаптации алгоритмов к локальным условиям. Между тем для Северного Кавказа, особенно в горных и предгорных регионах, целостный подход, объединяющий методы машинного обучения и метеорологические данные с учётом уникальных электрофизических характеристик атмосферы, может иметь перспективы для прогноза осадков, включая диагностику типов осадков и количественное прогнозирование.

Таким образом, предлагаемое исследование фокусируется на разработке гибридной модели, которая объединит непрерывные измерения напряжённости электрического поля атмосферы и современные методы машинного обучения для прогноза суточных сумм осадков в летний период на примере одного наблюдательного пункта. Интеграция электрофизических данных позволит учесть интегральное влияние физических механизмов конденсации и заряженности частиц в облаках, а адаптация различных

архитектур машинного обучения под локальные климатические условия Северного Кавказа обеспечит высокую точность и надёжность прогноза. Новизна данной работы заключается в комплексном использовании данных разных типов и разработке алгоритма, способного обрабатывать разреженные и зашумленные измерения электрического поля в условиях ограниченного наблюдательного охвата.

Поэтому, развитие и внедрение методов машинного обучения и спектрального анализа в прогнозирование осадков является актуальной задачей, способствующей повышению точности прогнозов и обеспечению безопасности населения, что особенно важно в условиях изменяющегося климата.

Материалы и методы

В исследовании использованы архивные данные метеорологической станции аэропорта Нальчик (WMO_ID=37212) за период с 2020 по 2025 год. Первичный набор включает ежедневные измерения температуры воздуха, атмосферного давления, относительной влажности, скорости и направления ветра, а также суточные суммы осадков. Дополнительно привлекались данные о напряжённости атмосферного электрического поля атмосферы (НЭПА) с прибора измерителя EFM-100 установленного на крыше здания ФГБУ «ВГИ». Работа с данными и программирование моделей производились на языке высокого уровня python.

Для повышения качества модели произведено расширение исходного пространства признаков. Были сформированы лаговые признаки за один, два и, соответственно, три предыдущих дня. Выбор временных лагов 1–3 суток мотивирован физикой атмосферных процессов: облачные структуры формируются на фоне медленно эволюционирующей циклонической активности [6, 13], а электрическая нестабильность способна проявляться за несколько дней до подхода фронта. Также рассчитаны производные характеристики с использованием методов спектрального и вейвлет-анализа, позволяющих выделять частотные и локальные временные особенности сигналов. Эти признаки отражают инерционную динамику атмосферы и существенно увеличивают информативность модели при машинном обучении. То есть целевая переменная – осадки в миллиметрах – прогнозируется на основе предшествующих значений метеорологических параметров, напряжённости атмосферного электрического поля и производных характеристик: среднее, максимум, минимум, асимметрия, эксцесс, коэффициент вариации, время наступления максимума и минимума, интервал между максимальным и минимальным значением, интеграл по времени, относительный размах (размах отнесенный к среднему), полу-ширина пика, скорость роста до пика, количество пиков выше 75-го перцентиля, среднее расстояние между пиками, коэффициент вариации интервалов между пиками, энтропия распределения, межквартильный размах, автокорреляция, центр масс спектра, ширина спектра, спектральная энтропия, суммарная

мощность, соотношение мощности в коротких и длинных диапазонах, доминирующая частота, амплитуда доминирующей частоты, частота второго по величине пика, амплитуда второго по величине пика, период доминирующей волны, энтропия распределения энергии по масштабам и др. Есть еще признаки, учитывающие время: например, номер месяца, номер недели в году, день недели, синус и косинус дня года. В предположении о наличии солнечно-земных связей для исследования рассчитывался такой параметр, как продолжительность светового дня.

Обучение проводилось только на данных с мая по сентябрь каждого года с 2020 по 2023 (для фокусировки на тёплый сезон и уменьшения сезонной дисперсии).

Для повышения качества и воспроизводимости анализа, данные по напряжённости электрического поля проходили этапы предобработки, включающие медианную фильтрацию по окну в минуту для сглаживания шумов, а также восстановление малых пропусков линейной интерполяцией. Балансировка классов для экстремальных осадков не производилась, что может влиять на чувствительность метрик при редких явлениях, однако именно такое распределение отражает реальные условия оперативного прогноза.

Перед обучением модели все числовые признаки были масштабированы с использованием стандартизатора `StandardScaler`. Для учёта временной структуры данных применена стратегия временной кросс-валидации (`TimeSeriesSplit`), что позволило сохранить временную непрерывность между обучающей и валидационной выборками. Оптимизация гиперпараметров моделей проводилась с помощью библиотеки `Optuna`.

Протестирован широкий спектр методов искусственного интеллекта. Наиболее эффективными оказались ансамблевые методы деревьев решений (`LightGBM`, `XGBoost`, `CatBoost`, `Random Forest`) и глубокие нейронные сети (`GRU`, `LSTM`, `CNN DR RF`).

