НАУКИ ОБ АТМОСФЕРЕ И КЛИМАТЕ/ATMOSPHERIC AND CLIMATE SCIENCES

DOI: https://doi.org/10.60797/IRJ.2025.162.119

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В ПРОГНОЗИРОВАНИИ УРОЖАЙНОСТИ СЕЛЬХОЗКУЛЬТУР

Научная статья

Аджиева А.А.[1,] *, Шаповалов В.А.[2], Беккиев А.М.[3], Хитиева А.Ж.[4]
[1] ORCID : 0000-0002-1047-8417;
[2] ORCID : 0000-0002-9701-6820;
[1, 4] Кабардино-Балкарский государственный аграрный университет имени В. М. Кокова, Нальчик, Российская Федерация
[2, 3] Высокогорный геофизический институт, Нальчик, Российская Федерация

* Корреспондирующий автор (aida-adzhieva[at]mail.ru)

Аннотация

В статье выполнена систематизация и дана критическая оценка современных методов машинного обучения для прогнозирования урожайности сельскохозяйственных культур на основе агрометеорологических и дистанционных данных с акцентом на регионы Северного Кавказа. В рамках обзора проанализированы классы моделей (регрессионные, ансамблевые деревья, градиентный бустинг, глубокие нейронные сети CNN/LSTM/ConvLSTM), источники данных (метеостанции, реанализы, почвенно-агрохимические обследования, индексы растительности NDVI/EVI, орографические предикторы) и протоколы валидации.

Показано, что комплексирование наземных и спутниковых признаков, фазовая агрегация метеопараметров и пространственная стратификация существенно повышают точность прогнозов относительно классических регрессионных подходов, а методы XAI (SHAP/LIME) повышают доверие к результатам за счёт объяснимости. Для Северного Кавказа ключевыми предикторами выступают распределение осадков по критическим фазам, температурные аномалии, запасы продуктивной влаги и орография; целесообразны кластерные схемы по агроклиматическим зонам и интеграция вероятностных сезонных прогнозов погоды. Обсуждаются ограничения такие как качество данных, переобучение, переносимость в условиях климатических трендов и направления дальнейшего развития.

Ключевые слова: агрометеорология, урожайность, машинное обучение, XGBoost, Random Forest, CNN, LSTM, ConvLSTM, дистанционное зондирование, NDVI, SHAP, районирование, Северный Кавказ, сезонные прогнозы.

MACHINE LEARNING METHODS IN CROP YIELD FORECASTING

Research article

Adzhieva A.A.[1,] *, Shapovalov V.A.[2], Bekkiev A.M.[3], Khitieva A.Z.[4]
[1] ORCID : 0000-0002-1047-8417;
[2] ORCID : 0000-0002-9701-6820;
[1, 4] Kabardino-Balkar State Agrarian University, Nalchik, Russian Federation
[2, 3] High-Mountain Geophysical Institute, Nalchik, Russian Federation

* Corresponding author (aida-adzhieva[at]mail.ru)

Abstract

The article systematises and critically evaluates modern machine learning methods for predicting crop yields based on agrometeorological and remote sensing data, with a focus on the North Caucasus region. The review analyses model classes (regression, ensemble trees, gradient boosting, deep neural networks CNN/LSTM/ConvLSTM), data sources (weather stations, reanalyses, soil and agrochemical surveys, NDVI/EVI vegetation indices, orographic predictors) and validation protocols.

It has been shown that the combination of ground-based and satellite features, phase aggregation of meteorological parameters, and spatial stratification significantly improve the accuracy of forecasts compared to classical regression approaches, while XAI methods (SHAP/LIME) increase confidence in the results due to their explainability. For the North Caucasus, the key predictors are precipitation distribution across critical phases, temperature anomalies, productive moisture reserves, and orography; cluster schemes by agroclimatic zones and the integration of probabilistic seasonal weather forecasts are advisable. Limitations such as data quality, retraining, transferability in the context of climate trends, and directions for further development are discussed.

Keywords: agrometeorology, crop yield, machine learning, XGBoost, Random Forest, CNN, LSTM, ConvLSTM, remote sensing, NDVI, SHAP, zoning, North Caucasus, seasonal forecasts.

