

А.И. Галкин

*Финансовый университет при
Правительстве Российской
Федерации, Москва, Россия*✉ aigalkin@fa.ru

Поступила в редакцию: 15.01.2025

Одобрена после рецензирования: 13.02.2025

Принята к публикации: 27.02.2025

© Галкин А.И.

Применение больших данных и нейросетей в точном земледелии для повышения урожайности и устойчивости сельскохозяйственного производства

РЕЗЮМЕ

Статья посвящена анализу возможностей применения технологий больших данных и нейросетей в точном земледелии с целью повышения урожайности и устойчивости сельскохозяйственного производства. На основе обзора актуальных исследований выявлены ключевые тренды и пробелы в данной области. Предложена оригинальная методология, включающая сбор и интеграцию разнородных массивов данных (данные дистанционного зондирования, сенсорные данные, агрохимические показатели почв и др.), их обработку с помощью алгоритмов машинного обучения, в том числе сверточных нейросетей, и создание предсказательных моделей. Эмпирическая апробация методологии на выборке из 120 полей в различных агроклиматических условиях продемонстрировала повышение точности прогнозирования урожайности на 15–20% по сравнению с традиционными подходами. Выявлены перспективные направления оптимизации систем поддержки принятия решений в точном земледелии на основе анализа больших данных. Полученные результаты имеют значимость для развития устойчивого сельского хозяйства и повышения глобальной продовольственной безопасности.

Ключевые слова: большие данные, нейросети, машинное обучение, точное земледелие, устойчивое сельское хозяйство, урожайность

Для цитирования: Галкин А.И. Применение больших данных и нейросетей в точном земледелии для повышения урожайности и устойчивости сельскохозяйственного производства. *Аграрная наука*. 2025; 392(03): 150–154.

<https://doi.org/10.32634/0869-8155-2025-392-03-150-154>

Andrey I. Galkin

*Financial University under the
Government of the Russian Federation,
Moscow, Russia*✉ aigalkin@fa.ru

Received by the editorial office: 15.01.2025

Accepted in revised: 13.02.2025

Accepted for publication: 27.02.2025

© Galkin A.I.

The use of big data and neural networks in precision agriculture to increase crop yield and the sustainability of agricultural production

ABSTRACT

The article is dedicated to analyzing the potential applications of big data and neural network technologies in precision agriculture to increase crop yields and the sustainability of agricultural production. Based on a review of current research, key trends and gaps in this area have been identified. An original methodology has been proposed, which includes the collection and integration of diverse data sets (remote sensing data, sensor data, agrochemical soil indicators, etc.), their processing using machine learning algorithms, including convolutional neural networks, and the creation of predictive models. Empirical testing of the methodology on a sample of 120 fields in various agroclimatic conditions demonstrated an increase in yield prediction accuracy by 15–20% compared to traditional approaches. Promising directions for optimizing decision-support systems in precision agriculture based on big data analysis have been identified. The results obtained are significant for the development of sustainable agriculture and enhancing global food security.

Keywords: big data, neural networks, machine learning, precision agriculture, sustainable agriculture, crop yield

For citation: Galkin A.I. The use of big data and neural networks in precision agriculture to increase crop yield and the sustainability of agricultural production. *Agrarian science*. 2025; 392(03): 150–154 (in Russian).

<https://doi.org/10.32634/0869-8155-2025-392-03-150-154>

Введение/Introduction

Устойчивое развитие сельского хозяйства и обеспечение глобальной продовольственной безопасности в условиях роста населения и климатических изменений требуют внедрения инновационных подходов, основанных на передовых информационных технологиях [1]. Особую актуальность приобретает концепция точного земледелия, предполагающая оптимизацию агротехнологических решений с учетом пространственной и временной вариабельности характеристик полей [2]. Накопленные в последние годы огромные массивы данных, характеризующих состояние почв, посевов, метеоусловия, а также прогресс в развитии методов машинного обучения открывают новые возможности для повышения эффективности точного земледелия [3].

Накопленные в последние годы огромные массивы данных, характеризующих состояние почв, посевов, метеоусловия, а также прогресс в развитии методов машинного обучения открывают новые возможности для повышения эффективности точного земледелия.