Ансамбли деревьев решений:

- `LightGBM`: высокоэффективный алгоритм градиентного бустинга, использующий рост деревьев по листьям (`leaf-wise`), что ускоряет обучение и повышает точность;

- `CatBoost`: устойчивый к переобучению бустинг, обладающий встроенной поддержкой категориальных признаков и высокой стабильностью на разреженных данных;

- `XGBoost`: регуляризованный градиентный бустинг с гибкой системой настройки и устойчивостью к переобучению;

- `Random Forest`: ансамбль решающих деревьев, обладающий высокой стабильностью и хорошей интерпретируемостью.

Глубокие нейросетевые модели:

- `GRU (Gated Recurrent Unit)`: компактная рекуррентная архитектура с возможностью моделирования временных зависимостей, требующая меньше параметров, чем `LSTM`;

- LSTM (Long Short-Term Memory): расширенная RNN-модель с долгосрочной памятью, хорошо справляющаяся с сезонными зависимостями;
- CNN DR RF (Dilated Receptive Field CNN): сверточная нейросеть с расширенным полем (receptive field) за счёт растущих коэффициентов «конволюции с отверстиями» (dilation rate). Позволяет эффективно выявлять паттерны в временных рядах на различных масштабах.

Обучение производилось на основе скользящего окна фиксированной длины (look-back), с прогнозом целевой переменной (осадков) на следующий день.

Ансамблевые методы деревьев решений LightGBM, CatBoost, XGBoost и Random Forest были реализованы с использованием стандартных API библиотек lightgbm, catboost, xgboost и scikit-learn. В качестве целевой функции, где не указано иначе, использовалась минимизация MAE. Все модели обучались на табличных признаках с лагами.

- LightGBM: использован boosting type = "gbdt", рост деревьев — leaf-wise. Обучение ускорено за счёт histogram-based binning. Проведена настройка num_leaves, min_child_samples, learning_rate и feature_fraction.

- CatBoost: применена схема Ordered Boosting с автокодировкой категориальных признаков. Гиперпараметры – depth, iterations, learning_rate, l2_leaf_reg.

- XGBoost: регуляризованный бустинг с параметрами gamma, lambda, max_depth. Использована функция потерь reg:squarederror и ранний останов (early_stopping_rounds).

- Random Forest: количество деревьев – 500, критерий – MSE. Каждое дерево обучалось на случайной подвыборке признаков (max_features=sqrt), использован bootstrap=True.

Все нейросетевые архитектуры реализованы с использованием библиотеки Keras (TensorFlow backend). Входные данные представляют собой тензоры формы (batch_size, sequence_length, num_features), полученные скользящим окном размером 3 дня.

- GRU-модель имеет два скрытых слоя с возвратом последовательностей. Последний слой GRU возвращает единственный выход, поступающий на полносвязный выходной слой. Оптимизатор: Adam. Функция потерь: MAE. Регуляризация: Dropout (0.2).

- Архитектура LSTM аналогична GRU, но использует три вентиля – входной, забывания и выходной.

- CNN DR RF – эта архитектура представляет собой сверточную нейросеть с увеличивающимся dilation_rate, что позволяет расширять область восприятия без увеличения глубины модели. После трёх сверточных слоёв применяется нормализация, уплощение и плотный выходной слой. LayerNormalization стабилизирует динамику градиентов и ускоряет сходимость. Такой подход особенно эффективен для временных рядов с выраженными шаблонами, где необходим захват как кратко-, так и долгосрочной зависимости.

Оценка качества предсказаний выполнялась на периоде наблюдений с 2023 по 2025 год с использованием следующих метрик:

- MAE (Mean Absolute Error) – средняя абсолютная ошибка;
- RMSE (Root Mean Squared Error) – корень из средней квадратичной ошибки;
- MAPE (Mean Absolute Percentage Error) – средняя абсолютная процентная ошибка;
- sMAPE (Symmetric MAPE) – симметричная средняя абсолютная процентная ошибка;
- Median RE – медианная относительная ошибка.

Для интерпретации вклада признаков в итоговое предсказание применялся метод SHAP (SHapley Additive exPlanations), позволяющий оценить влияние каждого входного параметра на результат модели. Следует отметить, что метод SHAP, используемый для оценки вклада признаков, чувствителен к мультиколлинеарности и шуму в данных.

Прогноз оценивается как по всем данным, так и по случаям с осадками более 3 мм, что полезно для оценки экстремальных событий.

Таким образом, результаты предсказаний моделей оценивались по средним абсолютным и процентным ошибкам: MAE, RMSE, MAPE и sMAPE. Это позволило получить объективную картину точности прогноза как в абсолютном выражении, так и относительно наблюдаемых величин.