Введение

Колебания урожайности сельскохозяйственных культур в регионах с контрастными метеоусловиями и сложной орографией, таких как Северный Кавказ, определяются комплексным влиянием теплового режима, влагообеспеченности, солнечной радиации и применяемых агротехнологий [1]. Для аграрного сектора точные прогнозы урожайности имеют стратегическое значение: они позволяют планировать распределение ресурсов, оптимизировать систему удобрений и орошения, снижать риски для фермеров и агрохолдингов, а также поддерживать продовольственную безопасность региона и страны в целом.

Традиционные методы агрометеорологических прогнозов включают многомерные регрессионные модели и процессные модели роста растений (CERES-Maize, CERES-Wheat, Sirius, DSSAT/CSM). Эти подходы заложили методологическую основу, однако ограничены предположением линейности зависимостей, высокой требовательностью к параметризации и слабой переносимостью между агроклиматическими зонами [5], [7]. В условиях климатического дрейфа и учащения экстремальных явлений (засухи, периоды аномальной жары, неравномерное распределение осадков) данные ограничения становятся особенно критичными.

Развитие технологий дистанционного зондирования Земли (MODIS, Sentinel, «Электро»), появление специализированных сервисов («Вега») и доступность многоспектральных и реанализных данных создали возможность формирования обширных массивов гетерогенных предикторов. Одновременно прогресс в методах машинного обучения (Random Forest, градиентный бустинг) и глубоких нейронных сетях (CNN, LSTM, ConvLSTM) позволил извлекать сложные нелинейные зависимости, учитывать пространственно-временную динамику и адаптировать модели к локальным условиям [2], [3], [11]. Важным этапом стало внедрение методов объяснимого ИИ (SHAP, LIME), которые обеспечивают интерпретацию прогнозов и повышают доверие к результатам.

Актуальность использования машинного обучения обусловлена современными тенденциями: растущей неопределённостью климатических условий и высокой межгодовой изменчивостью урожайности [14]. На Северном Кавказе соседствуют засушливые степные зоны и увлажнённые предгорья, что требует моделей, учитывающих пространственную неоднородность агроландшафтов и чувствительность к критическим фазам вегетации. Традиционные статистические и процессные модели оказываются недостаточными, тогда как современные подходы машинного обучения позволяют интегрировать метеорологические, почвенные и спутниковые данные, строить адаптивные региональные прогнозы и формировать вероятностные сценарии урожайности [4], [6]. Это делает задачу развития и внедрения методов машинного обучения ключевой как для научного сообщества, так и для практики агропромышленного комплекса региона.

Методы и принципы исследования

Базовой физической метрикой сезонного развития сельскохозяйственных культур является сумма активных температур (Growing Degree Days, GDD):

$$\mathrm{GDD} = \sum_{i=1}^{n} \max\left(0, \frac{T_{\max,i}+T_{\min,i}}{2} - T_{\mathrm{base}}\right) \quad (1)$$

а для культур, чувствительных к длине светового дня (например, озимой пшеницы) — фототермическое время:

$$\mathrm{PTT} = \sum_{i=1}^{n} \left(T_i - T_{\mathrm{base}}\right) \cdot f(\mathrm{Photoperiod}_i) \quad (2)$$

где функция f учитывает длину светового дня [16], [17], [18].

В процессных моделях прирост биомассы часто формализуют через использование солнечной радиации:

$$\Delta W = \mathrm{RUE} \cdot \mathrm{PAR} \cdot \left(1 - e^{-k \cdot \mathrm{LAI}}\right) \quad (3)$$

где RUE — коэффициент использования радиации, PAR — фотосинтетически активная радиация, k — коэффициент экстинкции, LAI — листовой индекс [2].

В области статистики и машинного обучения переход от линейных моделей:

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \varepsilon \quad (4)$$

к ансамблевым методам и глубоким нейросетям значительно улучшил возможность учета нелинейных зависимостей [6], [14]:

Случайный лес (Random Forest):

$$\hat{Y} = \frac{1}{M} \sum_{m=1}^{M} T_m(X) \quad (5)$$

где $T_m(X)$ — отдельное дерево решений, M — количество деревьев в ансамбле,

Градиентный бустинг (Gradient Boosting):

$$F_m(x) = F_{m-1}(x) + \nu\, h_m(x) \quad (6)$$

где $F_m(x)$ — модель на шаге m, $h_m(x)$ — слабый ученик (обычно дерево решений), ν — темп обучения (learning rate).

