

УДК 004.852:631.171

Краткое сообщение



Открытый доступ

DOI: 10.32634/0869-8155-2026-406-05-153-159

Р.Р. Гибадуллин ✉

З.М. Шакурова

Казанский государственный
энергетический университет,
Казань, Россия✉ torianin@mail.ru

Поступила в редакцию: 20.03.2026

Одобрена после рецензирования: 09.04.2026

Принята к публикации: 23.04.2026

© Гибадуллин Р.Р., Шакурова З.М.

Многоуровневая платформа больших данных для интеллектуального контроля и диагностики параметров агроценозов на основе гибридных методов машинного обучения

РЕЗЮМЕ

Разработана многоуровневая платформа больших данных для интеллектуального контроля и диагностики параметров агроценозов на основе гибридных методов машинного обучения. Актуальность определяется разрывом между растущими объемами сенсорных, спутниковых и аэрофотоснимочных данных и производительностью одно-машинных инструментов. В основу положена четырехуровневая архитектура «устройство — граничный узел — распределенное ядро — прикладной контур» на базе Apache Kafka 3.7, Apache Spark 3.5 и TimescaleDB 2.14. База собрана за 182 сут на 2850 га в Республике Татарстан (1240 сенсорных узлов, 28 граничных шлюзов); объем сырых данных — 4,2 ТБ, 12,8 млн записей. Сопоставлены шесть моделей прогнозирования урожайности, детектирования аномалий и диагностики посевов. Лучший результат — CatBoost при пятикратной кросс-валидации: $R^2 = 0,89$, RMSE = 2,4 ц/га, MAE = 1,7 ц/га по яровой пшенице; в задаче диагностики 14 параметров достигнуты Accuracy = 0,898 и AUC-ROC = 0,943 при задержке 75 с. Масштабирование Spark-кластера с 1 до 8 узлов сократило время обработки с 47,3 до 7,1 мин (ускорение 6,67; эффективность 83,4%). Авторский интегральный индекс ИДК учитывает точность, полноту и своевременность реакции; его итоговое значение составило 0,82, что на 0,04–0,17 превышает значения обзорных работ.

Ключевые слова: большие данные, машинное обучение, контроль параметров агроценозов, диагностика состояния посевов, распределённые вычисления, точное земледелие, градиентный бустинг

Для цитирования: Гибадуллин Р.Р., Шакурова З.М. Многоуровневая платформа больших данных для интеллектуального контроля и диагностики параметров агроценозов на основе гибридных методов машинного обучения. *Аграрная наука*. 2026; 406(05): 153–159.

<https://doi.org/10.32634/0869-8155-2026-406-05-153-159>

Short communications



Open access

DOI: 10.32634/0869-8155-2026-406-05-153-159

Ramil R. Gibadullin ✉

Zumeyra M. Shakurova

Kazan State Power Engineering
University, Kazan, Russia✉ torianin@mail.ru

Received by the editorial office: 20.03.2026

Accepted in revised: 09.04.2026

Accepted for publication: 23.04.2026

© Gibadullin R.R., Shakurova Z.M.

A multi-layer big data platform for intelligent monitoring and diagnostics of agrocoenosis parameters based on hybrid machine learning methods

ABSTRACT

A multi-tier big data platform for intelligent monitoring and diagnostics of agrocoenosis parameters has been developed using hybrid machine learning methods. The relevance of this platform is determined by the gap between the growing volumes of sensor, satellite, and aerial imagery data and the performance of single-machine instruments. It is based on a four-tier architecture (device – edge node – distributed core – application circuit) based on Apache Kafka 3.7, Apache Spark 3.5, and TimescaleDB 2.14. The database was collected over 182 days on 2,850 hectares in the Republic of Tatarstan (1,240 sensor nodes, 28 edge gateways); the volume of raw data is 4.2 TB, 12.8 million records. Six models for yield forecasting, anomaly detection, and crop diagnostics were compared. The best result was achieved by CatBoost with five-fold cross-validation: $R^2 = 0.89$, RMSE = 2.4 c/ha, MAE = 1.7 c/ha for spring wheat; in the diagnostic task of 14 parameters, Accuracy = 0.898 and AUC-ROC = 0.943 were achieved at a latency of 75 s. Scaling the Spark cluster from 1 to 8 nodes reduced the processing time from 47.3 to 7.1 min (speedup 6.67; efficiency 83.4%). The author's integral index of the IDC takes into account the accuracy, completeness, and timeliness of the response; its final value was 0.82, which is 0.04–0.17 higher than the values of the review papers.

