

Прогнозирование NDVI на основе интеграции спутниковых и метеорологических данных для мониторинга сельскохозяйственных территорий

К.А. Маковейчук, А.В. Олифинов, Я.Т. Маковейчук

Аннотация— В статье рассматривается задача мониторинга состояния растительности сельскохозяйственных территорий на основе анализа и прогнозирования индекса NDVI с использованием спутниковых данных Sentinel-2 и метеорологических наблюдений. Обоснована целесообразность интеграции данных дистанционного зондирования Земли и измерений, получаемых с использованием технологий Интернета вещей (IoT).

Сформирован конвейер подготовки данных, включающий извлечение, синхронизацию и предобработку спутниковых и метеорологических временных рядов (интерполяция пропусков, сглаживание, агрегация), что позволяет получить устойчивый обучающий датасет.

На его основе реализованы и исследованы модели прогнозирования, включая градиентный бустинг и рекуррентную нейронную сеть LSTM. Проведён сравнительный анализ, показавший, что при использовании агрегированных временных рядов ансамблевые методы обеспечивают более высокую точность по сравнению с рекуррентными моделями.

Установлено, что ключевой вклад в прогноз вносят лаговые значения NDVI, что свидетельствует о доминирующей роли автокорреляционной структуры временного ряда и определяет ограничения интерпретации влияния внешних факторов.

Предложенный подход может быть использован при построении интеллектуальных систем мониторинга и поддержки принятия решений в сельском хозяйстве, включая системы с использованием IoT-инфраструктуры.

Ключевые слова— NDVI, временные ряды, дистанционное зондирование Земли, интеграция данных, метеорологические данные.

I. ВВЕДЕНИЕ

В конце XX века, в связи с развитием технологических и программных средств для сбора и анализа данных о состоянии посевных площадей, почвы, растений и агроклиматических условий, сформировалось новое

направление — точное земледелие (precision agriculture).

Современные системы точного земледелия обеспечивают:

- повышение эффективности сельскохозяйственного производства за счёт оптимизации затрат;
- дифференцированное управление агротехнологическими процессами;
- экологическую устойчивость за счёт рационального использования ресурсов.

Ключевыми технологическими компонентами точного земледелия являются системы глобального позиционирования, геоинформационные системы, технологии дистанционного зондирования Земли (ДЗЗ) и решения на основе Интернета вещей (IoT) [1].

Каждое из направлений активно развивается, в том числе за счёт применения методов глубокого обучения [20-21].

Дистанционное зондирование Земли является одним из ключевых инструментов мониторинга сельскохозяйственных территорий [2]. Развитие спутниковых систем, таких как Sentinel и Landsat, обеспечило доступ к регулярным мультиспектральным данным с высоким пространственным и временным разрешением.

Одним из наиболее распространённых показателей, используемых при анализе состояния растительности, является индекс NDVI, позволяющий оценивать вегетационную активность растений.

На практике NDVI широко применяется в цифровых платформах мониторинга (например, OneSoil), а также в специализированных решениях, включая индекс плодородия почвы ASF [3], основанный на многолетних значениях NDVI.

Современные системы мониторинга сельскохозяйственных объектов всё чаще строятся на основе интеграции разнородных источников данных. Существенную роль здесь играют технологии Интернета вещей (IoT), обеспечивающие сбор данных о состоянии среды в режиме, близком к реальному времени.

Использование IoT-датчиков (температура, влажность, осадки, давление) позволяет дополнить спутниковые наблюдения локальными измерениями и повысить точность анализа. Интеграция IoT и данных ДЗЗ формирует основу для построения интеллектуальных систем мониторинга, способных

Статья получена 7 марта 2026.

К.А. Маковейчук – ФГБОУ ВО "Финансовый университет при Правительстве Российской Федерации", г. Москва (e-mail: kamakovejchuk@fa.ru)

А.В. Олифинов – Гуманитарно-педагогическая академия (филиал) ФГАОУ ВО "Крымский федеральный университет им. В. И. Вернадского" в г. Ялте (e-mail: alex.olifirov@gmail.com)

Я.Т. Маковейчук – ФГБОУ ВО "Финансовый университет при Правительстве Российской Федерации", г. Москва (e-mail: ytmakovejchuk@fa.ru).

учитывать как пространственные, так и временные особенности агроэкосистем.

В то же время, несмотря на наличие отдельных решений, задача комплексного использования спутниковых и IoT-данных для прогнозирования показателей состояния растительности остаётся недостаточно проработанной.

Существующие подходы к анализу NDVI преимущественно основаны на ретроспективных данных или сравнении со среднесезонными значениями. Это ограничивает возможности оперативного принятия решений в условиях изменяющихся погодных факторов.

В связи с этим актуальной является задача разработки методов прогнозирования NDVI с учётом совокупности спутниковых и метеорологических данных, включая данные IoT-наблюдений.

Новые методы и методики мониторинга создают возможности реализации приоритетных направлений фундаментальных и поисковых научных исследований (rubрика 1.5.10.5. «Экономическая, социальная и политическая география», [4]) по выявлению факторов и тенденций пространственного развития России и ее регионов, по выбору оптимальных моделей территориальной организации населения и хозяйства.

«Своевременно реагировать на изменение условий вегетации и оперативно принимать оптимальные решения по проведению агрономических мероприятий» агрономы могут на основе сравнения среднесезонных значений индекса NDVI с его текущими сезонными значениями [5].

Это подчеркивает актуальность разработки моделей прогнозирования индекса NDVI, основанных на интеграции спутниковых и метеорологических данных. Использование методов машинного обучения позволяет учитывать сложные нелинейные зависимости между параметрами среды и динамикой растительности, что расширяет возможности анализа и прогноза состояния сельскохозяйственных территорий.

Целью работы является разработка метода сбора, предобработки и интеграции спутниковых и метеорологических данных, включая данные IoT-датчиков, а также построение модели прогнозирования индекса NDVI для мониторинга состояния сельскохозяйственных полей и поддержки принятия решений.

II. АНАЛИЗ НАУЧНЫХ ИССЛЕДОВАНИЙ В ОБЛАСТИ СБОРА И ОБРАБОТКИ ДАННЫХ СО СПУТНИКОВЫХ ИЗОБРАЖЕНИЙ

На основе анализа отечественных и зарубежных исследований, касающихся использования информации спутниковых снимков, для формирования исходных данных был проведён анализ методов, спутников и платформ доступа к спутниковым снимкам (см. таблицу I). Очевидно, что активно используются классические модели МО и модели глубокого обучения [20-21], при этом способов анализа, по сути, два: анализ собственно изображений со спутников и анализ агрегированной статистической информации о различных индексах (растительности), прежде всего индекса NDVI. Можно выделить несколько типовых направлений в разработке

моделей, основываясь на проведенном анализе.

A. Модель прогнозирования NDVI и построения вегетационных карт полей

В работе [6] предложен подход к прогнозированию индекса NDVI на основе рекуррентной нейронной сети с долгой краткосрочной памятью LSTM (Long Short-Term Memory). Использование в моделях обработки спутниковых данных сети LSTM является частым, так как LSTM хорошо подходит для прогнозирования временных рядов. Кроме извлеченных с изображений полей Волгоградской области значений NDVI за 5 лет, датасет для обучения модели, как и в ряде подобных работ, был дополнен признаками, соответствующими характеристикам погоды. Интересным направлением использования результатов прогноза является построение вегетационных карт полей NDVI. Карты могут быть построены на основе спутниковых снимков и дополнены вегетационными изображениями, сформированными по спрогнозированным значениям NDVI. Такие карты можно использовать для выявления проблемных полей.

B. Модель прогнозирования NDVI на основе изображений и статистических расчетных характеристик

В работе [7] были скомбинированы две модели: модель CNN (Convolutional Neural Network) для прогнозирования краткосрочного NDVI по снимкам, и модель SF-CNN (statistical features), для которой на основании NDVI рассчитываются статистические характеристики, и подаются на ее вход вместе с исходными данными. Такая методика выявляет закономерности из анализируемых данных со снимков, и позволяет повысить точность прогнозов NDVI на следующие 3 месяца с 30-дневным интервалом в нескольких сложных областях.