Эти два подхода представляют разные философии построения ансамблей [10]:

Случайный лес использует бэггинг (bootstrap aggregating) — параллельное построение независимых деревьев на различных подвыборках данных с последующим усреднением их предсказаний. Это уменьшает дисперсию модели и повышает устойчивость к переобучению.

Градиентный бустинг использует последовательное обучение — каждое новое дерево строится для исправления ошибок предыдущих деревьев. Параметр ν (обычно 0,01-0,1) контролирует скорость обучения и предотвращает переобучение.

Использование нейронных сетей, в частности рекуррентные сети (Recurrent Neural Network, RNN), сети с долговременной и кратковременной памятью (Long Short-Term Memory, LSTM), сети содержащие рекуррентный блок с вентилем (Gated Recurrent Unit, GRU), сверточные нейронные сети (Convolutional Neural Networks, CNN), пространственно-временные ансамбли на основе сверточной сети с долгосрочной и краткосрочной памятью (Convolutional LSTM, ConvLSTM) позволяют учитывать сезонную динамику метеорологических данных и вегетационных индексов [8], [14], [15].

2.1. Используемые данные и предикторы

Метеопараметры. Суточные и декадные значения T_{min}, T_{max}, вид и интенсивность осадков, суммы активных температур (GDD), суммы осадков по фазам развития растений (кущение, выход в трубку, колошение, налив), индексы засушливости (гидротермический коэффициент Селянинова (ГТК), эвапорометрический коэффициент (Кэ) А.А. Скворцова, «коэффициент соответствия тепла и влаги» Д.И. Абрамовича, гидротермический показатель В.Г. Нестерова, коэффициентов увлажнения Н.В. Бова, Е.С. Улановой и др.), продолжительность засушливых и жарких периодов.

Почвенно-агрохимические данные. Содержание гумуса, pH, содержание азота, фосфора и калия (N-P-K), гранулометрический состав, влагоемкость. Их анализ предшествует внесению удобрений и выбору способов проведения орошения.

Данные дистанционного зондирования. Вегетационные индексы NDVI/EVI — их максимальные значения и интеграл за сезон, скорость нарастания биомассы, фотосинтетический потенциал; текстурные признаки со спутниковых снимков; орографические предикторы, например, высота, уклон, экспозиция и др. [2], [3], [6].

Основные ограничения связаны с неполнотой и неоднородностью исторических данных (особенно 90-х годов), риском переобучения на малых выборках, климатическим дрейфом и проблемами совместимости разнородных источников информации [11], [12]. Для их преодоления необходимы регуляризация, строгие протоколы валидации и локальная адаптация моделей.

Предобработка и агрегация данных включала:

– очистку от выбросов и заполнение пропусков;

– агрегацию погодных признаков по фазам развития растений;

– расчет скользящих статистик на окнах 7–30 суток для экстремальных значений;

– нормализацию или бинаризацию категориальных факторов (предшествующая культура, мероприятия по защите растений), сезонные фенологические метрики (даты начала и конца сезона вегетации, амплитуда, площадь под кривой).

Модели и методы:

1) фенологические модели. Формулы GDD/PTT и фотосинтетического прироста биомассы ΔW;

2) базовые модели. Множественная регрессия и регуляризованные обобщенные линейные модели LASSO/ElasticNet;

3) ансамблевые методы и бустинг. Random Forest; XGBoost/LightGBM/CatBoost градиентный бустинг, с функцией потерь $L = \sum \ell(y_i, F(x_i)) + \Omega(F)$ и регуляризацией Ω [16], [17];

4) глубокие нейросети. MLP для табличных данных; LSTM/GRU для временных рядов погоды и вегетационных индексов; CNN/ConvLSTM для пространственно-временных данных [8], [15];

5) интерпретация моделей (XAI). SHAP-анализ для объяснения вклада признаков [13];

6) гибридные подходы. Комбинация процессных моделей DSSAT/Sirius с корректорами систематических ошибок машинного обучения; ансамбли "бустинг + LSTM" [9], [18];