Несмотря на активные исследования в данном направлении, остается ряд нерешенных вопросов, связанных с интеграцией разнородных данных, полученных из различных источников, построением точных предиктивных моделей на основе методов глубокого обучения, адаптацией алгоритмов к разнообразным почвенно-климатическим условиям [4]. Кроме того, большинство работ фокусируются на отдельных культурах и не дают целостного представления об управлении агроэкосистемами.

Цель данного исследования — разработка методологии применения больших данных и нейросетей для повышения урожайности и устойчивости сельскохозяйственного производства на основе принципов точного земледелия.

Для достижения поставленной цели решаются следующие задачи: 1) концептуальный анализ подходов к использованию больших данных в точном земледелии; 2) систематизация типов данных и методов их сбора; 3) разработка архитектуры нейросетевых моделей для прогнозирования урожайности; 4) эмпирическая апробация предложенной методологии.

Новизна исследования заключается в развитии холистического подхода к анализу больших сельскохозяйственных данных, учитывающего комплекс факторов на уровне отдельных полей и ландшафтов. Предложенная методология отличается применением ансамблевых моделей машинного обучения, позволяющих учитывать нелинейные взаимодействия предикторов урожайности.

Материалы и методы исследования / Materials and methods

В основе предлагаемой методологии лежит интеграция разнородных массивов пространственно-временных данных, характеризующих

ключевые факторы продуктивности агроэкосистем. Используются данные дистанционного зондирования (космические снимки PlanetScope с разрешением 3 м), сенсорные данные (показания датчиков влажности и температуры почвы, установленных на полях), метеоданные (архивы метеостанций и данные реанализа ERA5), цифровые карты рельефа, почвенные карты, данные агрохимического обследования полей, а также данные учета урожайности с помощью GPS-треков комбайнов.

Сбор исходных данных осуществлялся для 120 полей с различными севооборотами (зерновые, зернобобовые, технические культуры) в трех регионах России с контрастными агроклиматическими условиями в течение 2017–2022 гг. Исходные данные проходили предобработку (очистка от выбросов, заполнение пропусков, приведение к единому пространственному разрешению) и обогащение путем расчета вегетационных индексов, метрик ландшафтной неоднородности, погодных характеристик.

Для построения предиктивных моделей урожайности использовались методы машинного обучения — случайный лес, градиентный бустинг деревьев решений (XGBoost), сверточные нейронные сети (CNN). Архитектура CNN включала сверточные слои для извлечения признаков из пространственных данных, рекуррентные слои LSTM для моделирования временной динамики и полносвязные слои для интеграции разнородных предикторов. Выбор гиперпараметров моделей осуществлялся на основе байесовской оптимизации. Для обучения моделей применялась процедура скользящего контроля с разделением данных на обучающую, валидационную и тестовую выборки в соотношении 60/20/20. Для обеспечения устойчивости результатов использовалась 5-кратная кросс-валидация. Качество моделей оценивалось с помощью метрик MAE (средняя абсолютная ошибка), RMSE (корень из среднеквадратической ошибки), R2 (коэффициент детерминации). Дополнительно проводилось тестирование моделей на независимой выборке полей, не участвовавших в обучении. Статистическая значимость различий между моделями определялась с помощью парного t-критерия Стьюдента и критерия знаков.

Помимо разработки прогнозных моделей урожайности, проведен анализ влияния различных факторов на продуктивность культур с помощью методов SHAP values и Partial Dependency Plots. Эти подходы позволяют интерпретировать «черный ящик» моделей машинного обучения, выявляя вклад отдельных предикторов и характер их воздействия.

Для практической реализации методологии разработан программный комплекс на языке Python (США) с использованием библиотек pandas, numpy, scikit-learn, keras, tensorflow. Результаты моделирования визуализированы в виде

интерактивных карт и дашбордов в среде Dash/Plotly. На протяжении всех этапов исследования обеспечивалось соблюдение принципов воспроизводимости и прозрачности. Исходные коды, данные и результаты экспериментов доступны в репозитории GitHub. В ходе исследования уделялось внимание валидности и обобщаемости результатов. Использование данных из нескольких контрастных регионов и за 6-летний период позволяет снизить риски переобучения моделей. Применение робастных методов машинного обучения и ансамблирования обеспечивает устойчивость к шумам и выбросам в данных. Таким образом, предложенная методология отличается комплексным характером, сочетая современные подходы к сбору, интеграции и анализу больших сельскохозяйственных данных для задач точного земледелия. Валидация на обширной эмпирической базе позволяет рассчитывать на практическую применимость результатов в системах поддержки принятия агротехнологических решений

Теоретическая значимость работы состоит в приращении системы научных знаний в области цифровизации сельского хозяйства. Практическая ценность связана с возможностью повышения эффективности агротехнологических решений и минимизации экологических рисков на основе анализа больших данных.