Результаты и обсуждение

На обучающей и валидационной выборке наибольшую точность в прогнозировании суточных осадков продемонстрировала модель LightGBM, при этом средняя абсолютная ошибка (MAE) составила около 0,015 мм, средняя процентная ошибка (MAPE) – порядка 1 %.

Далее результаты прогноза оценивались на тестовом наборе выделенном в периоде наблюдений с 2023 по 2025 год и не участвовавшем ни в обучающей, ни в валидационной выборке.

Сравнение результатов строится через столбчатую диаграмму (рис. 1) и таблицу (табл. 1), метрики оценки – MAE, RMSE, MAPE, sMAPE, Median Relative Error – позволяют охарактеризовать точность как на всем множестве, так и на подмножестве интенсивных осадков (> 3 мм).

Сравнительный анализ охватывал модели lgbm, catboost, xgboost, randomforest, gru, lstm и гибридную модель cnn_dnnf, и выполнялся по следующим метрикам: MAE (средняя абсолютная ошибка), RMSE (средне-квадратичная ошибка), MAPE (средняя абсолютная процентная ошибка), sMAPE (симметричная MAPE), а также медианное относительное отклонение.

Общая точность моделей на всех данных (all). Модель CatBoost продемонстрировала наилучшие результаты по метрикам MAE (1.16) и RMSE (3.51), что указывает на высокую точность при общей выборке. Она также показала умеренные значения MAPE (98.71 %) и sMAPE (136.54 %), что важно в условиях высокой изменчивости данных.

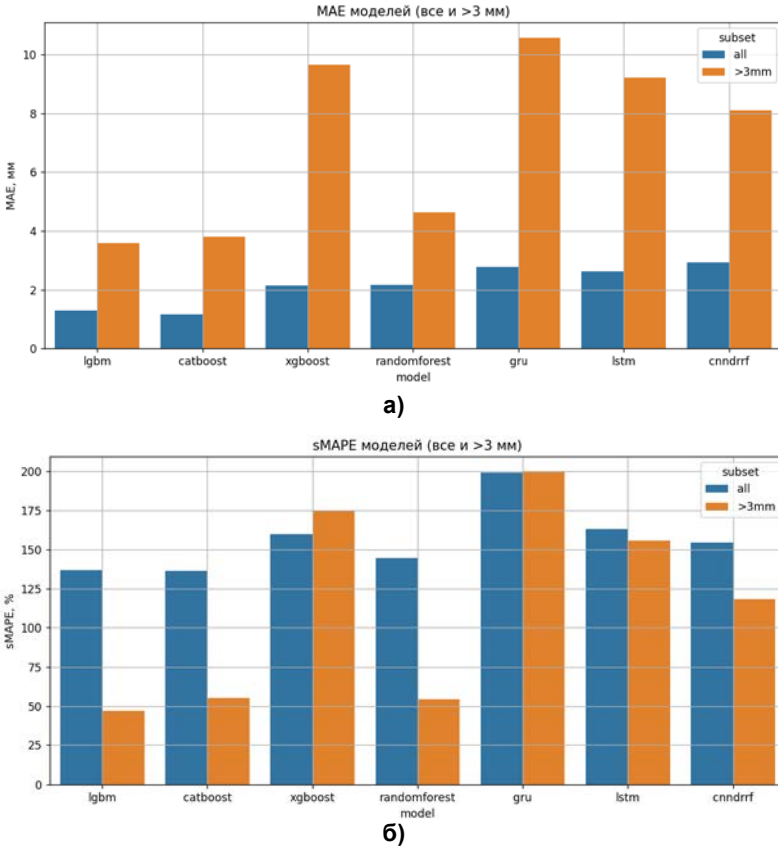


Рис. 1. Графики основных метрик, используемых для сравнения полученных моделей: средняя абсолютная ошибка (MAE) (а); симметричная средняя абсолютная процентная ошибка (sMAPE), устойчивая к нулям в данных (б).

Fig. 1. Plots of the main metrics used for the comparison of the resulting models: Mean Absolute Error (MAE) (a); Symmetric Mean Absolute Percentage Error (sMAPE), robust to zero values in the data (b).

Модель LightGBM оказалась сопоставимой по MAE (1.28) и имела чуть более высокую RMSE (3.67), но значительно хуже справлялась с относительными ошибками (MAPE = 147.85 %).

Модели xgboost, randomforest, gru, lstm и cnnrrf демонстрировали заметно худшие значения всех метрик, особенно XGBoost и GRU, где наблюдаются значительные ошибки (MAPE > 140 %, медианные ошибки достигают 90–114 %).

Производительность моделей на данных с интенсивными осадками (осадки > 3 мм).

В сегменте значительных осадков модель LightGBM достигла наилучших абсолютных метрик (MAE = 3.59, RMSE = 7.60) и продемонстрировала наименьшие значения MAPE (30.33 %) и sMAPE (47.06 %). Медианное

отклонение при этом оставалось нулевым, что свидетельствует о стабильности прогноза в наиболее критических ситуациях.