7) целевая переменная. Урожайность (т/га) по культурам и административным районам/хозяйствам;

8) метрики оценки: R^2, RMSE, MAE, MAPE;

9) валидация. Сквозная проверка по годам Leave-One-Year-Out (LOYO), пространственно-стратифицированная кросс-валидация, тестирование на экстремальные годы (засухи, сильная жара) [16];

10) учет особенностей региона. Разделение Северного Кавказа на подзоны (засушливые степи востока, предгорные степи, увлажненные предгорья и долины) с обучением отдельных моделей для каждой зоны или введением индикаторов зоны. Учитывая особенности Северо-Кавказского региона с его разнообразием микроклиматов и рисками (жара, засуха, град) необходимо сочетать биофизические модели с методами машинного обучения с учетом пространственной стратификации.

Основные результаты

Проведённый сравнительный анализ показал преимущество ансамблевых методов и гибридных подходов над линейными базовыми моделями. Включение нелинейных зависимостей позволило снизить RMSE на 15–35% и увеличить коэффициент детерминации R^2 до диапазона 0,75–0,90 в зависимости от культуры, горизонта прогноза и набора признаков. Особенно устойчивым оказалось использование вегетационных индексов NDVI/EVI и фенологических метрик: добавление этих предикторов к метеопараметрам (в моделях XGBoost и CatBoost) повышало точность прогноза на 0,05–0,15 по R^2.

Динамические архитектуры (LSTM, GRU) показали преимущество на среднесрочных горизонтах прогноза (T2–T3), благодаря учёту временной структуры погодных рядов и фаз вегетации. Пространственно-временные модели ConvLSTM дополнительно учитывали градиенты увлажнения и орографические факторы. Они показали лучшую адаптацию в условиях Северного Кавказа при учёте орографических и почвенных инвариантов (высота над уровнем моря, уклон, типы почв, влагоёмкость, ёмкость катионного обмена), а также при использовании калиброванных и стабильных индексов ДЗЗ (NDVI/EVI).

Объединение статичных (почвенно-агрохимических, орографических) и динамических (погода, индексы вегетации) предикторов сформировало основу устойчивой фичеризации.

SHAP-анализ подтвердил биофизическую обоснованность признаков: для озимой пшеницы и кукурузы систематически выделялись суммы осадков по фазам развития, амплитуда NDVI и количество жарких дней (T_max > 30 °C) в период май–июль. В целом, в интерпретации моделей регулярно фигурировали 5–7 ключевых предикторов, среди которых осадки, температурные экстремумы и запасы влаги весной входили в число наиболее значимых. Для повышения объяснимости и практического применения результатов работы нейросетевых моделей проводилась интеграция методов объяснимого ИИ (XAI), что значительно повышает доверие к результатам, позволяя связывать прогнозы с конкретными агрометеорологическими условиями, создавая отчеты вида: «снижение урожайности связано с дефицитом осадков в период дата1–дата2, повышенным числом жарких дней и низкими запасами влаги весной».

Использование единой модели для всего региона оказалось менее эффективно, чем кластеризация по агроклиматическим подзонам (по показателям ГТК, высоте, распределению осадков). Такой подход снижал RMSE на 7–12% и повышал устойчивость прогноза в экстремальные годы (засуха, жара).

Комбинированные модели, совмещающие DSSAT/Sirius и машинное обучение, обеспечивают физически обоснованную динамику роста культур с одновременной коррекцией систематических ошибок. Такой подход позволил сохранять интерпретируемость модели и использовать сценарии погодных изменений для вероятностной оценки урожайности.

Гибридизация снимает часть ограничений «черного ящика»: биофизическая динамика задает рамки экстраполяции при климатических дрейфах, а машинное обучение позволяет скорректировать статистические смещения, опираясь на фактические наблюдения. Особенно продуктивным этот подход оказывается в условиях дефицита длинных репрезентативных рядов и при наличии выраженных трендов.

Основные сложности создавали переобучение на малых и неоднородных выборках, утечка целевой переменной при извлечении признаков, климатический дрейф и несовместимость пространственных масштабов (поле — район — спутниковый пиксель). Для их минимизации в какой-то мере эффективными оказались: сквозная и пространственно-стратифицированная валидация, присваивание версий разнородным данным, мониторинг дрейфа, перекалибровка, регуляризация и методы доменной адаптации, согласование пространственного разрешения через разбиение на тайловые агрегаты.