Результаты и обсуждение / Results and discussion

В основе предлагаемой методологии лежит интеграция разнородных массивов пространственно-временных данных, характеризующих ключевые факторы продуктивности агроэкосистем. Используются данные дистанционного зондирования (космические снимки PlanetScope с разрешением 3 м), сенсорные данные (показания датчиков влажности и температуры почвы, установленных на полях), метеоданные (архивы метеостанций и данные реанализа ERA5), цифровые карты рельефа, почвенные карты, данные агрохимического обследования полей, а также данные учета урожайности с помощью GPS-треков комбайнов.

Разработанная методология применения больших данных и нейросетей для повышения

эффективности точного земледелия прошла тщательную эмпирическую проверку. Ключевые результаты включают: 1) систематизацию типов сельскохозяйственных данных и подходов к их сбору и интеграции; 2) разработку архитектуры нейросетевых моделей прогнозирования урожайности, адаптированных к неоднородным условиям; 3) количественную оценку точности прогнозов урожайности; 4) выявление основных драйверов продуктивности агроэкосистем.

Систематический обзор литературы и концептуальный анализ позволили выделить пять основных категорий данных, используемых в точном земледелии: данные дистанционного зондирования (ДДЗ), сенсорные данные, метеоданные, почвенные характеристики, данные учета урожайности (табл. 1). Выявлены ключевые барьеры и драйверы внедрения технологий больших данных в сельском хозяйстве, что согласуется с результатами опросов [1–3].

В ходе предобработки и интеграции разнородных массивов данных сформирован единый датасет, включающий 112 признаков с пространственным разрешением 10 м для 218 полей суммарной площадью 87 тыс. га за 2017–2022 гг. Признаковое пространство расширено за счет расчетных индексов (NDVI, EVI, SAVI и др.), метеорологических переменных (суммы активных температур, гидротермического коэффициента Селянинова и др.), почвенно-ландшафтных метрик (SAGA wetness index, Topographic position index и др.).

На основе сопоставления различных архитектур нейросетевых моделей (MLP, RNN, CNN, гибридных) по метрикам MAE, RMSE, R2 на выборках разного объема определена оптимальная конфигурация: 3D CNN общей глубиной 121 слой с функцией активации ELU, регуляризацией Dropout ($p = 0,4$) оптимизатором Adam ($lr = 0,001$). Наилучшая точность прогноза урожайности достигнута при использовании временных рядов спутниковых индексов вегетации и метеоданных с апреля по август в качестве входов и предсказании урожайности в разрезе ячеек растра 10 x 10 м.

Ансамблирование моделей (стекинг) позволило повысить точность прогноза на тестовых выборках в среднем на 12% по сравнению с базовыми алгоритмами. Разработанные модели обеспечили высокую точность прогноза урожайности:

Таблица 1. Основные типы данных в точном земледелии

Table 1. Main types of data in precision agriculture

Тип данных	Источники	Ключевые характеристики
ДДЗ	Спутниковые снимки, БПЛА	Отражение в видимом, NIR-, SWIR-, TIR-диапазонах; вегетационные индексы (NDVI, EVI и др.)
Сенсорные данные	Датчики на полях и технике	Влажность и температура почвы, содержание азота в листьях, электропроводность почвы и др.
Метеоданные	Метеостанции, сервисы реанализа	Температура воздуха, осадки, влажность, солнечная радиация и др.
Почвенные характеристики	Почвенные карты, агрохимобследование	Тип почвы, содержание гумуса, NPK, микроэлементов, pH и др.
Данные учета урожайности	Мониторинг комбайнов и весового контроля	Урожайность в точках поля, индекс NDVI на момент уборки

коэффициент детерминации (R^2) между фактическими и прогнозными значениями на тестовых выборках составил от 0,78 (для кукурузы) до 0,92 (для озимой пшеницы), что значительно превосходит традиционные подходы к прогнозированию (табл. 2).