Таблица 1. Метрики для набора данных с признаками на основе НЭПА
Table 1. Metrics for the dataset with features based on the atmospheric electric field strength

Model	Subset	MAE	RMSE	MAPE	sMAPE	Median RE
lgbm	all	1.28	3.67	147.85	137.02	0.00
lgbm	>3mm	3.59	7.60	30.33	47.06	0.00
catboost	all	1.16	3.51	98.71	136.54	0.29
catboost	>3mm	3.80	7.67	32.05	55.26	0.29
xgboost	all	2.13	5.24	82.43	160.06	90.61
xgboost	>3mm	9.66	11.80	93.05	175.05	94.71
randomforest	all	2.17	4.03	238.18	144.63	48.27
randomforest	>3mm	4.64	7.66	37.91	54.38	29.14
gru	all	2.78	5.70	152.02	198.89	114.10
gru	>3mm	10.57	12.55	106.12	199.35	104.43
lstm	all	2.62	5.20	141.37	163.13	85.91
lstm	>3mm	9.22	11.44	86.88	155.98	88.36
cnndrrf	all	2.92	4.96	225.19	154.68	78.32
cnndrrf	>3mm	8.10	10.50	72.01	118.21	77.39

Модель CatBoost оказалась близкой по точности (MAE = 3.80, RMSE = 7.67), но имела худшие относительные метрики (MAPE = 32.05 %, sMAPE = 55.26 %).

Остальные модели, особенно XGBoost и рекуррентные нейросети (GRU, LSTM), показывали неудовлетворительные результаты в этом подмножестве – MAPE превышал 86 %, что делает их менее пригодными для задач оперативного прогнозирования опасных явлений.

Модель CNN DR RF, показала конкурентоспособные абсолютные метрики, но не превосходила LightGBM и CatBoost. При осадках >3 мм, она демонстрировала MAE = 8.10 и RMSE = 10.50, а относительные ошибки были умеренными (MAPE = 72.01 %, sMAPE = 118.21 %).

Резкое падение качества на тестовой выборке, более чем в 70 раз по MAE, – очень важный результат, который сейчас недостаточно обсуждается. Это может быть связано с тем, что тестовый период 2023–2025 гг. имел данные по тем или иным причинам, сильно отличающиеся от обучающего периода, возможно из-за аномальных погодных условий. Высокие значения процентных ошибок MAPE и sMAPE обусловлены большим количеством дней с нулевыми или почти нулевыми осадками, и именно поэтому для оценки также важны абсолютные метрики – MAE, RMSE.

Нулевая медианная ошибка у моделей LightGBM на обоих подмножествах данных, а также у CatBoost на всех данных, подтверждает их устойчивость к выбросам и делает их предпочтительными в практических приложениях, особенно в контексте раннего предупреждения.

Для изучения вклада признаков на основе НЭПА в результативность прогноза осадков были выполнены эксперименты с набором данных, где эти признаки были исключены. Результаты экспериментов приведены на рис. 2 и в табл. 2.

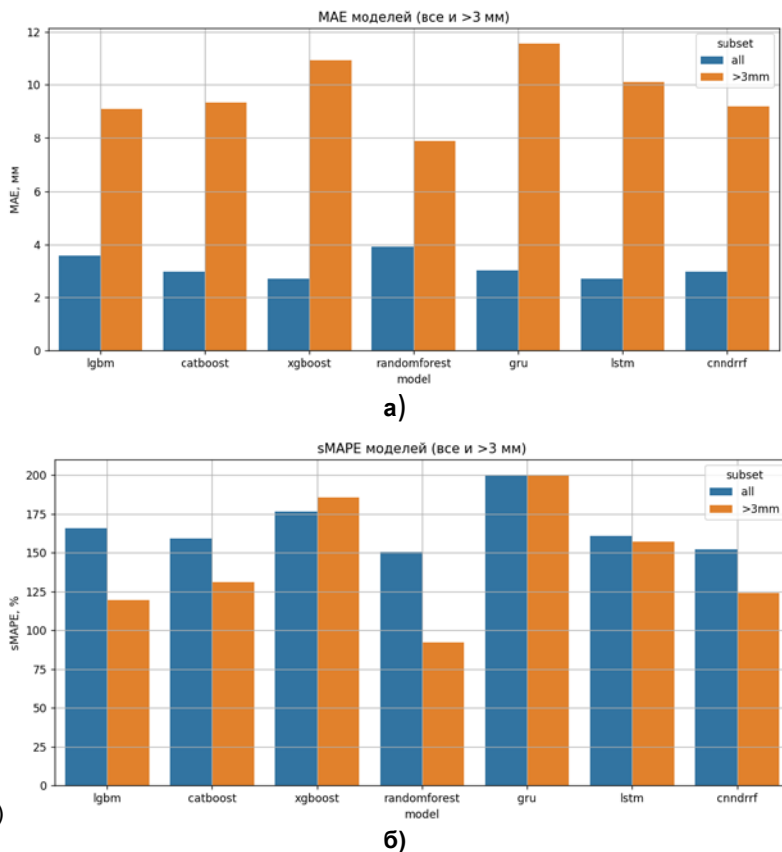


Рис. 2. Графики основных метрик для набора данных без признаков на основе напряженности электрического поля атмосферы, используемых для сравнения полученных моделей: средняя абсолютная ошибка (MAE) (а); симметричная средняя абсолютная процентная ошибка (sMAPE), устойчивая к нулям в данных (б).