Обсуждение

Для практического получения адаптируемых моделей и их внедрения в региональные структуры агропромышленного комплекса целесообразно использовать «минимальный жизнеспособный контур», включающий:

– модуль сбора и очистки метеорологических и спутниковых данных;
– библиотеку фазовых и экстремальных признаков (осадки, жара, засушливость, NDVI/EVI);
– ансамблевый бустинг как основной предиктор и LSTM/ConvLSTM для сезонных траекторий;
– XAI-слой (SHAP) с отчётами;
– двойной вывод (точечный прогноз и вероятностные квантили);
– регулярную перекалибровку и стресс-тестирование на экстремальных годах.

Для Северного Кавказа показала эффективность параллельная модельная настройка по подзонам (степи, предгорья, увлажнённые долины) с последующей «мягкой» ансамблизацией.

Заключение

Современные методы машинного обучения, в частности градиентный бустинг и гибридные подходы с глубоким обучением, демонстрируют высокую эффективность в прогнозировании урожайности сельскохозяйственных культур. Интеграция наземных и спутниковых данных позволяет существенно повысить точность и заблаговременность прогнозов, при этом сохраняя их интерпретируемость.

Региональная специфика Северного Кавказа требует особого подхода к построению прогнозных моделей. Наилучшие результаты достигаются при реализации следующих методологических принципов:

1. Комплексный анализ метеоданных с учетом фазовой агрегации показателей и экстремальных значений.
2. Пространственная дифференциация территорий по агроклиматическим зонам.
3. Мониторинг вегетации через анализ сезонной динамики индексов NDVI/EVI с учетом орографических факторов.
4. Объяснимый искусственный интеллект на базе современных методов интерпретации моделей (SHAP, LIME).
5. Интегративный подход к прогнозированию с учетом вероятностных сезонных прогнозов погоды.

Выполненный обзор и тесты подтверждают, что интеграция машинного обучения с дистанционным зондированием и агрометеорологическими предикторами формирует на текущем этапе наиболее успешную комбинацию для оперативных и предсезонных оценок урожайности. По сравнению с традиционными регрессионными подходами ансамблевые методы (XGBoost/LightGBM/CatBoost) и гибриды с глубокими нейронными сетями (LSTM/ConvLSTM) устойчиво снижают RMSE на 15–35% и поднимают R^2 до 0,75–0,90, особенно при корректной фазовой агрегации погоды и использовании фенологических метрик NDVI/EVI. Для Северного Кавказа наибольший вклад вносят суммы осадков в критические фазы, число жарких дней, весенние запасы влаги и орография; это согласуется с агрономической теорией и подтверждается XAI-разбором вкладов признаков (SHAP), повышающим доверие пользователей.

Таким образом, сочетание бустингов и гибридных нейронных сетей с физически обоснованной фичеризацией и XAI-интерпретацией позволяет уже сегодня получить устойчивые, полезные для принятия решений прогнозы урожайности. Выигрыш максимален при фазовой агрегации погодных факторов, учёте экстремумов, использовании NDVI/EVI и орографических предикторов, кластерной стратификации и вероятностном выводе. Эти принципы обеспечивают баланс точности, объяснимости и переносимости — ключевых требований к прикладным агропроизводственным решениям в условиях климатической изменчивости региона.

Перспективные направления для дальнейших исследований, согласно обзору, связаны с развитием пространственно-временных моделей (например, ансамбли на основе ConvLSTM), внедрением федеративного обучения для работы с данными конкретных хозяйств без их централизации, а также унификацией прочих региональных данных с применением робастных методов валидации. Реализация этих подходов будет способствовать созданию устойчивых и эффективных систем поддержки принятия решений в агропромышленном комплексе.

Конфликт интересов

Не указан.

Рецензия

Сообщество рецензентов Международного научно-исследовательского журнала
DOI: https://doi.org/10.60797/IRJ.2025.162.119.1

Conflict of Interest

None declared.