C. Метод крупномасштабной оценки урожайности сельскохозяйственных культур

В работе [8] разработана модель - конвейер обработки для прогнозирования урожайности на основе данных спутниковых временных рядов. Модель извлекает со снимков Sentinel-2 по픽сельно информацию, период снимков – 11 дней, затем, после этапа маскировки облаков, выполняется построение гистограмм и медиан признаков для нейромодели TCN (temporal convolutional network). TCN получала на вход тензор 32×10 (гистограмма), либо тензор 1×10 (медиана) для каждой временной точки. Также для сравнения была обучена модель RF (Random forests), которая получала на вход $\text{time-steps} \times 32 \times 10$ (гистограмма) или $\text{time-steps} \times 1 \times 10$ (медиана). Оценки и прогноз урожайности сельскохозяйственных культур на уровне фермы показали, что модель TCN превосходит базовую модель RF. Однако, на уровне региона TCN достигла наилучшей точности прогнозирования.

В других рассмотренных работах [30, 31] также исследователи ставили своими целями прогнозирование и восполнение данных по NDVI либо оценку и прогнозирование урожайности, при этом основываясь на сверточных нейронных сетях, рекуррентных нейронных

сетях, классических моделях, таких, как случайный лес, улучшать качество прогнозов. и производных статистических признаках, позволяющих

Таблица 1: Анализ исследований растительного покрова для сельского хозяйства, основанных на данных аэрокосмических (спутниковых) снимков

Источник информации (спутник)	Способ обработки (методы КЗ / извлечение индексов по длинам волн)	Модель МО / метод	Сфера применения результатов	Автор(ы), страна
Sentinel-2, данные с портала apps.sentinel-hub.com	На основе временных рядов прогнозирование значений NDVI на каждый пиксель спутникового снимка	LSTM (Long Short-Term Memory - Долговременная краткосрочная память)	Предложена модель прогнозирования NDVI и построения на его основе вегетационных карт, основанная на рекуррентных нейронных сетях с архитектурой LSTM	Шурлаева Е.А. и др. [6]
Terra (США), программа для извлечения данных IDL	Данные предварительно обрабатываются по извлеченным статистическим признакам, которые затем вводятся в нейронную сеть вместе с необработанными данными	CNN (Convolutional Neural Network - Сверточная нейронная сеть) + SF-CNN (Statistical Features CNN - CNN со статистическими признаками)	Предложена модель прогнозирования NDVI на основе CNN + SF-CNN, объединившая как преимущества по-пиксельного прогноза CNN, так и прогноза на статистических признаках на основе исторических данных, что повышает точность прогнозов NDVI	Cui, C. et al. [7]
Sentinel-2, информация с платформы EODIE	Агрегирование статистической информации о различных индексах (растительности) или простых интервалах длин волн	TCN (Temporal Convolutional Net-work - временная сверточная сеть) + RF (Random Forest - Случайный лес)	Предложен метод крупно-масштабной сезонной оценки урожайности сельскохозяйственных культур на основе временных рядов спутниковых изображений (satellite image time series, SITS) для статистического растениеводства и прогнозирования	Yli-Heikkilä, M. et al. [8] Wittke, S. et al. [9]
Sentinel-2, данные собраны и обработаны THEIA	Анализ дескрипторов спектрального распределения наблюдений путем случайной выборки пикселей + извлечение временных функций = классификация с учетом временных различий	PSE+LTAE (Pixel-Set Encoder + Light-weight Temporal Attention Encoder)	Предложена модель меж- и внутригодовой сельскохозяйственной динамики ежегодной классификации участков с помощью подхода глубоко-го обучения	Quinton, F. et al. [10] Nyborg, J. et al. [11]
Sentinel-1 + Sentinel-2, данные собраны и обработаны THEIA	Исследование набора мультимедийных и мультимедийных патчей на основе объединения пространственно-временных и пространственно-спектральных признаков для семантической сегментации (классификации)	ConvLSTM + Inception-S1S2 (Convolutional Long Short-Term Memory - Сверточная долго-временная кратко-срочная память + начальный модуль по дате)	Предложена модель глубокого обучения, позволяющая извлекать пространственно-спектральные и пространственно-временные характеристики для семантической сегментации	Wenger, R. et al. [12]
Sentinel-1 + Sentinel-2, на основе платформы Google Earth Engine	Классификация типов культур путем комбинирования оптических изображений временных рядов (Sentinel-2) и радиолокационных изображений с синтезированной апертурой (Sentinel-1)	RF (Random Forest - Случайный лес)	Предложена модель для картографирования по типам культур путем классификации комбинированных изображений, объединяющая пиксельную классификацию с сегментацией изображений. Как признаки использованы месячные медианные значения EVI, LSWI. и NDVI.	Guo, L. et al. [13]

III. АНАЛИЗ ВЕГЕТАЦИОННОГО ИНДЕКСА NDVI

Направления исследований спутниковых снимков очень широки и основываются на показателях, которые могут быть получены со снимков.

Существует более ста индексов растительности, полученных на основе мультиспектральных снимков [14]. В 1969 году в работе Kriegler FJ. et al., было предложено простое преобразование полос [15]:

$$NDVI = \frac{NIR - RED}{NIR + RED}, \quad (1)$$

где RED и NIR - измерения спектрального излучения (или отражения), зарегистрированные датчиками в красной (видимой) и ближней инфракрасной областях, соответственно. В результате получается новое упрощенное изображение, называемое нормализованным вегетационным индексом - Normalized Difference Vegetation Index, NDVI.

NDVI является наиболее популярным показателем состояния растения. Значения NDVI варьируются от -1 до 1, независимо от того, что используется в качестве входных данных - излучение, отражение или цифровые измерения. В целом, его значения отрицательны для водных объектов, близки к нулю для скал, песков или бетонных поверхностей, и положительны для растительности, включая сельскохозяйственные культуры, кустарники, травы и леса [2]. Если значение NDVI близко к 1, это говорит о том, что территория покрыта здоровой растительностью. Значения NDVI от 0 до 0,2 указывают на бесплодную или разреженную растительность, а значения NDVI от 0,2 до 0,5 - на сухую растительность. Значения NDVI от 0,5 до 0,8 указывают

на здоровую растительность, а значения NDVI выше 0,8 - на густую растительность.

NDVI можно рассчитать с помощью любого мультиспектрального датчика, имеющего видимый и ближний ИК-диапазон. Большинство спутниковых оптических датчиков используют несколько широких спектральных диапазонов.

IV. РАЗРАБОТКА КОНЦЕПТУАЛЬНОЙ МОДЕЛИ

Согласно поставленной цели, модель для прогнозирования показателя NDVI для мониторинга состояния полей сельскохозяйственной отрасли в регионе должна быть обучена на данных NDVI, объединенных с показателями условий среды (температура, влажность, осадки, атмосферное давление). Для разработки, обучения и инференса такой модели необходимо реализовать следующие этапы:

- выполнить сбор и обработку данных;
- разработать архитектуру модели;
- выполнить обучение и тестирование модели, валидацию и оценку производительности модели;
- выполнить интеграцию модели в веб-приложение.

Концептуальная модель предсказания показателя NDVI для мониторинга состояния полей сельскохозяйственной отрасли представлена на рис. 1.

Концептуальная модель (рис. 1) отражает полный конвейер обработки данных — от этапа сбора спутниковых и метеорологических данных до построения и оценки моделей прогнозирования NDVI. В нее не включен блок «использование» - он носит прикладной характер и отражает интеграцию модели в информационные системы мониторинга.



Рис.1: Концептуальная модель прогнозирования показателя NDVI для мониторинга состояния полей сельскохозяйственной отрасли

V. ИНТЕГРАЦИЯ ИОТ-ИСТОЧНИКОВ ДАННЫХ И ЗАДАЧИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Современные системы мониторинга сельскохозяйственных территорий, наряду с использованием спутниковых данных и метеорологических наблюдений, могут включать распределённые IoT-устройства, такие как почвенные датчики влажности, температуры, уровня освещённости, а также локальные агрометеостанции.

Интеграция IoT-устройств позволяет повысить детализацию наблюдений за состоянием агроэкосистем, обеспечивая получение данных с высокой временной

частотой и локальной пространственной привязкой.

В то же время использование IoT-инфраструктуры приводит к появлению дополнительных рисков, связанных с качеством и достоверностью поступающих данных. К таким рискам относятся:

- шумы и сбои измерительных устройств;
- потери пакетов данных при передаче;
- некорректные показания вследствие аппаратных ошибок;
- аномалии, вызванные внешним воздействием или компрометацией устройств.