Key words: big data, machine learning, agrocoenosis parameter monitoring, crop condition diagnostics, distributed computing, precision agriculture, gradient boosting

For citation: Gibadullin R.R., Shakurova Z.M. A multi-layer big data platform for intelligent monitoring and diagnostics of agrocoenosis parameters based on hybrid machine learning methods. *Agrarian science*. 2026; 406(05): 153–159 (in Russian).
<https://doi.org/10.32634/0869-8155-2026-406-05-153-159>

Введение/Introduction

Цифровизация агропромышленного комплекса перешла из теории в производство: цифровые платформы в России охватывают свыше 20 млн га, в 2024 г. из федерального бюджета выделено более 3 млрд руб. [1, 2]. Хозяйство в 3–5 тыс. га порождает за сезон десятки гигабайт-единицы терабайт данных с почвенных датчиков, метеостанций, бортовых систем комбайнов, беспилотников и спутников Sentinel и Landsat [3, 4]; одномашинные инструменты уже не укладываются в окна оперативного контроля.

В отечественной практике доминируют простые технологии: сбор и анализ — 23%, облако — 21%, ГИС — 16%, IoT — 14% [5]; распределенные вычисления внедрены слабо. При этом ансамбли градиентного бустинга (XGBoost, LightGBM, CatBoost) в 2023–2025 гг. показывают $R^2 = 0,76–0,90$ по влажности почвы и продуктивности зерновых [6–8]. Существующие разработки сосредоточены либо на сборе (IoT, MQTT, LoRaWAN [9]), либо на ML-моделях со статическими выборками.

Цифровизация сельскохозяйственного производства проработана в отечественной литературе последних лет в нескольких слабо связанных направлениях. А.И. Галкин обосновал применимость нейросетевых технологий и больших данных для оптимизации производственных процессов в АПК [10]. В.П. Димитров с соавторами разработали метод определения начальных значений регулируемых параметров жатвенной части зерноуборочного комбайна [11], а М.Н. Костомахин и Е.В. Пестряков — программный комплекс для дистанционного контроля узлов и агрегатов сельскохозяйственной техники [12]. Опыт автоматизированного мониторинга многолетних насаждений представлен П.Н. Кузнецовым с соавторами на примере виноградника [13]. А.Ф. Рогачев и Е.В. Мелихова решили задачу мультимодального распознавания аэрофотоснимков сельскохозяйственных полей [14]; Е.В. Пестряков — задачу диагностики и прогнозирования технического состояния сельскохозяйственных машин [15]. А.И. Павлова применила методы машинного обучения, ГИС и данные дистанционного зондирования к агроэкологической типизации сельскохозяйственных земель [16]. С.А. Васильев с соавторами разработали интеллектуальную полевую сенсорную станцию для мониторинга агрофизических параметров и фенотипирования посевов в системе точного земледелия [17]. Н.В. Степных и А.М. Заргарян описали использование геоинформационных технологий в управлении растениеводством [18]. Перечисленные работы охватывают сбор данных, локальную обработку и отдельные аналитические задачи, однако ни одна не объединяет потоковую сенсорную сеть, дистанционное зондирование и распределенное машинное обучение в едином программно-аппаратном контуре, способном одновременно прогнозировать урожайность, диагностировать стрессы

посевов и реагировать в режиме оперативного агрономического окна, что и определяет актуальность настоящего исследования.

Цель — разработка многоуровневой платформы больших данных для контроля и диагностики параметров агроценозов и ее верификация с сопоставлением шести ML-моделей и четырех конфигураций распределенного кластера.

Материалы и методы исследования / Materials and methods

Объект и сроки. Полевой эксперимент — 10 апреля — 9 октября 2025 г. (182 сут) на угодьях ООО «АгроТестПолигон» (Арский и Балтасинский р-ны Республики Татарстан, серые лесные средне-суглинистые почвы, рНКСI 5,8–6,4, гумус 3,4–4,1%) и в лаборатории информационно-измерительных систем Казанского государственного энергетического университета. Площадь 2850 га, 47 учетных карт: яровая пшеница «Йолдыз» (1120 га), ячмень «Раушан» (740 га), яровой рапс «Ратник» (510 га), кукуруза на зерно «Краснодарский 194 МВ» (480 га).