Fig. 2. Plots of the main metrics for the dataset without features based on the atmospheric electric field strength, used for the comparison of the resulting models: Mean Absolute Error (MAE) (a); Symmetric Mean Absolute Percentage Error (sMAPE), robust to zero values in the data (b).

Анализ результатов показал, что включение признаков на основе измерений напряжённости электрического поля атмосферы приводит к статистически значимому снижению ошибок прогнозирования. Например, для моделей CatBoost и LightGBM средняя абсолютная ошибка (MAE) по всей выборке снижается с 2.97 до 1.16 мм и с 3.59 до 1.28 мм

соответственно, а аналогичное улучшение отмечается и для других метрик (RMSE, MAPE, sMAPE). Для подмножества интенсивных осадков (>3 мм) наблюдается аналогичное преимущество: MAE и RMSE для LightGBM уменьшаются почти вдвое при добавлении электрических данных. Это свидетельствует о наличии дополнительной прогностической информации, вносимой электрическими параметрами, даже несмотря на то, что они не всегда входят в топ по SHAP-важности.

Таблица 2. Метрики для набора данных без признаков на основе НЭПА
Table 2. Metrics for the dataset without features based on the atmospheric electric field strength

Model	Subset	MAE	RMSE	MAPE	sMAPE	Median RE
lgbm	all	3.59	6.13	356.19	166.04	116.35
lgbm	>3mm	9.11	12.11	76.91	119.33	77.80
catboost	all	2.97	5.81	220.91	159.29	90.28
catboost	>3mm	9.34	12.17	75.35	131.06	86.35
xgboost	all	2.71	6.23	94.94	176.51	96.95
xgboost	>3mm	10.94	13.36	96.13	185.73	97.65
randomforest	all	3.92	6.25	370.83	150.48	87.35
randomforest	>3mm	7.89	11.47	57.46	92.20	62.65
gru	all	3.01	6.12	165.39	200.00	122.38
gru	>3mm	11.55	13.76	106.79	200.00	105.40
lstm	all	2.71	5.58	168.03	160.69	87.11
lstm	>3mm	10.09	12.59	86.94	156.99	92.03
cnndrrf	all	2.96	5.45	261.29	152.12	81.35
cnndrrf	>3mm	9.19	12.02	73.33	124.03	81.35

Таким образом, проведено обучение нескольких моделей машинного обучения, среди которых наилучшими показателями точности выделяются LightGBM, CatBoost, LSTM.

Относительная ошибка прогноза осталась небольшой даже в ситуациях с высоким уровнем осадков, что говорит о способности модели предсказывать выбросы – экстремальные события.

Использование метода SHAP обеспечивает интерпретируемость модели, что важно для практического применения в метеорологии. Характеристики НЭПА полученные с помощью Фурье и вейвлет-преобразования, а также лаги температуры и влажности оказались наиболее значимыми признаками, подтверждая их физическую обоснованность в контексте прогноза осадков. Применение лаговых признаков и спектральных характеристик позволяет учитывать как краткосрочные, так и долгосрочные зависимости в данных, что улучшает точность прогноза. Среди спектральных и вейвлет-признаков наибольший вклад в качество прогноза внесли следующие характеристики: доминирующая частота, спектральная энтропия и суммарная мощность в коротковолновой области, а также амплитуда

второй по величине спектральной компоненты. Это соответствует физической природе осадков, так как характерная частота и вариабельность электрических сигналов и метеопараметров связаны с фазами развития облачных структур и фронтальной активностью. При исключении электрических данных метрики прогноза ухудшаются, что подтверждает их дополнительную информативность и синергетический эффект при совместном использовании с классическими метеопризнаками.

Полученные результаты подчёркивают потенциал использования методов машинного обучения в оперативном прогнозе осадков.

Заключение

В настоящем исследовании была проведена оценка эффективности различных моделей машинного обучения и глубокого обучения для прогнозирования осадков.

Модель LightGBM продемонстрировала наилучшее сочетание точности и устойчивости как на всей выборке, так и при фокусе на интенсивных осадках. CatBoost также является надёжной альтернативой. Модели на основе нейросетей уступают по ряду метрик, что, возможно, связано с ограниченностью обучающей выборки или их архитектурными особенностями. Результаты подтверждают перспективность градиентного бустинга в задачах прогнозирования осадков при использовании комплексной информации.