В условиях распределённых IoT-систем указанные факторы приводят к искажению входных данных, что

непосредственно влияет на качество последующего анализа и прогнозирования показателей состояния растительности, включая индекс NDVI.

Для повышения надёжности системы мониторинга целесообразно применение методов интеллектуального анализа данных для выявления аномалий в потоках IoT-данных. В рамках данной задачи возможны различные постановки, включая бинарную классификацию (норма/аномалия) и обнаружение выбросов во временных рядах [25]. В качестве признаков пространства могут использоваться как исходные сенсорные измерения, так и их производные характеристики (скользящие средние, дисперсия, градиенты изменений).

На практике для решения задач обнаружения аномалий применяются следующие классы моделей:

- рекуррентные нейронные сети (LSTM, GRU), учитывающие временную динамику процессов [27-29];
- автоэнкодеры, выявляющие отклонения от типичного распределения данных [28];
- ансамблевые методы (например, Isolation Forest, Random Forest), ориентированные на выявление статистических аномалий [26].

Особую сложность представляет выраженный дисбаланс между нормальными и аномальными наблюдениями, характерный для IoT-сред, при котором редкие события (сбои, атаки, некорректные измерения) оказывают ограниченное влияние на процесс обучения моделей. Формально это проявляется в смещении функционала эмпирического риска:

$$\hat{R}(\theta) = \sum_{k=1}^K \pi_k \mathbb{E}_{x|y=k} [\ell(f_\theta(x), k)], \quad (2)$$

где π_k — априорная вероятность класса, $\ell(\cdot)$ — функция потерь. При $\pi_k \ll \pi_j$ вклад редких классов в оптимизацию оказывается существенно заниженным.

Полученные постановки задачи согласуются с подходами, используемыми в задачах обнаружения вторжений в IoT-сетях, где влияние межклассового дисбаланса рассматривается как один из ключевых факторов, снижающих чувствительность моделей к редким аномалиям и требующих применения генеративных методов для корректировки обучающих выборок [16, 17].

В таких условиях перспективным является применение генеративных подходов, основанных на аппроксимации условных распределений признаков вида $p(x|y)$. Это позволяет синтезировать данные для малочисленных классов и управляемо корректировать априорные вероятности без разрушения статистической структуры исходных данных. В отличие от классических методов oversampling, такие подходы обеспечивают более адекватное воспроизведение многомерных зависимостей и повышают чувствительность моделей к редким аномалиям.

Таким образом, в рамках интегрированной системы мониторинга сельскохозяйственных территорий IoT-компонент может рассматриваться не только как источник дополнительных данных, но и как объект

интеллектуального контроля. Это позволяет формировать многоуровневую архитектуру, включающую спутниковые данные, метеорологические наблюдения и IoT-датчики, дополненную контуром обеспечения достоверности данных на основе методов машинного обучения.

VI. ПОДГОТОВКА ДАННЫХ ДЛЯ ИССЛЕДОВАНИЯ

А. Источники и сбор спутниковых данных

Для формирования исходных данных был выполнен сравнительный анализ методов, спутниковых систем и платформ доступа к данным (см. таблицу 1). При этом были учтены такие факторы, как спектральный диапазон и периодичность съёмки, пространственное разрешение и свободный доступ к данным спутников. В результате анализа установлено, что заданным требованиям по спектральным характеристикам, пространственному разрешению и периодичности съёмки соответствует спутник Sentinel-2 проекта Copernicus, к тому же его снимки доступны через ряд специализированных платформ.

Дополнительно выполнен анализ платформ по критериям доступности, функциональности и удобства использования.

Базовой платформой для получения информации со спутников проекта Copernicus является Copernicus Data Space Ecosystem Browser. Однако ее использование на территории России требует применения VPN.

Широкое распространение получила платформа Google Earth Engine (США), обладающая развитым интерфейсом визуализации и обработки данных. Для ее использования нужен Google аккаунт. Формирование пользовательских запросов требует использования специализированных скриптовых языков (JS/Python), систематизация API слабая.

В последние годы наибольшее распространение среди исследователей получила платформа Sentinel Hub (Словения). Платформа имеет широкий функционал, работает в России без VPN. Sentinel Hub предоставляет 30 000 кредитов для исследовательской или учебной работы (не коммерческой).

Платформа Sentinel Hub также включает удобный EO Browser. В нем можно, как и в Copernicus Browser, выбирать слои и показатели (в т. ч. NDVI), ставить пины (отметки) на рабочие области снимков.

Кроме работы со снимками, Sentinel Hub позволяет работать со специальным набором API. API-инструментарий платформы включает восемь сервисов, обладающих унифицированной структурой и встроенным конструктором запросов, который интегрирует визуальный парсер интересующих областей и редактор скриптов на python, sh-py, curl со взаимной ретрансляцией.

Для получения значений NDVI достаточно воспользоваться инструментом API Sentinel Hub Statistical [18].

Конструктор запросов Statistical к базе снимков со спутников семейств Sentinel, Landsat и др. позволяет получить исторические данные с 2017 года. Statistical, в частности, дает возможность получить усредненный

временной ряд NDVI для определенной области земной поверхности по API запросу.

Таким образом, сравнительный анализ платформ показал, во-первых, отсутствие в свободном доступе функционала для прогнозирования показателей, полезных для фермерства, в частности индекса NDVI, по имеющимся входным данным. Во-вторых, при наличии доступа к снимкам, достаточно сложным у ряда платформ был механизм расчета либо экспорта текущих значений NDVI и других признаков. В-третьих, для доступа к части платформ требовался VPN.

В. Формирование NDVI временного ряда

С учётом проведённого анализа в качестве основного источника данных была выбрана платформа Sentinel Hub. Для составления API запроса в конструкторе Sentinel Hub Statistical необходимо ввести координаты

выбранного участка местности (см. рис. 2). В качестве тестовой территории выбраны окрестности г. Тихорецка Краснодарского края, поскольку данный регион характеризуется развитым сельскохозяйственным производством, является важным сельскохозяйственным районом РФ. Были заданы параметры запроса: период наблюдений (2017–2025 гг.) и интервал агрегации P10D (10 дней).

При выборе области исследования инструмент Statistical автоматически накладывает маску облачности на снимки, и с учетом маски рассчитывается средний NDVI. Площадь выбранного полигона составила 50.50 km². Запросы к архиву спутниковых данных формируются автоматически на основе заданных параметров.

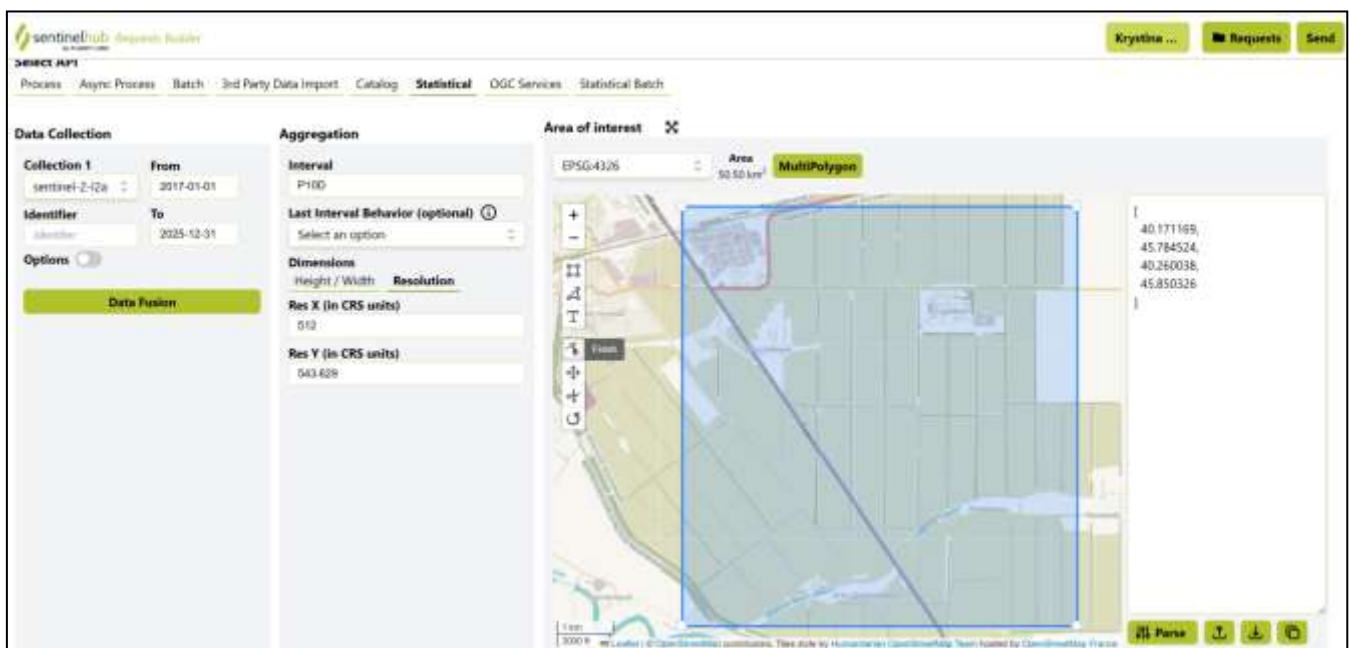


Рис.2: Платформа SentinelHub Requests Builder, API Statistical, для получения средних значений показателя NDVI для выбранной территории (по координатам и запросу)