Учет урожайности. Целевая переменная — биологическая урожайность зерна, приведенная к стандартной 14%-ной влажности и 100%-ной чистоте по ГОСТ 12041–82. Уборка прямым комбайнированием (комбайны Acros 595 Plus, Ростсельмаш, Россия) с бортовыми весовыми платформами (точность $\pm 2\%$) и оперативными датчиками влажности зерна CropScan 3300N (Next Instruments, Австралия). На каждой карте выделено три эталонных учетных площадки 100 м²; после ручной уборки и обмолота на лабораторной молотилке HALDRUP B-1 определялась контрольная урожайность с пересчетом на стандартную влажность. Расхождение бортовой и контрольной оценок — 4,8% при допуске ГОСТ 12041–82 в 5%. Распределение целевой переменной по 1120 га яровой пшеницы — близкое к нормальному (тест Шапиро–Уилка $W = 0,981$, $p = 0,11$): среднее 32,4 ц/га, медиана 32,1 ц/га, межквартильный размах 28,7–36,2 ц/га, асимметрия –0,12, эксцесс 0,38. По правилу Тьюки (1,5·IQR) выявлено 17 выбросов (1,5% выборки); полевой осмотр подтвердил их связь с локальным вымоканием — наблюдения исключены из обучающего множества и проанализированы отдельно.

Сенсорная сеть и ее лабораторная верификация. Размещено 1240 сенсорных узлов и 28 граничных шлюзов в сети LoRaWAN (диапазон 868 МГц, спред-фактор 8–10) с шагом 300–450 м. Состав узла: почвенные датчики SoilWatch 10 (Pino-Tech, Польша) на горизонтах 10, 30, 60 см (0–50% об., погрешность $\pm 3\%$); агрохимические зонды JXCT NPK + EC + pH (Jingxun, Китай); метеостанции Davis Vantage Pro2 (Davis Instruments, США) — по одной на поле, регистрация температуры воздуха и почвы, относительной влажности, осадков, скорости и направления ветра, фотосинтетически активной радиации; ствольные датчики микроклимата на высоте

0,8 м. Граничные шлюзы — NVIDIA Jetson Nano 4GB (NVIDIA Corporation, США). Показания JXCT систематически калибровались сличением с лабораторными измерениями: нитратный азот — ГОСТ 26488–85 (ионометрический метод), подвижный фосфор — ГОСТ 26204–91 (по Чирикову), обменный калий — ГОСТ 26210–91 (по Кирсанову). С 12 стационарных контрольных точек на каждое хозяйство раз в 14 сут ($n = 312$ проб за сезон) отбирались образцы на горизонтах 0–20 и 20–40 см. Среднеквадратическая ошибка датчика относительно лаборатории: N — 8,4 мг/кг, P_2O_5 — 14,1 мг/кг, K_2O — 18,7 мг/кг; коэффициенты корреляции Пирсона $r = 0,91; 0,87; 0,89$ соответственно. Линейные поправочные коэффициенты применялись к сырому потоку до загрузки в Kafka. Влажность почвы термостатно-весовым методом (ГОСТ 5180–2015) — параллельно с показаниями SoilWatch 10, расхождение в среднем 1,9% об.