Review

International Research Journal Reviewers Community
DOI: https://doi.org/10.60797/IRJ.2025.162.119.1

Список литературы / References

1. Алибий Ф.М. Влияние температурного режима и осадков на динамику урожайности основных сельскохозяйственных культур на Северном Кавказе в 2010–2016 гг. / Ф.М. Алибий, А.В. Гятов, Б.А. Кушхова // Аграрный вестник Урала. — 2018. — № 8. — С. 10–16. — URL: https://cyberleninka.ru/article/n/vliyanie-temperaturnogo-rezhima-i-osadkov-na-dinamiku-urozhaynosti-osnovnyh-selskohozyaystvennyh-kultur-na-severnom-kavkaze-v-2010-2016 (дата обращения: 08.09.2025).
2. Денисов П.В. Прогнозирование урожайности озимой пшеницы с использованием дистанционного зондирования / П.В. Денисов, А.Б. Иванов, Н.П. Мишуров [и др.] // Управление рисками в АПК. — 2021. — № 39. — С. 37–45. — DOI: 10.53988/24136573-2021-01-03.
3. Ерошенко Ф.В. Дистанционная оценка урожайности озимой пшеницы (вегетационный индекс фотосинтетического потенциала) / Ф.В. Ерошенко, С.А. Барталев, И.Г. Сторчак [и др.] // Современные проблемы дистанционного зондирования Земли из космоса. — 2016. — Т. 13, № 4. — С. 99–112.
4. Страшная А.И. Прогнозирование урожайности зерновых при комплексировании наземных и спутниковых данных / А.И. Страшная, О.В. Береза, П.С. Кланг // Гидрометеорологические исследования и прогнозы. — 2021. — № 2 (380). — С. 111–137. — DOI: 10.37162/2618-9631-2021-2-111-137.
5. Aydoğdu M. Evaluation of yield prediction performance of DSSAT CSM-CERES-Wheat model in some bread wheat varieties / M. Aydoğdu, H. Yildiz, H. Gürkan [et al.] // International Journal of Environment and Geoinformatics. — 2023. — Vol. 10, No. 1. — P. 51–66. — DOI: 10.30897/ijegeo.1087591.
6. Bali N. Emerging trends in machine learning to predict crop yield and study its influential factors: A survey / N. Bali, A. Singla // Archives of Computational Methods in Engineering. — 2021. — Vol. 29, No. 1. — P. 95–112. — DOI: 10.1007/s11831-021-09569-8.
7. Elsayed M.L. Modeling current and future climate effects on winter wheat production in Colorado, USA / M.L. Elsayed, S. Anapalli, L.R. Ahuja [et al.] // Bridging Among Disciplines by Synthesizing Soil and Plant Processes. — 2019. — Vol. 8. — P. 171–199.
8. Gavahi K. A combined convolutional neural network with long short-term memory network for crop yield forecasting / K. Gavahi, P. Abbaszadeh, H. Moradkhani // Expert Systems with Applications. — 2021. — Vol. 184. — P. Article 115511. — DOI: 10.1016/j.eswa.2021.115511.
9. Huber F. Extreme Gradient Boosting for yield estimation compared with Deep Learning approaches / F. Huber, A. Yushchenko, B. Stratmann [et al.] // arXiv preprint arXiv:2208.12633. — 2022. — URL: https://arxiv.org/abs/2208.12633 (accessed: 09.15.2025).
10. Klompenburg T. Crop yield prediction using machine learning: A systematic literature review / T. Klompenburg, A. Kassahun, C. Catal // Computers and Electronics in Agriculture. — 2020. — Vol. 177. — Art. 105709. — DOI: 10.1016/j.compag.2020.105709.
11. Muruganantham P. A systematic literature review on crop yield prediction using remote sensing and deep learning / P. Muruganantham, K. Srikumar, V. Ramasamy [et al.] // Remote Sensing. — 2022. — Vol. 14, No. 9. — Art. 1990. — P. 28. — DOI: 10.3390/rs14091990.
12. Oikonomidis A. Deep learning for crop yield prediction: A systematic literature review / A. Oikonomidis, A. Gkelios, N. Doulamis [et al.] // New Zealand Journal of Crop and Horticultural Science. — 2023. — Vol. 51, No. 2. — P. 135–164. — DOI: 10.1080/01140671.2022.2032213.
13. Ryo M. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis / M. Ryo // Artificial Intelligence in Agriculture. — 2022. — Vol. 6. — P. 257–265. — DOI: 10.1016/j.aiia.2022.11.003.
14. Shawon S.M. Crop yield prediction using machine learning: an extensive review / S.M. Shawon, M.A. Islam, M.M. Hasan [et al.] // Smart Agricultural Technology. — 2025. — No. 10. — P. 17. — DOI: 10.1016/j.atech.2024.100718.
15. Sun J. County-level soybean yield prediction using deep CNN-LSTM model / J. Sun, L. Di, Z. Sun [et al.] // Sensors. — 2019. — Vol. 19, No. 20. — Art. 4363. — DOI: 10.3390/s19204363.
16. Sun J. Global de-trending significantly improves the accuracy of XGBoost-based county-level maize and soybean yield prediction in the Midwestern United States / J. Sun, Y. Li, Y. Xie [et al.] // GIScience & Remote Sensing. — 2024. — Vol. 61, No. 1. — DOI: 10.1080/15481603.2024.2349341.
17. Yang X. Multi-source information fusion-driven corn yield prediction using the Random Forest from the perspective of Agricultural and Forestry Economic Management / X. Yang, Z. Hua, L. Li [et al.] // Scientific Reports. — 2024. — Vol. 14. — Art. 54354. — DOI: 10.1038/s41598-024-54354-9.
18. Yan Y. Crop Yield Time-Series Data Prediction Based on Multiple Hybrid Machine Learning Models / Y. Yan, Y. Wang, J. Li [et al.] // arXiv preprint arXiv:2502.10405. — 2025. — URL: https://arxiv.org/abs/2502.10405 (accessed: 09.15.2025).