Для извлечения данных был реализован базовый скрипт запроса, состоящий из двух функций (см. рис. 3).

Функция `setup()` получает на вход каналы B04 и B08, соответствующие видимому красному спектру и ближнему инфракрасному спектру, а также внутренний параметр `dataMask` для маскирования облачности снимков, и определяет, что выходным показателем будет NDVI.

Функция `evaluatePixel` обрабатывает каждый пиксель снимка и вычисляет NDVI для каждого пикселя снимка.

Полученные данные передаются в API Statistical, где вычисляются минимальный, максимальный, средний, стандартное отклонение для показателя NDVI для каждого снимка (см. рис. 4).

Параметр `sampleCount`, возвращаемый API Statistical, отражает количество валидных пикселей, участвующих в расчёте статистических характеристик NDVI с учетом маски облачности (`dataMask`).

```
function setup() {
  return {
    input: [{
      bands: [
        "B04",
        "B08",
        "dataMask"
      ]
    }],
    output: [{
      id: "ndvi",
      bands: 1
    }, {
      id: "dataMask",
      bands: 1
    }]
  };
}

function evaluatePixel(samples) {
  return {
    ndvi: [index(samples.B08, samples.B04)],
    dataMask: [samples.dataMask]
  };
}
```

Рис. 3: Скрипт для запроса видимого красного спектра и ближнего инфракрасного спектра и расчета NDVI



Рис. 4: Результаты расчета статистических параметров за 10-дневные интервалы доступного периода

Значение данного параметра может варьироваться в зависимости от доли облачного покрытия и геометрии выбранного полигона. Параметр может использоваться как индикатор надежности статистических оценок, так как малое значение sampleCount свидетельствует о значительном влиянии облачности или частичном перекрытии полигона.

Результаты могут быть экспортированы в формате JSON для последующей обработки.

С. Метеорологические данные

С учётом результатов ранее проведённых исследований [6], для обеспечения учета внешних факторов, влияющих на динамику вегетационного индекса, в модель дополнительно включены метеорологические параметры. Для формирования признаков, характеризующих условия внешней среды, использованы архивные данные метеорологических наблюдений по станции г. Тихорецка, полученные с ресурса RP5 (Reliable Prognosis) [19].

Ресурс представляет собой специализированную систему доступа к архивам метеорологических наблюдений в форматах SYNOP и METAR. Данные формируются на основе регулярных измерений наземных метеостанций с высокой временной дискретностью (до одного измерения каждые 3 часа), что обеспечивает высокую временную детализацию данных.

Выбор данного источника обусловлен следующими факторами:

- наличием архивных наблюдений с высокой временной детализацией;
- доступностью исторических данных за длительный период (с 2017 года);
- наличием полного набора метеорологических параметров, необходимых для моделирования (температура, влажность, давление, осадки, ветер);
- согласованностью формата данных и возможностью их автоматизированной обработки.

Из исходного массива отобраны признаки, оказывающие значимое влияние на состояние растительности и, соответственно, на значение индекса NDVI. К таким признакам отнесены:

- температура воздуха (средняя, минимальная и максимальная);
- относительная влажность воздуха;
- скорость ветра;
- количество осадков;
- температура точки росы;
- атмосферное давление.

Выбор указанных параметров обусловлен их физической интерпретацией в контексте процессов вегетации. Температура и влажность определяют интенсивность фотосинтеза и транспирации, осадки характеризуют водообеспеченность растений, а давление и ветер косвенно отражают изменение погодных режимов.

Параметр температуры почвы был исключен ввиду высокой доли пропусков (более 40%).

Д. Предобработка и синхронизация

Полученные из разных источников массивы данных требуют предобработки, включающей извлечение, форматирование, агрегацию и объединение с метеорологическими данными целевого показателя NDVI.

Поскольку метеорологические данные представлены с внутрисуточной дискретностью (наблюдения каждые 3 часа), для обеспечения сопоставимости с временным разрешением данных NDVI была выполнена агрегация до суточного уровня. Такой подход позволяет снизить влияние внутрисуточной вариабельности и согласовать масштаб данных с задачей прогнозирования.

Агрегация осуществлялась следующим образом:

- для температурных показателей вычислялись среднее, минимальное и максимальное значения за сутки;
- для относительной влажности, скорости ветра и атмосферного давления рассчитывались средние значения;
- для осадков вычислялась суммарная величина за сутки.

В результате была сформирована таблица ежедневных наблюдений, содержащая агрегированные значения метеорологических параметров.

Полученный временной ряд NDVI с шагом 10 дней был приведен к ежедневному разрешению. Для этого применена линейная интерполяция, позволяющая восстановить промежуточные точки временного ряда.

С целью уменьшения влияния шумов и выбросов дополнительно применено сглаживание временного ряда с использованием фильтра Савицкого — Голея. Параметры окна сглаживания подбирались эмпирически на основе минимизации шума без потери трендовой компоненты.

Е. Формирование итогового датасета

После предварительной обработки спутниковых данных и формирования временного ряда NDVI, полученного с использованием платформы Sentinel Hub, выполнено объединение спутниковых и метеорологических данных. Объединение осуществлялось по дате наблюдения с приведением данных к единому временному разрешению

(ежедневному). Для дат, по которым отсутствовали спутниковые наблюдения (вследствие облачности или отсутствия съёмки), значения NDVI были восстановлены с использованием интерполяции, что позволило сформировать непрерывный временной ряд.

Итоговый датасет представляет собой многомерный временной ряд, используемый далее в качестве

обучающей выборки для моделей прогнозирования. В датасете каждому дню соответствует вектор признаков, включающий метеорологические параметры и NDVI.

Предобработка и объединение данных реализованы с использованием Python.

Результаты слияния данных о погоде и NDVI представлены в датафрейме на рис. 5

date	T_mean	T_min	T_max	RH_mean	Wind_mean	Precip_sum	DewPoint_mean	SoilTemp_min	Pressure_mean	ndvi_smooth
2017-01-01	-2.8875	-3.8	-1.9	93.250	1.625	0.0	-3.8250	2.0	764.6375	0.168045
2017-01-02	2.8625	-1.4	4.9	77.000	4.000	0.0	-0.9375	2.0	762.8625	0.168045
2017-01-03	4.7750	1.8	7.2	70.000	4.125	0.0	-0.3625	2.0	761.2000	0.168045
2017-01-04	0.3250	-1.0	3.7	87.250	2.000	0.3	-1.5875	2.0	762.4125	0.168045
2017-01-05	3.1125	0.8	5.9	74.875	2.750	0.0	-1.1125	2.0	759.6875	0.168045
...	...	...	...	...	...	...	...	...	...	...
2025-12-27	-1.1500	-2.9	0.1	86.875	1.250	2.0	-3.0875	9.0	756.4250	0.179405
2025-12-28	-0.2625	-1.1	0.9	88.125	2.500	0.2	-2.0500	9.0	749.9000	0.150688
2025-12-29	-0.6000	-3.6	0.9	86.875	2.250	2.0	-2.5375	9.0	748.9375	0.104433
2025-12-30	-0.3375	-1.7	1.3	86.000	3.000	4.0	-2.4875	9.0	752.8500	0.040640
2025-12-31	-2.1625	-4.0	-0.1	91.625	0.375	5.0	-3.3750	9.0	754.1375	-0.040690

Рис.5: Итоговый объединенный датасет метеорологических параметров и значений NDVI для выбранной территории

VII. АРХИТЕКТУРА МОДЕЛЕЙ И МЕТОДИКА ОБУЧЕНИЯ

Для решения задачи прогнозирования временного ряда индекса NDVI была выбрана архитектура рекуррентной нейронной сети типа Long Short-Term Memory, поскольку данный класс моделей эффективно учитывает временные зависимости и широко применяется при анализе динамических процессов [22-23]. Кроме LSTM, была реализована альтернативная модель — градиентный бустинг [24].