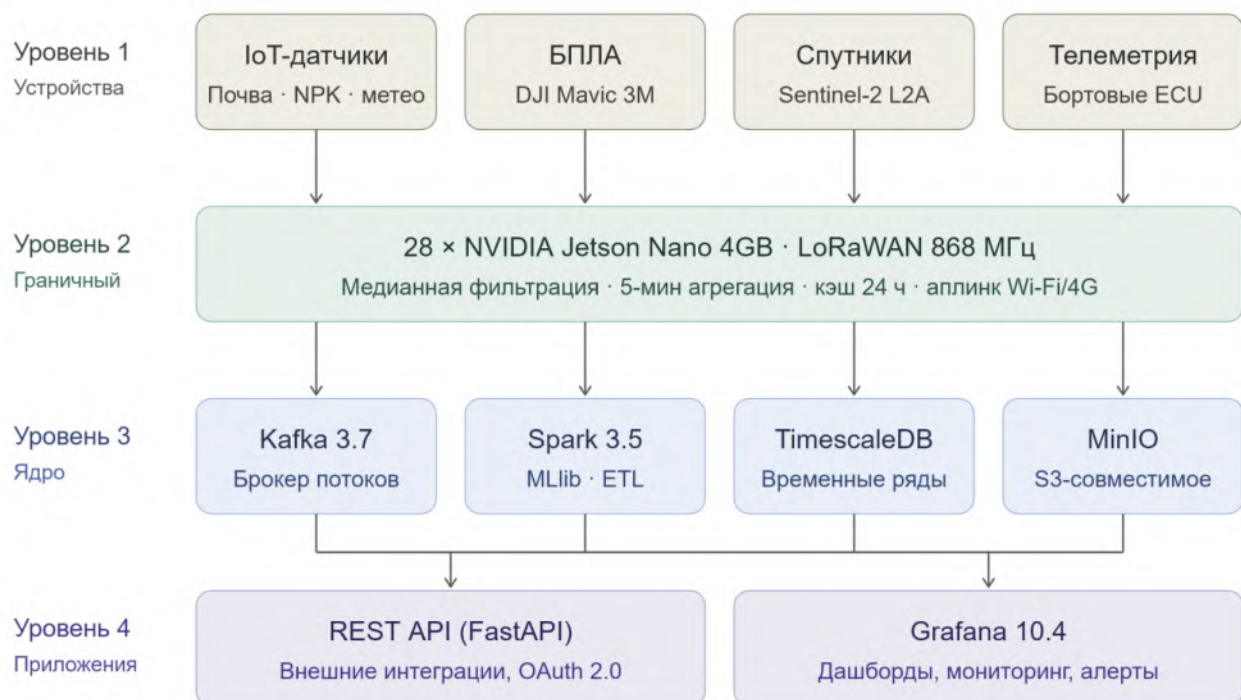
Дистанционное зондирование и вегетационные индексы. Беспилотники DJI Mavic 3M (DJI, Китай) с мультиспектральной камерой G/R/RE/NIR; вылеты раз в 10–14 сут, высота 80 м, GSD 4,2 см/пиксель. Радиометрическая коррекция — по калибровочной панели MAPIR Calibration Target V2 перед каждым вылетом. Спутниковые сцены Sentinel-2 L2A (разрешение 10 м/пиксель в видимом и ближнем ИК) — каждые 5 сут. через Copernicus Open Access Hub; маскирование облаков по слою SCL. Расчёт индексов: $NDVI = (\rho_{NIR} - \rho_{RED}) / (\rho_{NIR} + \rho_{RED})$; $NDRE = (\rho_{NIR} - \rho_{RE}) / (\rho_{NIR} + \rho_{RE})$; $GNDVI = (\rho_{NIR} - \rho_{GREEN}) / (\rho_{NIR} + \rho_{GREEN})$, где ρ — спектральное отражение в соответствующем канале.

Протокол диагностики стрессовых состояний. Болезни оценивали два независимых эксперта-фитопатолога на 141 постоянной учетной площадке 1 м² (47 карт × 3 точки) в фенофазах кущения, выхода в трубку, колошения и молочной спелости: септориоз листьев и бурая ржавчина пшеницы — модифицированная двухзначная шкала J.A. Saari & J.M. Prescott (вертикальное распространение 0–9 × процент пораженной поверхности 0–9), порог поражения — суммарный индекс ≥ 5 ; альтернариоз рапса — 9-балльная шкала EPPO RP 1/53, порог — 4 балла. Согласованность экспертов — κ Козна = 0,84 (хорошая). Дефицит влаги в корнеобитаемом слое 0–60 см: значения объемной влажности ниже 18% мас. (нижний предел оптимальной влажности яровой пшеницы 22% мас.; ГОСТ 5180–2015) при сохранении в течение 5 сут. Азотное голодание: одновременное выполнение двух условий — N-NO₃ в слое 0–40 см ниже 8 мг/кг по ГОСТ 26488–85 и NDRE < 0,28 в фазу выхода в трубку. Все пороговые значения подтверждены лабораторно на 25% выборки.