Список литературы на английском языке / References in English

1. Alibiy F.M. Vliyaniye temperaturnogo rezhima i osadkov na dinamiku urozhaynosti osnovnykh sel'skokhozyaystvennykh kul'tur na Severnom Kavkaze v 2010–2016 gg. [The Influence of Temperature and Precipitation on the Dynamics of Yields of Major Agricultural Crops in the North Caucasus in 2010–2016] / F.M. Alibiy, A.V. Gyatov, B.A. Kushkhova // Agrarnyy vestnik Urala [Agrarian Bulletin of the Urals]. — 2018. — No. 8. — P. 10–16. — URL: https://cyberleninka.ru/article/n/vliyanie-temperaturnogo-rezhima-i-osadkov-na-dinamiku-urozhaynosti-osnovnyh-selskohozyaystvennyh-kultur-na-severnom-kavkaze-v-2010-2016 (accessed: 09.08.2025). [in Russian]

2. Denisov P.V. Prognozirovaniye urozhaynosti ozimoy pshenitsy s ispol'zovaniyem distantsionnogo zondirovaniya [Forecasting Winter Wheat Yield Using Remote Sensing] / P.V. Denisov, A.B. Ivanov, N.P. Mishurov [et al.] // Upravleniye riskami v APK [Risk Management in the Agro-Industrial Complex]. — 2021. — No. 39. — P. 37–45. — DOI: 10.53988/24136573-2021-01-03. [in Russian]

3. Yeroshenko F.V. Distantsionnaya otsenka urozhaynosti ozimoy pshenitsy (vegetatsionnyy indeks fotosinteticheskogo potentsiala) [Remote Assessment of Winter Wheat Yield (Vegetation Index of Photosynthetic Potential)] / F.V. Yeroshenko, S.A. Bartalev, I.G. Storchak [et al.] // Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa [Modern Problems of Remote Sensing of the Earth from Space]. — 2016. — Vol. 13, No. 4. — P. 99–112. [in Russian]

4. Strashnaya A.I. Prognozirovaniye urozhaynosti zernovykh pri kompleksirovanii nazemnykh i sputnikovykh dannykh [Forecasting Grain Yields by Combining Ground-Based and Satellite Data] / A.I. Strashnaya, O.V. Bereza, P.S. Klang // Gidrometeorologicheskiye issledovaniya i prognozy [Hydrometeorological Research and Forecasts]. — 2021. — No. 2 (380). — P. 111–137. — DOI: 10.37162/2618-9631-2021-2-111-137. [in Russian]

5. Aydoğdu M. Evaluation of yield prediction performance of DSSAT CSM-CERES-Wheat model in some bread wheat varieties / M. Aydoğdu, H. Yildiz, H. Gürkan [et al.] // International Journal of Environment and Geoinformatics. — 2023. — Vol. 10, No. 1. — P. 51–66. — DOI: 10.30897/ijegeo.1087591.