В результате проведенного исследования был разработан и апробирован подход к краткосрочному прогнозированию суточных осадков на основе метеорологических данных и измерений напряженности электрического поля с применением методов машинного обучения.

Полученные значения MAE для лучших моделей (например, CatBoost – 1.16 мм, LightGBM – 1.28 мм на всей выборке) соответствуют критериям высокой точности для оперативного суточного прогноза осадков, согласно публикациям [15, 19]. Тем не менее, нейросетевые модели (GRU, LSTM) показывают существенно худшие результаты, что указывает на необходимость дальнейшей оптимизации их архитектур и расширения обучающей выборки.

Анализ важности признаков, проведённый методом SHAP, выявил ключевые факторы, влияющие на точность модели, среди которых выделяются лаговые значения температуры и влажности, а также спектральные характеристики временных рядов напряженности поля. Это подтверждает эффективность использования таких производных признаков в задачах прогнозирования атмосферных явлений.

Дальнейшее развитие исследований предполагает интеграцию дополнительных данных, например, таких как спутниковые наблюдения и другие методы дистанционного зондирования атмосферы, а также использование более сложных нейросетевых архитектур, способных учитывать одновременно пространственные и временные закономерности.

Работа демонстрирует эффективность подхода и может быть полезна для анализа возможностей методов искусственного интеллекта в задачах краткосрочного прогноза. Её продолжение может быть полезно для оперативного прогнозирования погоды и принятия решений в различных отраслях, зависящих от погодных условий, например, для оперативного прогнозирования погодных условий в авиации, сельском хозяйстве и для предупреждения чрезвычайных ситуаций, связанных с экстремальными осадками.

Практические результаты работы могут быть использованы для оперативного прогноза опасных явлений в регионах с ограниченной наблюдательной сетью, а также при внедрении результатов измерений онлайн-систем мониторинга электрических параметров атмосферы. Планируется расширить спектр анализируемых признаков и протестировать разработанный подход на других станциях с различными типами климата. В дальнейшем возможно использование методов балансировки классов и интеграция данных дистанционного зондирования для повышения точности предсказаний экстремальных событий.

Список литературы

1. *Аджиев А.Х., Кондратьева Н.В., Кумукова О.А., Сейнова И.Б., Богаченко Е.М.* Метод фонового прогнозирования селевой опасности на Центральном Кавказе и результаты его апробирования // Селевые потоки: катастрофы, риск, прогноз, защита. 2008. С. 263-266.
2. *Белякова П.А., Борщ С.В., Христофоров А.В., Юмина Н.М.* Вероятностное прогнозирование максимальных расходов и уровней воды и критических осадков на реках Черноморского побережья Кавказа // Труды Гидрометцентра России. 2013. № 349. С. 104-121.
3. *Бисحوков Р.М.* Прогнозирование изменений режима атмосферных осадков на Северном Кавказе // Вестник Курганской ГСХА. 2018. № 1 (25). С. 14-17.
4. *Болгов М.В., Арефьева Е.В.* О некоторых вопросах, связанных с прогнозированием чрезвычайных ситуаций, вызванных гидрологическими опасными явлениями и их последствиями // Научные и образовательные проблемы гражданской защиты. 2017. №. 4 (35). С. 102-110.
5. *Лубков А.С., Вышваркова Е.В., Воскресенская Е.Н.* Исследование возможности нейросетевого прогнозирования осадков в Крымском регионе // Тезисы докладов международной научно-технической конференций «Системы контроля окружающей среды – 2019», Севастополь, 12–13 сентября 2019 г. Севастополь: ИПТС, 2019. С. 152.
6. *Митин В.Ю., Аиткулов В.Н.* Математические модели прогнозирования температуры и осадков с использованием фрактального и Фурье-анализа метеорологических рядов // Вестник Пермского университета. Серия: Математика. Механика. Информатика. 2024. №. 1 (64). С. 33-42.
7. *Морейдо В.М., Гарцман Б., Соломатин Д.П., Сучилина З.А.* Возможности краткосрочного прогнозирования стока малой реки с использованием методов машинного обучения // Гидросфера. Опасные процессы и явления. 2020. Т. 2, №. 4. С. 375-390.
8. *Таланов Е.А.* Технология краткосрочного вероятностного прогнозирования паводков и селевых потоков на основе прогноза количества осадков на 3 и 6 часов // Селевые потоки: катастрофы, риск, прогноз, защита. 2016. С. 229-233.
9. *Фрумин Г.Т., Иванов М.Э., Куликова Л.А., Еремина А.В.* Подходы к решению проблемы диагноза и прогноза атмосферных осадков в интересах городских служб водоотведения // Общество. Среда. Развитие (Тегга Нумана). 2016. №. 2 (39). С. 97-100.
10. *Шайдуллина А.Ф.* Сравнительный ретроспективный анализ методов прогнозирования лесных пожаров по метеорологическим данным // Проблемы науки. 2016. №. 7 (8). С. 23-26.