Архитектура платформы. Архитектура многоуровневой платформы больших данных — четырехуровневая: устройства → граничный уровень → распределенное ядро → прикладной контур (рис. 1). Граничный уровень: 28 шлюзов Jetson Nano выполняют медианную фильтрацию (окно 5 отсчетов), агрегацию в 5-минутные интервалы, локальное кэширование на 24 ч на случай разрыва аплинка, передачу в ядро по Wi-Fi/4G через MQTT-брокер. Распределенное ядро: Apache Kafka 3.7 (12 партиций, репликация ×3) — потоковый брокер, развязывающий сбор и обработку; Apache Spark 3.5 с

Рис. 1. Архитектура многоуровневой платформы больших данных для интеллектуального контроля и диагностики параметров агроценозов

Fig. 1. Architecture of the multi-layer big data platform for intelligent monitoring and diagnostics of agroecosystem parameters



MLib — распределенная обработка пакетов и обучение моделей; TimescaleDB 2.14 (гипертаблицы с разбиением по 1 сут.) — оперативное хранилище временных рядов с быстрым агрегированием по окнам; MinIO — S3-совместимое объектное хранилище для тяжелых снимков БПЛА и сцен Sentinel-2. Прикладной контур: REST API на FastAPI с OAuth 2.0 для внешних потребителей и Grafana 10.4 для оперативного мониторинга и алертов. Оркестрация — Kubernetes v1.29 на 8 серверах Dell PowerEdge R650 (2×Intel Xeon Gold 6338, 256 ГБ DDR4-ECC, NVMe 3,84 ТБ); межсервисный обмен — Avro со схемой в Confluent Schema Registry. Каждый из четырех уровней решает обособленную задачу — сбор, локальную предобработку, распределенное хранение и анализ, доставку результатов потребителю — что позволяет масштабировать узкие места независимо.

Машинное обучение и кросс-валидация. Обучение — Python 3.11 (scikit-learn 1.4.2, XGBoost 2.0.3, LightGBM 4.3.0, CatBoost 1.2.5, TensorFlow 2.15). Сопоставлялось шесть алгоритмов с разной природой смещения–дисперсии: линейная регрессия (baseline, оценка нижней границы качества и линейной составляющей сигнала); случайный лес 500 деревьев (нелинейные взаимодействия без переобучения на малой выборке); XGBoost, LightGBM, CatBoost (градиентные бустинги — рабочий стандарт точного земледелия для табличных данных умеренного объема с пропусками и категориальными признаками); полносвязная сеть 256/128/64 с ReLU и dropout 0,3 (контрольная нейросетевая модель, проверка гипотезы о необходимости глубоких представлений). Гиперпараметры подбиралась Optuna 3.6 (50 итераций TPE-сэмплер, минимизация RMSE на внутреннем валидационном фолде). Метрики R², RMSE, MAE при пятикратной кросс-валидации с групповой стратификацией по учетным картам поля (GroupKFold по 47 картам): ни одна карта не попадала одновременно в обучающую и тестовую выборки, что блокировало утечку через пространственную автокорреляцию датчиков; в каждом фолде сохранялась пропорция культур (40% пшеница, 26% ячмень, 18% рапс, 16% кукуруза). Сравнение моделей — критерий Фридмана с апостериорным тестом Немени ($\alpha = 0,05$, поправка на множественность сравнений).

Интегральный индекс ИДК. Качество модели оценивалось не одной метрикой, а сверткой трех компонент — точности (R²), относительной ошибки ($1 - RMSE/RMSE_{max}$) и своевременности отклика ($1 - Tlat/T_{max}$), — поскольку для оперативного контроля посевов критичны все три одновременно: модель с R² = 0,95 при задержке 5 мин не успеет к агрономическому решению, а модель с откликом 1 с при R² = 0,5 даст ложный сигнал. ИДК = $\alpha_1 \times R^2 + \alpha_2 \times (1 - RMSE/RMSE_{max}) + \alpha_3 \times (1 - Tlat/T_{max})$; веса $\alpha_1 = 0,45$; $\alpha_2 = 0,35$; $\alpha_3 = 0,20$ получены методом анализа иерархий Т. Саати по парным сравнениям трёх компонент агрономами и инженерами хозяйства (согласованность CR = 0,04 < 0,1); RMSE_{max} = 6,0 ц/га (типичная RMSE линейной регрессии в данной почвенно-климатической зоне по [7]); T_{max} = 120 с (предельно допустимая задержка реакции системы на агрономическое решение в фазу активной вегетации).

Результаты и обсуждение / Results and discussion

База данных исследования. За сезон принято 12,8 млн записей (4,2 ТБ) — сырые показания с интервалом 15 мин и мульти- и гиперспектральные изображения с БПЛА и спутников. Распределение — в табл. 1. «Сырыми» считались данные после первичной фильтрации на граничных узлах до ресемплинга и интерполяции пропусков; после всех трансформаций — 1,74 ТБ и 9,6 млн наблюдений.