В дополнение к концептуальной модели (рис. 1), отражающей общий конвейер обработки данных, представлена архитектура системы прогнозирования NDVI как совокупность функциональных компонентов.

Архитектура (рис. 6) включает этапы сбора, предобработки и синхронизации данных, формирование обучающего набора и модуль моделирования.

Результаты моделирования передаются в блок оценки качества, где выполняется расчет метрик и сравнительный анализ, после чего прогнозные значения могут быть использованы в прикладных задачах мониторинга и поддержки принятия решений.

Входные данные моделей — датасет, полученный в результате объединения спутниковых и метеорологических наблюдений и являющийся многомерным временным рядом. Каждому моменту времени соответствует вектор признаков, включающий температурные характеристики, влажность, скорость ветра, атмосферное давление и количество осадков. Таким образом, вход модели может быть представлен в виде тензора размерности

$$X \in \mathbb{R}^{N \times T \times F} \quad (3)$$

где N — число наблюдений, T — длина временного окна, F — количество признаков.

В рамках исследования использовано минимальное временное окно ($T = 1$). Несмотря на нетипичность такого выбора для рекуррентных моделей, он является обоснованным вследствие предварительной агрегации и сглаживания временных рядов на этапе предобработки. В результате временной ряд приобретает квазистационарный характер, а задача прогнозирования в значительной степени определяется автокорреляционной структурой самого ряда NDVI. В результате основной вклад в прогноз вносят лаговые значения NDVI, тогда как влияние внешних факторов носит вспомогательный характер.

LSTM-модель включает два последовательных слоя по 50 нейронов каждый. Первый слой возвращает последовательность скрытых состояний, обеспечивая передачу информации на следующий уровень, тогда как второй слой формирует компактное представление входных данных. Выходной слой реализован в виде полносвязного слоя с линейной функцией активации и формирует итоговое прогнозное значение NDVI.

Формально модель можно представить в виде отображения

$$\hat{y}_t = f(X_t, \theta), \quad (4)$$

где X_t — вектор входных признаков, а θ — параметры

модели.

Динамика LSTM-слоя задается рекуррентным обновлением скрытого состояния

$$h_t = LSTM(X_t, h_{t-1}), \quad (5)$$

что позволяет учитывать накопленную информацию о процессе [22-23].

Обучение выполнялось с использованием функции потерь MSE и оптимизатора Adam. Для повышения устойчивости модели применялась нормализация входных признаков, а также механизм ранней остановки, основанный на контроле ошибки на валидационной выборке. Обучающая выборка формировалась в соотношении 80% от общего объема данных, оставшаяся часть использовалась для тестирования.

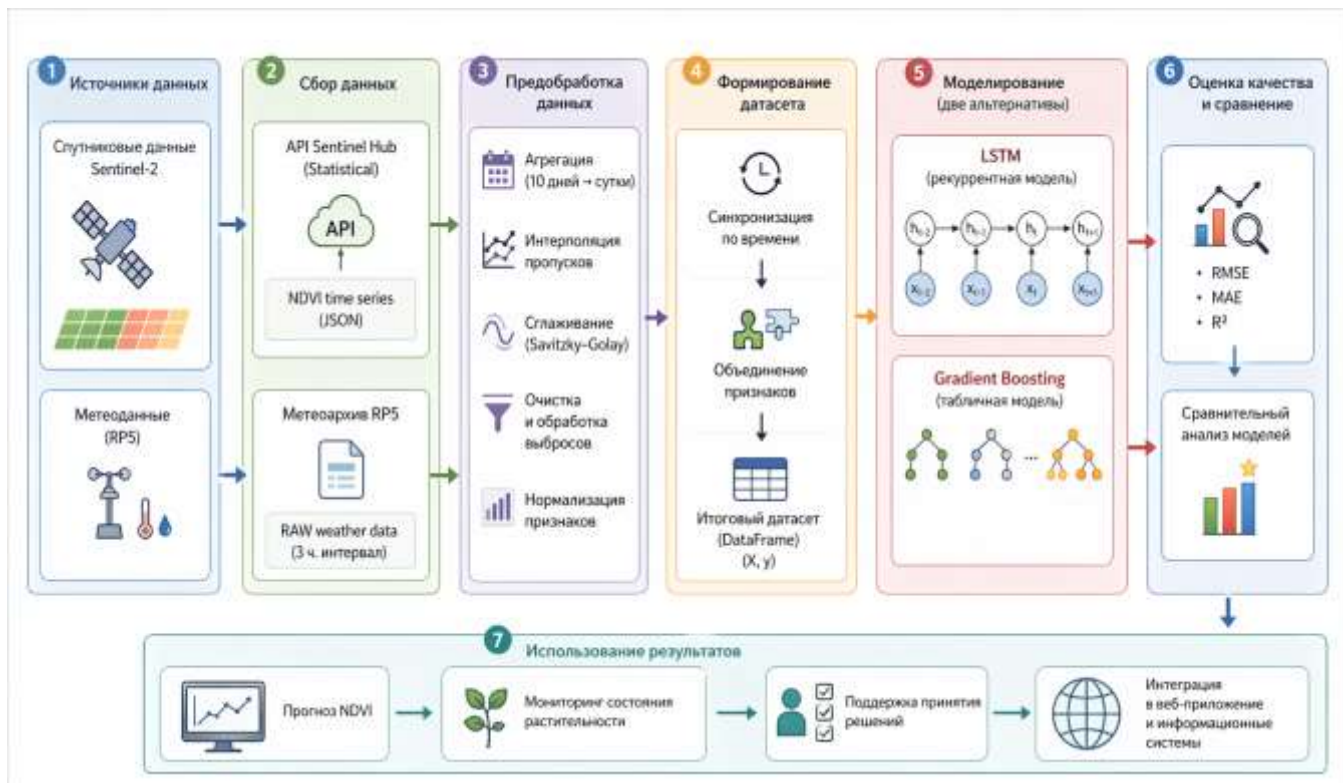


Рис.6: Архитектура системы обработки и прогнозирования индекса NDVI на основе спутниковых и метеорологических данных

В качестве альтернативного подхода была реализована модель градиентного бустинга, использующая табличное представление данных без явного учета временной структуры. Такая модель представляет собой композицию решающих деревьев

$$\hat{y}_t = \sum_{m=1}^M \gamma_m h_m(X_t), \quad (6)$$

где h_m — базовые модели, а γ_m — их веса [24].

Это позволило выполнить корректное сравнительное исследование эффективности различных классов алгоритмов в задаче прогнозирования NDVI.

Предложенный подход отличается ориентацией на предварительно агрегированные и сглаженные временные ряды, что обеспечивает снижение уровня шумов и устойчивость моделей к пропускам наблюдений. Это позволяет эффективно применять как рекуррентные, так и ансамблевые методы в рамках единого конвейера обработки данных.

VIII. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНОГО ИССЛЕДОВАНИЯ

Экспериментальное исследование направлено на

сравнительный анализ качества прогнозирования индекса NDVI с использованием рекуррентной нейронной сети LSTM и модели градиентного бустинга. Оценка выполнялась на отложенной тестовой выборке, не использовавшейся в процессе обучения, что обеспечивает объективность оценки.

В качестве основной метрики использовалась среднеквадратичная ошибка (RMSE), позволяющая интерпретировать величину отклонения прогноза в тех же единицах, что и целевая переменная. Дополнительно были рассчитаны средняя абсолютная ошибка (MAE) и коэффициент детерминации R^2 , что позволяет более полно оценить качество моделей с точки зрения устойчивости к выбросам и способности объяснять вариацию данных.