11. Яровая Д.А., Посошков В.Л. Параметры атмосферных осадков в Азово-Черноморском регионе на основе модельных и натуральных данных // Морской гидрофизический журнал. 2017. № 1 (193). С. 11-26.
12. Balasmeh O. Al, Babbar R., Karmaker T. Trend analysis and ARIMA modeling for forecasting precipitation pattern in Wadi Shueib catchment area in Jordan // Arabian Journal of Geosciences. 2019. Vol. 12. P. 1-19.
13. Dong J., Zeng W., Wu L., Huang J., Gaiser T., Srivastava A.K. Enhancing short-term forecasting of daily precipitation using numerical weather prediction bias correcting with XGBoost in different regions of China // Engineering Applications of Artificial Intelligence. 2023. Vol. 117. P. 105579.
14. Kusiak A., Wei X., Verma A.P., Roz E. Modeling and Prediction of Rainfall Using Radar Reflectivity Data: A Data-Mining Approach // IEEE Transactions on Geoscience and Remote Sensing. 2013. Vol. 51, no. 4. P. 2337–2342.
15. Nastos P.T., Paliatatos A.G., Koukoultsos K.V., Larissi I.K., Moustiris K.P. Artificial neural networks modeling for forecasting the maximum daily total precipitation at Athens, Greece // Atmospheric Research. 2014. Vol. 144. P. 141-150.
16. Partal T., Cigizoglu H.K. Prediction of daily precipitation using wavelet–neural networks // Hydrological sciences journal. 2009. Vol. 54, no. 2.P. 234-246.
17. Ponnoprat D. Short-term daily precipitation forecasting with seasonally-integrated auto-encoder // Applied Soft Computing. 2021. Vol. 102. P. 107083.
18. Sodoudi S., Noorian A., Geb M., Reimer E. Daily precipitation forecast of ECMWF verified over Iran. Theoretical and applied climatology, 2010, Vol. 99, pp. 39-51.
19. Valipour M., Khoshkam H., Bateni S.M., Jun C. Machine-learning-based short-term forecasting of daily precipitation in different climate regions across the contiguous United States // Expert Systems with Applications. 2024. Vol. 238. P. 121907.
20. Waqas M., Humphries U.W., Hlaing P.T., Wangwongchai A., Dechpichai P. Advancements in daily precipitation forecasting: a deep dive into daily precipitation forecasting hybrid methods in the tropical climate of Thailand // MethodsX. 2024. Vol. 12. P. 102757.
21. Weerasinghe H.D.P., Premaratne H.L., Sonnadara D.U.J. Performance of neural networks in forecasting daily precipitation using multiple sources // Journal of the National Science Foundation of Sri Lanka. 2010. Vol. 38, no. 3. P. 163-170.

References

1. Adzhiev A.H., Kondrat'eva N.V., Kumukova O.A., Seynova I.B., Bogachenko E.M. Metod fonovogo prognozirovaniya selevoy opasnosti na Central'nom Kavkaze i rezul'taty ego aprobirovaniya. *Selevye potoki: katastrofy, risk, prognoz, zashchita*, 2008, pp. 263-266 [in Russ.].
2. Belyakova P.A., Borsch S.V., Khristoforov A.V., Yumina N.M. Probabilistic forecasting of maximal discharges, water levels and critical precipitation for the rivers of the Black sea coast of Caucasus. *Trudy Gidromettsentra Rossii [Proceedings of the Hydrometcentre of Russia]*, 2013, vol. 349, pp. 104-121 [in Russ.].
3. Bischokov R.M. Prognozirovaniye izmeneniy rezhima atmosferynykh osadkov na Severnom Kavkaze. *Vestnik Kurganskoj GSKHA*, 2018, no. 1 (25), pp. 14-17 [in Russ.].
4. Bolgov M.V., Arefeva E.V. O nekotorykh voprosah, svyazannykh s prognozirovaniem chrezvychaynykh situatsiy, vyzvannykh gidrologicheskimi opasnymi yavleniyami i ih posledstviyami. *Nauchnye i obrazovatel'nye problemy grazhdanskoj zashchity*, 2017, no. 4 (35), pp. 102-110 [in Russ.].
5. Lubkov A.S., Vyshkvarikova E.V., Voskresenskaya E.N. Issledovanie vozmozhnosti neyrossetevogo prognozirovaniya osadkov v Krymskom regione. Tezisy dokladov mezhdunarodnoy nauchno-tehnicheskoy konferentsiy «Sistemy kontrolya okruzhayushchey sredy – 2019», Sevastopol', 12–13 sentyabrya 2019 g, Sevastopol', IPTS, 2019, 152 p. [in Russ.].
6. Mitin B. Iu., Aptukov B.H. Mathematical Models of Temperature and Precipitation Forecasting Using Fractal and Fourier-Analysis of Meteorological Series. *Vestnik Permskogo universiteta. Seriya: Matematika. Mekhanika. Informatika. [Bulletin of Perm University. Mathematics. Mechanics. Computer Science]*, 2024, vol. 64, no. 1, pp. 33-42 [in Russ.]. DOI: 10.17072/1993-0550-2024-1-33-42.