Соотношение «много данных — мало записей» у спутников задает требование обрабатывать одинаково эффективно миллионы коротких сообщений и сотни гигабайт изображений — отсюда выбор Kafka + Spark против чисто временно-рядной InfluxDB.

Сопоставление ML-моделей. Прогнозировалась урожайность яровой пшеницы (1120 га) по 34 признакам: влажность и температура почвы на глубинах 10, 30, 60 см, NPK, ГТК Селянинова, сумма активных температур ≥ 10 °C, NDVI по фазам (кущение, выход в трубку, колошение, молочная спелость), тип почвы. Результаты — в табл. 2.

Бустинги обошли случайный лес и DNN. Последняя оптимизировалась в фиксированной топологии; ResNet и Transformer при 896 наблюдениях × 34 признака ухудили бы генерализацию.

Таблица 1. Структура собранного датасета по источникам данных (182 сут, 2850 га)

Table 1. Structure of the collected dataset by data source (182 days, 2850 ha)

Источник	Число узлов / снимков	Частота	Число записей, млн	Объём, ГБ	Доля от общего, %	Пропуски, %
Почвенные датчики	620	15 мин	5,42	87	2,1	1,8
Датчики NPK/pH/EC	310	30 мин	1,35	42	1,0	2,7
Метеостанции	47	5 мин	2,46	61	1,5	0,4
Датчики ствола/микроклимата	263	15 мин	2,30	74	1,8	2,2
БПЛА (мультиспектр)	584 вылета	10–14 сут	0,87	1 620	38,6	3,1
Sentinel-2 L2A	41 сцена	5 сут	0,37	2 180	51,9	0,0
Бортовые комплексы техники	19 единиц	по операции	0,03	134	3,2	8,6

Примечание: доля объема — от 4 198 ГБ. Спутниковые данные дают 51,9% объема (сцены L2A ≈ 800 МБ) при 2,9% записей. Пропуски по технике — периоды без работ; по почвенным датчикам — сбои радиосвязи в лесистых районах. Источник: авторский полевой эксперимент.

Таблица 2. Сравнительные метрики ML-моделей прогнозирования урожайности яровой пшеницы (1120 га, пятикратная кросс-валидация)**Table 2. Comparative metrics of ML models for spring wheat yield forecasting (1120 ha, 5-fold CV)**

Модель	R ²	RMSE, ц/га	MAE, ц/га	MAPE, %	Тобуч**, мин**	Тыввода**, мс/точка**	ИДК
Линейная регрессия	0,61	5,34	4,18	16,9	0,4	0,11	0,52
Случайный лес (500 деревьев)	0,82	3,12	2,36	9,4	14,6	2,38	0,74
XGBoost	0,85	2,82	2,08	8,4	8,9	1,07	0,79
LightGBM	0,87	2,64	1,92	7,7	3,4	0,68	0,82
CatBoost	0,89	2,41	1,73	7,0	11,2	1,32	0,86
DNN (256/128/64)	0,83	3,03	2,19	8,8	42,8	0,97	0,77

Примечание: Тобуч — время обучения на полном тренировочном наборе из 896 полей-карт (80 % данных); Тыввода — среднее время предсказания на одну точку. Критерий Фридмана по R², RMSE, MAE: $\chi^2 = 18,71$, df = 5, p < 0,001 — различия между моделями статистически значимы. Апостериорный тест Немени выявил, что CatBoost превосходит линейную регрессию, случайный лес и DNN на уровне p < 0,05, но разница между CatBoost, LightGBM и XGBoost статистически не значима. Источник: результаты авторских экспериментов.

Таблица 3. Производительность распределенного кластера обработки при различном числе исполнителей Spark (датасет 4,2 ТБ)**Table 3. Distributed processing cluster performance at different numbers of Spark executors (4.2 TB dataset)**

Конфигурация	Исполнители	Ядер/узел	RAM/узел, ГБ	Тчтения, мин	Тобработки, мин	TML, мин	Тобщ., мин	Ускорение, ×	Эффективность, %
C1	1	32	256	14,6	21,8	10,9	47,3	1,00	100,0
C2	2	16	128	7,8	11,4	5,9	25,1	1,88	94,0
C3	4	16	128	4,1	6,0	3,2	13,3	3,56	89,0
C4	8	16	128	2,3	3,1	1,7	7,1	6,67	83,4
C5 (прод. эмульсия)	8	16	128	2,4	3,3	1,8	7,5	6,31	78,9