Количественные результаты сравнительного анализа моделей представлены в таблице 2. Полученные результаты показывают, что модель градиентного бустинга демонстрирует существенно более низкие значения ошибок по метрикам RMSE и MAE, а также более высокое значение коэффициента детерминации R^2 по сравнению с моделью LSTM.

Таблица 2: Сравнительные результаты моделей прогнозирования NDVI

Модель	RMSE	MAE	R ²
LSTM	0.124	0.094	0.423
Gradient Boosting	0.026	0.019	0.974
Gradient Boosting (без лагов)	0.127	0.097	0.388

Результаты сравнительного анализа представлены на рис. 7.

Значение коэффициента детерминации R² для модели градиентного бустинга близко к единице, что свидетельствует о высокой степени согласования прогнозных и фактических значений.

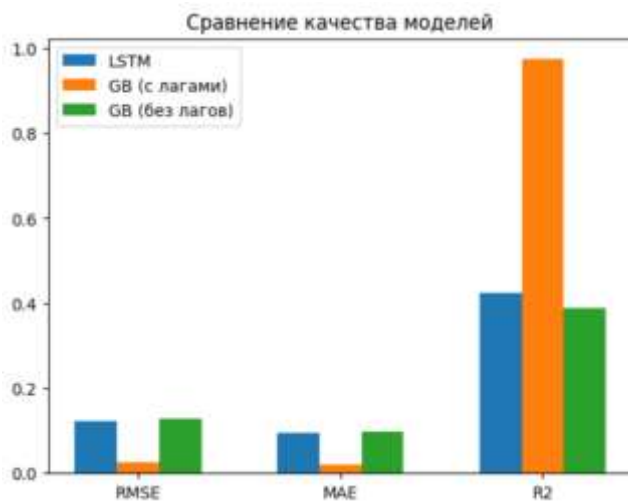


Рис. 7: Сравнение моделей LSTM и градиентного бустинга (с лагами и без) по метрикам RMSE, MAE и R²

Преимущество ансамблевой модели обусловлено особенностями предварительной обработки данных. Агрегация временных рядов и их сглаживание приводят к снижению выраженности временных зависимостей, в результате чего задача приобретает характер регрессии по табличным данным. В таких условиях модели градиентного бустинга, эффективно работающие с нелинейными зависимостями признаков, оказываются более предпочтительными по сравнению с рекуррентными архитектурами.

В то же время модель LSTM демонстрирует сопоставимую динамику изменения предсказаний и сохраняет потенциал для повышения качества при использовании более длинных временных окон или менее агрегированных исходных данных. Это свидетельствует о потенциале дальнейшего развития подхода в направлении более детального учета временной структуры.

Анализ важности признаков показывает (рис. 8), что доминирующий вклад в прогноз вносит лаговое значение NDVI с минимальным сдвигом (ndvi_lag_1), доля которого существенно превышает вклад остальных признаков. Это указывает на сильную автокорреляцию временного ряда и предсказуемость показателя на основе его предыдущих значений.

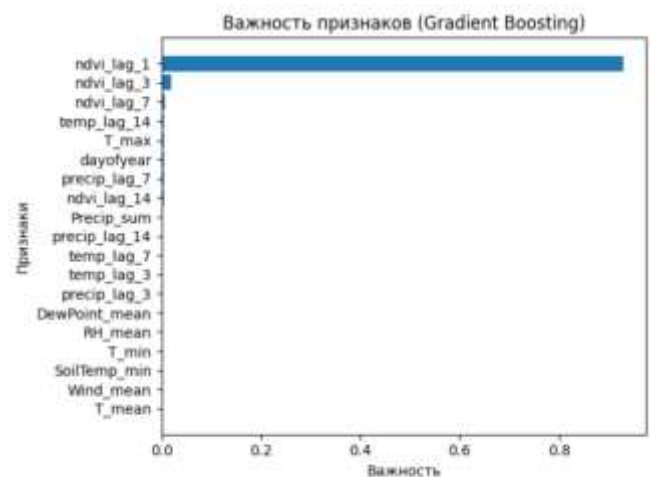


Рис. 8: Важность признаков в модели градиентного бустинга

Вклад метеорологических параметров существенно ниже, что свидетельствует о том, что в рамках сформированного датасета влияние погодных факторов носит вторичный характер по сравнению с собственной динамикой NDVI.

Качество прогнозирования существенно зависит от используемой структуры признаков. Модель градиентного бустинга, использующая исключительно метеорологические параметры (см. таблицу 2, модель «Gradient Boosting (без лагов)»), демонстрирует более высокую ошибку (RMSE $\approx$ 0.127), что указывает на ограниченную способность погодных факторов в одиночку объяснять динамику NDVI.

Введение лаговых признаков целевой переменной приводит к значительному снижению ошибки, однако при этом задача фактически переходит в постановку авторегрессионного прогнозирования. Таким образом, выделяются два режима работы модели: прогноз на основе внешних факторов и прогноз на основе инерционной динамики временного ряда.

IX. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полученные результаты показывают, что эффективность применяемых моделей в значительной степени определяется структурой исходных данных и характером предварительной обработки. В рассматриваемом случае агрегация временных рядов и применение методов сглаживания приводят к уменьшению выраженности временных зависимостей, вследствие чего преимущества моделей, ориентированных на последовательности, оказываются менее значимыми.

В этих условиях более простые и интерпретируемые модели, такие как градиентный бустинг, оказываются конкурентоспособными и даже превосходят нейросетевые подходы. Однако данный результат не следует рассматривать как универсальный, поскольку при переходе к более детализированным данным, включая исходные спутниковые изображения или последовательности с большим числом временных шагов, ситуация может существенно измениться.

Результаты также показывают, что использование

агрегированных временных рядов смещает акцент модели с выявления причинно-следственных связей на воспроизведение динамики. В этом случае модель становится эффективным инструментом краткосрочного прогноза, однако ее интерпретационные возможности в части влияния внешних факторов оказываются ограниченными.

Х. ЗАКЛЮЧЕНИЕ

Предложен подход к мониторингу состояния растительности на основе анализа и прогнозирования NDVI с использованием спутниковых данных и метеорологических наблюдений. Разработана методика формирования обучающего набора данных, включающая интеграцию разнородных источников информации, интерполяцию пропусков и сглаживание временных рядов.

На основе сформированного датасета реализованы и исследованы две модели прогнозирования — рекуррентная нейронная сеть LSTM и модель градиентного бустинга. Экспериментальное сравнение показало, что в условиях агрегированных временных рядов более высокую точность демонстрируют ансамблевые методы, тогда как применение рекуррентных архитектур требует более детализированных входных данных.

Результаты подтверждают возможность применения предложенного подхода для задач мониторинга состояния сельскохозяйственных угодий и поддержки принятия решений. Разработанная модель может быть адаптирована для других регионов при наличии сопоставимых климатических условий и доступности спутниковых и метеорологических данных.

Установлено, что при использовании лаговых признаков NDVI задача прогнозирования в значительной степени сводится к авторегрессионной модели, что ограничивает возможность интерпретации влияния метеорологических факторов. Это требует более строгого разделения сценариев применения моделей: краткосрочного прогноза на основе исторических значений и прогноза, основанного на внешних воздействиях.

Перспективным направлением дальнейших исследований является переход к гибридным архитектурам, объединяющим пространственные и временные признаки, а также отказ от избыточной агрегации данных, позволяющий в полной мере использовать потенциал моделей глубокого обучения.