7. Moreido V.M., Gartsman B.I., Solomatine D.P., Suchilina Z.A. Prospects for short-term forecasting of river streamflow from small watershed runoff using machine learning methods. *Gidrosfera. Opasnye processy i yavleniya [Hydrosphere. Hazard processes and phenomena]*, 2020, vol. 2, no. 4, pp. 375-390 [in Russ.]. DOI: 10.34753/HS.2020.2.4.375
8. Talanov E.A. Tekhnologiya kratkosrochnogo veroyatnostnogo prognozirovaniya padkov i selevyh potokov na osnove prognoza kolichstva osadkov na 3 i 6 chasov. *Selevye potoki: katastrofy, risk, prognoz, zashchita*. 2016, pp. 229-233 [in Russ.].
9. Frumin G.T., Ivanov M.E., Kulikova L.A., Eryomina A.V. Approaches to problem solving of diagnosis and forecasting of precipitation for the city services needs. *Obshchestvo. Sreda. Razvitiye (Terra Humana) [Society. Environment. Development (TERRA HUMANA)]*, 2016, vol. 39, no. 2, pp. 97-100 [in Russ.].
10. Shaydullina A.F. Sravnitel'nyy retrospektivnyy analiz metodov prognozirovaniya lesnyh pozharov po meteorologicheskim dannym. *Problemy nauki [Science problems]*, 2016, vol. 8, no. 7, pp. 23-26 [in Russ.].
11. Iarovaia D.A. Pososhkov V.L. Parameters of the Azov-Black Sea Region Precipitation Based on the Model and Observational Data. *Physical Oceanography*, 2017, no. 1, pp. 11-24 [in Russ.]. DOI: 10.22449/1573-160X-2017-1-11-24
12. Balasmeh O. Al, Babbar R., Karmaker T. Trend analysis and ARIMA modeling for forecasting precipitation pattern in Wadi Shueib catchment area in Jordan. *Arabian Journal of Geosciences*, 2019, vol. 12, pp. 1-19.
13. Dong J., Zeng W., Wu L., Huang J., Gaiser T., Srivastava A.K. Enhancing short-term forecasting of daily precipitation using numerical weather prediction bias correcting with XGBoost in different regions of China. *Engineering Applications of Artificial Intelligence*, 2023, vol. 117, pp. 105579.
14. Kusiak A., Wei X., Verma A.P., Roz E. Modeling and Prediction of Rainfall Using Radar Reflectivity Data: A Data-Mining Approach. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, vol. 51, no. 4, pp. 2337-2342.
15. Nastos P.T., Paliatsos A.G., Koukouletsos K.V., Larissi I.K., Moustiris K.P. Artificial neural networks modeling for forecasting the maximum daily total precipitation at Athens, Greece. *Atmospheric Research*, 2014, vol. 144, pp. 141-150.
16. Partal T., Cigizoglu H.K. Prediction of daily precipitation using wavelet-neural networks. *Hydrological sciences journal*, 2009, vol. 54, no. 2, pp. 234-246.
17. Ponnoprat D. Short-term daily precipitation forecasting with seasonally-integrated autoencoder. *Applied Soft Computing*, 2021, vol. 102, pp. 107083.
18. Sodoudi S., Noorian A., Geb M., Reimer E. Daily precipitation forecast of ECMWF verified over Iran. *Theoretical and Applied Climatology*, 2010, vol. 99, pp. 39-51.
19. Valipour M., Khoshkam H., Bateni S.M., Jun C. Machine-learning-based short-term forecasting of daily precipitation in different climate regions across the contiguous United States. *Expert Systems with Applications*, 2024, vol. 238, pp. 121907.
20. Waqas M., Humphries U.W., Hlaing P.T., Wangwongchai A., Dechpichai P. Advancements in daily precipitation forecasting: a deep dive into daily precipitation forecasting hybrid methods in the tropical climate of Thailand. *MethodsX*, 2024, vol. 12, pp. 102757.
21. Weerasinghe H.D.P., Premaratne H.L., Sonnadara D.U.J. Performance of neural networks in forecasting daily precipitation using multiple sources. *Journal of the National Science Foundation of Sri Lanka*, 2010, vol. 38, no. 3, pp. 163-170.

Поступила 20.06.2025; одобрена после рецензирования 01.10.2025;
принята в печать 15.10.2025.

Submitted 20.06.2025; approved after reviewing 01.10.2025;
accepted for publication 15.10.2025.