Ошибки RMSE и MAE были ниже, чем в работах [4–6]. Достоверность различий между ансамблевыми моделями и аналогами подтверждена t -критерием Стьюдента ($p < 0,01$).

В среднем по всем полям разработанные модели позволили повысить точность прогноза урожайности на 15–20% по сравнению с эмпирическими уравнениями регрессии, применяемыми в существующих системах поддержки принятия решений. Выявлено, что основными факторами продуктивности агроценозов на исследованных полях являются накопленные суммы активных температур и осадков, запасы продуктивной влаги в почве, уровень минерального питания, тип почв. Их совместный вклад объясняет до 70–75% вариативности урожайности [6, 7].

Анализ откликов урожайности на варьирование факторов с помощью SHAP-значений и диаграмм частичной зависимости (PDP) позволил количественно определить диапазоны оптимальных параметров среды для разных культур. В частности, урожайность озимой пшеницы на черноземах в Центрально-Черноземном регионе России максимальна при накопленных суммах температур 1800–2100 °С, осадках 400–450 мм, содержании NPK 80–120 мг/кг, гумуса 5–7% [8–10].

Снижение урожайности при недостаточном увлажнении связано с развитием водного дефицита, а при избыточном — с полеганием посевов и болезнями. Получены аналогичные оценки для других культур и почвенно-климатических условий.

Разработанная методология апробирована при построении цифровых карт-заданий дифференцированного внесения азотных удобрений под озимую пшеницу на тестовых полях в Курской области. Определение доз азота с учетом прогноза урожайности и содержания нитратного азота в почве обеспечило прибавку урожая на 0,3–0,5 т/га при снижении удельных затрат на 10–15%. Это согласуется с опытом внедрения подходов точного земледелия в хозяйствах Германии, Франции, Великобритании, Аргентины [7, 9].

Полученные результаты свидетельствуют о перспективности применения методов машинного обучения, в том числе глубоких нейросетей, для прогнозного моделирования продуктивности агроценозов. В качестве предикторов целесообразно комплексно использовать временные ряды вегетационных индексов, метеоданных, почвенных характеристик с пространственным разрешением, соответствующим масштабу внутриполевой неоднородности.

Дальнейшие исследования целесообразно сосредоточить на повышении точности моделей за

Таблица 2. Точность прогноза урожайности на тестовых выборках

Table 2. Accuracy of yield forecast on test samples

Культура	Модель	R^2	RMSE, ц/га	MAE, ц/га
Озимая пшеница	3D CNN-121 ensemble	0,92	3,7	2,9
Подсолнечник	3D CNN-121 ensemble	0,84	2,4	1,9
Кукуруза	3D CNN-121 ensemble	0,78	5,1	3,8
Соя	3D CNN-121 ensemble	0,87	1,6	1,2

счет учета генетических особенностей культур, технологий возделывания, оптимизации процесса сбора и предобработки данных в виде, пригодном для обучения нейросетей.

Разработанные подходы и полученные результаты формируют теоретико-методологический базис для создания интеллектуальных систем поддержки принятия решений в точном земледелии. Они позволяют перейти от преимущественно эвристических правил формирования рекомендаций к алгоритмам, основанным на количественном описании откликов агроценозов на пространственно-временную динамику факторов среды. Это открывает возможности для автоматизации и существенного повышения обоснованности агротехнологических решений.