БИБЛИОГРАФИЯ

- [1] Труфляк Е. В. Основные элементы системы точного земледелия / Е. В. Труфляк. — Краснодар : КубГАУ, 2016. — 39 с. URL: <https://kubsau.ru/upload/foresight/elements.pdf>
- [2] Тематическое дешифрирование и интерпретация космических снимков среднего и высокого пространственного разрешения: учебное пособие для студентов, обучающихся по направлению «Картография и геоинформатика» / А. Н. Шихов, А. П. Герасимов, А. И. Пономарчук, Е. С. Перминова ; Перм. гос. нац. исслед. ун-т. — Электрон. дан. (49,6 Мб). — Пермь, 2020. — 191 с. : ил. - Текст : электронный. URL: <https://elis.psu.ru/node/642172>
- [3] NDVI vs ASF // Михаил Воробьев, ГлавАгроном. URL: <https://glavagronom.ru/articles/ndvi-vs-asf-ili-pochemu-kompanii-razrabatyvayut-svoy-metod-ocenki-posevov>
- [4] Распоряжение Правительства Российской Федерации от 31.12.2020 г. № 3684-р «Об утверждении Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021 - 2030 годы)» URL: <http://government.ru/docs/all/132190/>
- [5] Родимцев С.А., Павловская Н.Е., Вершинин С.В., Зелюкин В.И., Горькова И.В. Моделирование условий вегетации с использованием отклонений текущих значений NDVI от среднеемноголетних показателей // Сельскохозяйственная биология, 2022, том 57, № 3, с. 591-603. DOI: 10.15389/agrobiology.2022.3.591rus
- [6] Шурлаева Е.А., Барышев Д.А., Зубанков А.С., Розалиев В.Л. Прогнозирование вегетационного индекса растительности сельскохозяйственных угодий Волгоградской области с помощью нейросетевых методов // Инженерный вестник Дона, №5 (2022). URL: <https://ivdon.ru/ru/magazine/archive/n5y2022/7664>
- [7] Cui, C., Zhang, W., Hong, Z., Meng, L. Forecasting NDVI in multiple complex areas using neural network techniques combined feature engineering. International Journal of Digital Earth, 2020, 13(12), 1733–1749. <https://doi.org/10.1080/17538947.2020.1808718>
- [8] Yli-Heikkilä, M., Wittke, S., Luotamo, M., Puttonen, E., Sulkava, M., Pellikka, P., Heiskanen, J., Klami, A. Scalable Crop Yield Prediction with Sentinel-2 Time Series and Temporal Convolutional Network. Remote Sensing, 2022, 14, no. 17: 4193. URL: <https://www.mdpi.com/2072-4292/14/17/4193>
- [9] Wittke, S., Fouilloux, A., Lehti, P., Varho, J., Kivimäki, A., Karhu, M., Karjalainen, M., Vaaja, M., Puttonen, E. EODIE — Earth Observation Data Information Extractor. SoftwareX, 2023, vol. 23: 101421. DOI: 10.1016/j.softx.2023.101421
- [10] Quinton, F., Landrieu, L. Crop Rotation Modeling for Deep Learning-Based Parcel Classification from Satellite Time Series. Remote Sensing, 2021, 13, no. 22: 4599. <https://doi.org/10.3390/rs13224599>
- [11] Nyborg, J., Pelletier, Ch., Assent, I. Generalized Classification of Satellite Image Time Series with Thermal Positional Encoding. // In proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. <https://doi.org/10.48550/arXiv.2203.09175>
- [12] Wenger, R., Puissant, A., Weber, J., Idoumghar, L., Forestier, G. Multimodal and Multitemporal Land Use/Land Cover Semantic Segmentation on Sentinel-1 and Sentinel-2 Imagery: An Application on a MultiSenGE Dataset. Remote Sensing, 2023, 15, no. 1: 151. <https://doi.org/10.3390/rs15010151>
- [13] Guo, L., Zhao, S., Gao, J., Zhang, H., Zou, Y., Xiao, X. A Novel Workflow for Crop Type Mapping with a Time Series of Synthetic Aperture Radar and Optical Images in the Google Earth Engine. Remote Sensing, 2022, 14, no. 21: 5458. <https://doi.org/10.3390/rs14215458>
- [14] Index DataBase. A database for remote sensing indices. Search indices for “vegetation” URL: <https://www.indexdatabase.de/search/i-search.php?s=vegetation>
- [15] Huang, S., Tang, L., Hupy, J.P. et al. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. J. For. Res. 32, 1–6 (2021). <https://doi.org/10.1007/s11676-020-01155-1>
- [16] Xu, L., Skoularidou, M., Cuesta-Infante, A., Veeramachaneni, K. Modeling Tabular Data using Conditional GAN. NeurIPS, 2019. DOI: 10.48550/arXiv.1907.00503
- [17] Alabdulwahab S, Kim Y-T, Seo A, Son Y. Generating Synthetic Dataset for ML-Based IDS Using CTGAN and Feature Selection to Protect Smart IoT Environments. Applied Sciences. 2023; 13(19):10951. DOI: 10.3390/app131910951
- [18] Copernicus Request Builder URL: <https://shapps.dataspace.copernicus.eu/requests-builder>
- [19] RP5 Расписание Погоды. “Архив погоды в Тихорецке.” URL: https://rp5.ru/Weather_archive_in_Tikhoretsk
- [20] Zhu X. X. et al. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources // IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 4, pp. 8-36, Dec. 2017. DOI: 10.1109/MGRS.2017.2762307
- [21] Reichstein, M., Camps-Valls, G., Stevens, B. et al. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204 (2019). <https://doi.org/10.1038/s41586-019-0912-1>
- [22] Sepp Hochreiter, Jürgen Schmidhuber; Long Short-Term Memory. Neural Comput 1997; 9 (8): 1735–1780. doi: <https://doi.org/10.1162/neco.1997.9.8.1735>
- [23] Bryan Lim, Stefan Zohren; Time-series forecasting with deep learning: a survey. Philos Trans A Math Phys Eng Sci 5 April 2021; 379 (2194): 20200209. <https://doi.org/10.1098/rsta.2020.0209>

- [24] Chen, T., Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. 2016. <https://doi.org/10.1145/2939672.2939785>
- [25] Chandola, V., Banerjee, A., Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv.* 41, 3, Article 15 (July 2009), 58 pages. <https://doi.org/10.1145/1541880.1541882>
- [26] Thamaraiselvi Dr. R. et al. Attack and Anomaly Detection in IoT Networks using Machine Learning. *International Journal of Computer Science and Mobile Computing*, Vol.9 Issue.10, October- 2020, pg. 95-103. DOI: 10.47760/ijcsmc.2020.v09i10.012
- [27] Dai, L.E., Wang, X., Xu, S.B. (2024). A Deep Learning Based Anomaly Detection Model for IoT Networks. In: Dong, J., Zhang, L., Cheng, D. (eds) *Proceedings of the 2nd International Conference on Internet of Things, Communication and Intelligent Technology. IoTCIT 2023. Lecture Notes in Electrical Engineering*, vol 1197. Springer, Singapore. https://doi.org/10.1007/978-981-97-2757-5_20
- [28] Fu A, Li J. Anomaly Behavior Detection Based on Deep Learning in an IoT Environment. *Sensors (Basel)*. 2025 Dec 15;25(24):7605. DOI: 10.3390/s25247605
- [29] Singh, A., Singh, S., Alam M. N., Singh, G. Deep Learning for Anomaly Detection in IoT Systems: Techniques, Applications, and Future Directions. *International Journal for Multidisciplinary Research (IJFMR)*, Volume 6, Issue 4, July-August 2024. DOI: <https://doi.org/10.36948/ijfmr.2024.v06i04.24601>
- [30] Gao, P., Du, W., Lei, Q. et al. NDVI Forecasting Model Based on the Combination of Time Series Decomposition and CNN – LSTM. *Water Resour Manage* 37, 1481–1497 (2023). <https://doi.org/10.1007/s11269-022-03419-3>
- [31] Gaber, Sh. Kh., Azmy, M., Alakkari, Kh., Abotaleb, M. El-kenawy, El-S. M. Deep Learning in IoT: An LSTM Approach for NDVI Forecasting. 2023, 1-7. DOI: 10.1109/ICEEM58740.2023.10319616

NDVI Forecasting Based on the Integration of Satellite and Meteorological Data for Agricultural Land Monitoring

K.A. Makovejchuk, A.V. Olifirov, Y.T. Makovejchuk

Abstract - This paper addresses the problem of monitoring vegetation conditions in agricultural areas through the analysis and forecasting of the NDVI index using Sentinel-2 satellite data and meteorological observations. The relevance of integrating remote sensing data with measurements obtained via Internet of Things (IoT) technologies is substantiated.

A data processing pipeline is developed, including extraction, synchronization, and preprocessing of satellite and meteorological time series (missing value interpolation, smoothing, and aggregation), enabling the construction of a robust training dataset.

Based on this dataset, forecasting models are implemented and analyzed, including gradient boosting and a recurrent neural network (LSTM). A comparative study shows that when using aggregated time series, ensemble methods provide higher accuracy compared to recurrent models.

It is established that lagged NDVI values contribute most significantly to prediction performance, indicating the dominant role of the autocorrelation structure of the time series and highlighting limitations in interpreting the influence of external factors.