6. Bali N. Emerging trends in machine learning to predict crop yield and study its influential factors: A survey / N. Bali, A. Singla // Archives of Computational Methods in Engineering. — 2021. — Vol. 29, No. 1. — P. 95–112. — DOI: 10.1007/s11831-021-09569-8.

7. Elsayed M.L. Modeling current and future climate effects on winter wheat production in Colorado, USA / M.L. Elsayed, S. Anapalli, L.R. Ahuja [et al.] // Bridging Among Disciplines by Synthesizing Soil and Plant Processes. — 2019. — Vol. 8. — P. 171–199.

8. Gavahi K. A combined convolutional neural network with long short-term memory network for crop yield forecasting / K. Gavahi, P. Abbaszadeh, H. Moradkhani // Expert Systems with Applications. — 2021. — Vol. 184. — P. Article 115511. — DOI: 10.1016/j.eswa.2021.115511.

9. Huber F. Extreme Gradient Boosting for yield estimation compared with Deep Learning approaches / F. Huber, A. Yushchenko, B. Stratmann [et al.] // arXiv preprint arXiv:2208.12633. — 2022. — URL: https://arxiv.org/abs/2208.12633 (accessed: 09.15.2025).

10. Klompenburg T. Crop yield prediction using machine learning: A systematic literature review / T. Klompenburg, A. Kassahun, C. Catal // Computers and Electronics in Agriculture. — 2020. — Vol. 177. — Art. 105709. — DOI: 10.1016/j.compag.2020.105709.

11. Muruganantham P. A systematic literature review on crop yield prediction using remote sensing and deep learning / P. Muruganantham, K. Srikumar, V. Ramasamy [et al.] // Remote Sensing. — 2022. — Vol. 14, No. 9. — Art. 1990. — P. 28. — DOI: 10.3390/rs14091990.

12. Oikonomidis A. Deep learning for crop yield prediction: A systematic literature review / A. Oikonomidis, A. Gkelios, N. Doulamis [et al.] // New Zealand Journal of Crop and Horticultural Science. — 2023. — Vol. 51, No. 2. — P. 135–164. — DOI: 10.1080/01140671.2022.2032213.

13. Ryo M. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis / M. Ryo // Artificial Intelligence in Agriculture. — 2022. — Vol. 6. — P. 257–265. — DOI: 10.1016/j.aiia.2022.11.003.

14. Shawon S.M. Crop yield prediction using machine learning: an extensive review / S.M. Shawon, M.A. Islam, M.M. Hasan [et al.] // Smart Agricultural Technology. — 2025. — No. 10. — P. 17. — DOI: 10.1016/j.atech.2024.100718.

15. Sun J. County-level soybean yield prediction using deep CNN-LSTM model / J. Sun, L. Di, Z. Sun [et al.] // Sensors. — 2019. — Vol. 19, No. 20. — Art. 4363. — DOI: 10.3390/s19204363.

16. Sun J. Global de-trending significantly improves the accuracy of XGBoost-based county-level maize and soybean yield prediction in the Midwestern United States / J. Sun, Y. Li, Y. Xie [et al.] // GIScience & Remote Sensing. — 2024. — Vol. 61, No. 1. — DOI: 10.1080/15481603.2024.2349341.

17. Yang X. Multi-source information fusion-driven corn yield prediction using the Random Forest from the perspective of Agricultural and Forestry Economic Management / X. Yang, Z. Hua, L. Li [et al.] // Scientific Reports. — 2024. — Vol. 14. — Art. 54354. — DOI: 10.1038/s41598-024-54354-9.

18. Yan Y. Crop Yield Time-Series Data Prediction Based on Multiple Hybrid Machine Learning Models / Y. Yan, Y. Wang, J. Li [et al.] // arXiv preprint arXiv:2502.10405. — 2025. — URL: https://arxiv.org/abs/2502.10405 (accessed: 09.15.2025).