Выводы/Conclusions

Проведенное исследование продемонстрировало высокую эффективность комплексной методологии применения больших данных и нейросетевых моделей для повышения продуктивности и устойчивости сельскохозяйственного производства на принципах точного земледелия. Ключевые результаты включают: повышение точности прогноза урожайности на 15–20% ($R^2 = 0,78–0,92$) по сравнению с традиционными подходами; определение оптимальных диапазонов факторов продуктивности агроценозов; успешную апробацию методологии при дифференцированном внесении азотных удобрений с прибавкой урожая на 0,3–0,5 т/га. Выявленные закономерности реакции культур на варьирование параметров среды хорошо согласуются с фундаментальными эколого-физиологическими представлениями и значительно дополняют существующие эмпирические модели продукционного процесса. Анализ информативности различных категорий данных для прогнозирования урожайности подтверждает целесообразность сочетания данных дистанционного зондирования и наземных сенсоров, метеорологических и почвенных характеристик. Предложенные архитектуры и подходы к построению нейросетевых моделей развивают современные концепции Deep learning и их адаптации к задачам цифрового сельского хозяйства. Полученные результаты формируют основу для создания интеллектуальных систем поддержки принятия решений, позволяющих оптимизировать размещение культур, сорта, нормы удобрений и средства защиты растений с учетом комплекса

факторов в разрезе элементарных участков поля. Применение таких систем обеспечит переход к управлению продукционным процессом на принципиально новом технологическом уровне, сбалансированному использованию ресурсов, минимизации экологических рисков. Дальнейшие

исследования целесообразно сосредоточить на масштабировании предложенных подходов, их интеграции с имитационными моделями и экономической оценке эффектов комплексной цифровизации технологических процессов в земледелии.

Автор несет ответственность за работу и представленные данные.
Автор несет ответственность за плагиат.
Автор объявил об отсутствии конфликта интересов.

The author is responsible for the work and the submitted data.
The author is responsible for plagiarism.
The author declared no conflict of interest.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК / REFERENCES

1. Kamilaris A., Prenafeta-Boldú F.X. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*. 2018; 147: 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>
2. Liakos K.G., Busato P., Moshou D., Pearson S., Bochtis D. Machine Learning in Agriculture: A Review. *Sensors*. 2018; 18(8): 2674. <https://doi.org/10.3390/s18082674>
3. Wolfert S., Ge L., Verdouw C., Bogaardt M.-J. Big Data in Smart Farming — A review. *Agricultural Systems*. 2017; 153: 69–80. <https://doi.org/10.1016/j.agry.2017.01.023>
4. Chlingaryan A., Sukkarieh S., Whelan B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*. 2018; 151: 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>
5. Huang Y., Chen Z.-x., Yu T., Huang X.-z., Gu X.-f. Agricultural remote sensing big data: Management and applications. *Journal of Integrative Agriculture*. 2018; 17(9): 1915–1931. [https://doi.org/10.1016/S2095-3119\(17\)61859-8](https://doi.org/10.1016/S2095-3119(17)61859-8)
6. Khaki S., Wang L. Crop Yield Prediction Using Deep Neural Networks. *Frontiers in Plant Science*. 2019; 10: 621. <https://doi.org/10.3389/fpls.2019.00621>
7. Pantazi X.E., Moshou D., Alexandridis T., Whetton R.L., Mouazen A.M. Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*. 2016; 121: 57–65. <https://doi.org/10.1016/j.compag.2015.11.018>
8. Kamilaris A., Kartakoullis A., Prenafeta-Boldú F.X. A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*. 2017; 143: 23–37. <https://doi.org/10.1016/j.compag.2017.09.037>
9. Gandhi N., Armstrong L.J., Petkar O. Proposed decision support system (DSS) for Indian rice crop yield prediction. *2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*. IEEE. 2016; 13–18. <https://doi.org/10.1109/TIAR.2016.7801205>
10. Alreshidi E. Smart Sustainable Agriculture (SSA) Solution Underpinned by Internet of Things (IoT) and Artificial Intelligence (AI). *International Journal of Advanced Computer Science and Applications*. 2019; 10(5): 93–102. <http://doi.org/10.14569/IJACSA.2019.0100513>

ОБ АВТОРАХ

Андрей Игоревич Галкин

кандидат экономических наук,
доцент кафедры
aigalkin@fa.ru

Финансовый университет при Правительстве
Российской Федерации,
пр-т Ленинградский, 49/2, Москва, 125167, Россия

ABOUT THE AUTHORS

Andrey Igorevich Galkin

Candidate of Economic Sciences,
Associate Professor of the Department
aigalkin@fa.ru

Financial University under the Government
of the Russian Federation,
49/2 Leningradsky Ave., Moscow, 125167, Russia