The proposed approach can be applied in the development of intelligent monitoring and decision support systems in agriculture, including those based on IoT infrastructures.

Keywords - NDVI, time series, remote sensing, data integration, meteorological data..

REFERENCES

- [1] Truflyak, Evgeny. Osnovnye elementy sistemy tochnogo zemledeliya [Basic Elements of Precision Agriculture System]. Krasnodar: KubGAU, 2016. URL: <https://kubsau.ru/upload/foresight/elements.pdf>
- [2] Shikhov, Alexander, et al. Tematicheskoe deshifirovanie i interpretatsiya kosmicheskikh snimkov srednego i vysokogo prostranstvennogo razresheniya [Thematic Interpretation of Medium- and High-Resolution Satellite Images]. Perm: Perm State National Research University, 2020. URL: <https://elis.psu.ru/node/642172>
- [3] Vorobyev, Mikhail. "NDVI vs ASF." GlavAgronom. URL: <https://glavagronom.ru/articles/ndvi-vs-asf-ili-pochemu-kompanii-razrabatyvayut-svoy-metod-ocenki-posevov>
- [4] Government of the Russian Federation. "On Approval of the Program of Fundamental Scientific Research in the Russian Federation for the Long-Term Period (2021–2030)." Decree No. 3684-r, 2020. URL: <http://government.ru/docs/all/132190/>
- [5] Rodimtsev, Sergey, et al. "Modeling Vegetation Conditions Using Deviations of Current NDVI Values from Long-Term Averages." Sel'skokhozyaystvennaya biologiya [Agricultural Biology], 57(3), 2022: 591–603. DOI: 10.15389/agrobiologiya.2022.3.591rus
- [6] Shurlaeva, Elena, et al. "Forecasting the Vegetation Index of Agricultural Lands in the Volgograd Region Using Neural Network Methods." Inzhenerny vestnik Dona [Engineering Journal of Don], no. 5 (2022). URL: <https://ivdon.ru/ru/magazine/archive/n5y2022/7664>
- [7] Cui, C., Zhang, W., Hong, Z., Meng, L. Forecasting NDVI in multiple complex areas using neural network techniques combined feature engineering. International Journal of Digital Earth, 2020, 13(12), 1733–1749. <https://doi.org/10.1080/17538947.2020.1808718>
- [8] Yli-Heikkilä, M., Wittke, S., Luotamo, M., Puttonen, E., Sulkava, M., Pellikka, P., Heiskanen, J., Klami, A. Scalable Crop Yield Prediction with Sentinel-2 Time Series and Temporal Convolutional Network. Remote

- Sensing, 2022, 14, no. 17: 4193. URL: <https://www.mdpi.com/2072-4292/14/17/4193>
- [9] Wittke, S., Fouilloux, A., Lehti, P., Varho, J., Kivimäki, A., Karhu, M., Karjalainen, M., Vaaja, M., Puttonen, E. EODIE — Earth Observation Data Information Extractor. SoftwareX, 2023, vol. 23: 101421. DOI: 10.1016/j.softx.2023.101421
- [10] Quinton, F., Landrieu, L. Crop Rotation Modeling for Deep Learning-Based Parcel Classification from Satellite Time Series. Remote Sensing, 2021, 13, no. 22: 4599. <https://doi.org/10.3390/rs13224599>
- [11] Nyborg, J., Pelletier, Ch., Assent, I. Generalized Classification of Satellite Image Time Series with Thermal Positional Encoding. // In proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. <https://doi.org/10.48550/arXiv.2203.09175>
- [12] Wenger, R., Puissant, A., Weber, J., Idoumghar, L., Forestier, G. Multimodal and Multitemporal Land Use/Land Cover Semantic Segmentation on Sentinel-1 and Sentinel-2 Imagery: An Application on a MultiSenGE Dataset. Remote Sensing, 2023, 15, no. 1: 151. <https://doi.org/10.3390/rs15010151>
- [13] Guo, L., Zhao, S., Gao, J., Zhang, H., Zou, Y., Xiao, X. A Novel Workflow for Crop Type Mapping with a Time Series of Synthetic Aperture Radar and Optical Images in the Google Earth Engine. Remote Sensing, 2022, 14, no. 21: 5458. <https://doi.org/10.3390/rs14215458>
- [14] Index DataBase. A database for remote sensing indices. Search indices for "vegetation" URL: <https://www.indexdatabase.de/search/i-search.php?s=vegetation>
- [15] Huang, S., Tang, L., Hupy, J.P. et al. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. J. For. Res. 32, 1–6 (2021). <https://doi.org/10.1007/s11676-020-01155-1>
- [16] Xu, L., Skoularidou, M., Cuesta-Infante, A., Veeramachaneni, K. Modeling Tabular Data using Conditional GAN. NeurIPS, 2019. DOI: 10.48550/arXiv.1907.00503
- [17] Alabdulwahab S, Kim Y-T, Seo A, Son Y. Generating Synthetic Dataset for ML-Based IDS Using CTGAN and Feature Selection to Protect Smart IoT Environments. Applied Sciences. 2023; 13(19):10951. DOI: 10.3390/app131910951
- [18] Copernicus Request Builder URL: <https://shapps.dataspace.copernicus.eu/requests-builder>
- [19] RP5 Weather Archive. "Weather Archive in Tikhoretsk." URL: https://rp5.ru/Weather_archive_in_Tikhoretsk
- [20] Zhu X. X. et al. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources // IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 4, pp. 8–36, Dec. 2017. DOI: 10.1109/MGRS.2017.2762307
- [21] Reichstein, M., Camps-Valls, G., Stevens, B. et al. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204 (2019). <https://doi.org/10.1038/s41586-019-0912-1>
- [22] Sepp Hochreiter, Jürgen Schmidhuber; Long Short-Term Memory. Neural Comput 1997; 9 (8): 1735–1780. doi: <https://doi.org/10.1162/neco.1997.9.8.1735>
- [23] Bryan Lim, Stefan Zohren; Time-series forecasting with deep learning: a survey. Philos Trans A Math Phys Eng Sci 5 April 2021; 379 (2194): 20200209. <https://doi.org/10.1098/rsta.2020.0209>
- [24] Chen, T., Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. 2016. <https://doi.org/10.1145/2939672.2939785>
- [25] Chandola, V., Banerjee, A., Kumar, V. Anomaly detection: A survey. ACM Comput. Surv. 41, 3, Article 15 (July 2009), 58 pages. <https://doi.org/10.1145/1541880.1541882>
- [26] Thamaraiselvi Dr. R. et al. Attack and Anomaly Detection in IoT Networks using Machine Learning. International Journal of Computer

Science and Mobile Computing, Vol.9 Issue.10, October- 2020, pg. 95-103. DOI: 10.47760/ijcsmc.2020.v09i10.012

[27] Dai, L.E., Wang, X., Xu, S.B. (2024). A Deep Learning Based Anomaly Detection Model for IoT Networks. In: Dong, J., Zhang, L., Cheng, D. (eds) Proceedings of the 2nd International Conference on Internet of Things, Communication and Intelligent Technology. IoTICIT 2023. Lecture Notes in Electrical Engineering, vol 1197. Springer, Singapore. https://doi.org/10.1007/978-981-97-2757-5_20

[28] Fu A, Li J. Anomaly Behavior Detection Based on Deep Learning in an IoT Environment. Sensors (Basel). 2025 Dec 15;25(24):7605. DOI: 10.3390/s25247605

[29] Singh, A., Singh, S., Alam M. N., Singh, G. Deep Learning for Anomaly Detection in IoT Systems: Techniques, Applications, and Future Directions. International Journal for Multidisciplinary Research (IJFMR), Volume 6, Issue 4, July-August 2024. DOI: <https://doi.org/10.36948/ijfmr.2024.v06i04.24601>

[30] Gao, P., Du, W., Lei, Q. et al. NDVI Forecasting Model Based on the Combination of Time Series Decomposition and CNN – LSTM. Water Resour Manage 37, 1481–1497 (2023). <https://doi.org/10.1007/s11269-022-03419-3>

[31] Gaber, Sh. Kh., Azmy, M., Alakkari, Kh., Abotaleb, M. El-kenawy, El-S. M. Deep Learning in IoT: An LSTM Approach for NDVI Forecasting. 2023, 1-7. DOI: 10.1109/ICEEM58740.2023.10319616