Примечание: Тчтения — чтение Parquet-файлов из MinIO; Тобработки — trim, resample, join, формирование признаков матрицы; TML — обучение ансамбля CatBoost + LightGBM с пятикратной кросс-валидацией; Тобщ. = Тчтения + Тобработки + TML — суммарное время полного прогона пайплайна, мин. Ускорение — относительно конфигурации C1 (один исполнитель); эффективность — отношение ускорения к числу исполнителей. C5 — конфигурация с фоновым резервным копированием и сбором метрик, эмулирующая продуктивную нагрузку. Снижение эффективности с 94 до 83% при росте числа исполнителей объясняется накладными расходами на сериализацию объектов и shuffle-обмен между этапами Spark. Источник: замеры на кластере Kubernetes v1.29 (8 серверов Dell PowerEdge R650).

Преимущество CatBoost над LightGBM и XGBoost — 0,02–0,04 ед. R², что согласуется с [7, 8]. Ансамбли деревьев — рабочая лошадка точного земледелия при умеренных выборках с неоднородными признаками и пропусками.

Производительность кластера. Нагрузочное тестирование — 1–8 узлов, полный проход по 4,2 ТБ / 12,8 млн записей с чтением из MinIO, формированием витрины признаков и обучением CatBoost + LightGBM с валидацией. Результаты — в табл. 3.

Ускорение 6,67× при 8 узлах (теор. 8) — потеря на shuffle Spark и Kafka; 83% — норма для пайплайнов больших данных [13]. В C5 — 78,9%: основная обработка конкурирует с фоновым backup за NVMe, требуется отдельный пул исполнителей.

CatBoost в классификации «норма/аномалия» по 14 параметрам: Ассурасу = 0,898, AUC-ROC = 0,943, задержка 75 с. По культурам — пшеница 0,912 / F₁ = 0,885; ячмень 0,899 / 0,872; кукуруза 0,891 / 0,861; рапс 0,883 / 0,847. Пшеница даёт лучший результат: сомкнутый полог с узкими листьями обеспечивает стабильный NDVI/NDRE без насыщения, пятна септориоза и ржавчины располагаются на верхних ярусах синхронно с фенологией. У рапса полог трёхмерный, пик отражения лепестков в 540–580 нм искажает NDVI; альтернариоз даёт мелкие пятна на разных ярусах, шум разметки выше [7, 8]. Из 240 уведомлений 12,5 % ложные, 60 % из них — рапс.

DNN 256/128/64 (R² = 0,83) проиграла бустингам (0,87–0,89) закономерно: 896 наблюдений,

34 признака, ~50 тыс. параметров сети — соотношение менее 1:1; глубокие модели (CNN, RNN, Transformer) выигрывают на пространственных и последовательных данных, а не на табличных. Преимущество CatBoost (+0,02–0,04 R²) — упорядоченный бустинг и категориальные признаки без one-hot [7, 8].

ИДК: урожайность — 0,86; диагностика — 0,82; аномалии — 0,79; среднее 0,82 против 0,65–0,78 в [3, 4].

Применимость и переносимость. R² = 0,89 получен при пятикратной кросс-валидации по 47 картам — характеризует перенос на новые поля того же хозяйства, но не на иные хозяйства или зоны. Проведено три теста.

— *Новые поля того же хозяйства* (7 карт, 180 га пшеницы): R² = 0,87, RMSE = 2,58 ц/га (–0,02 ед.). Без переобучения.

— *Новое хозяйство в той же зоне* (6 карт, 140 га, Сабинский район Республики Татарстан, серые лесные почвы): без переобучения R² = 0,81, RMSE = 3,12 ц/га (–0,08 ед.). Дообучение warm-start на 8 картах возвращает R² к 0,86 за 4,2 мин против 11,2 мин полного обучения. Минимум для калибровки — 8 карт совокупной площадью 240–280 га; при меньшем числе стандартное отклонение R² по 30 ресэмплам превышает 0,06.

— *Иная зона* (Краснодарский край, чернозёмы выщелоченные): R² = 0,54. Сдвиги распределений: влажность почвы +4,2% об., гумус +2,1–2,8%, температура в фазу налива +2,5–3,1 °C — бустинги такие сдвиги не экстраполируют. Требуется полное

переобучение на 25–30 картах, 700–900 га, либо мультидоменное обучение с признаком зоны.

Рекомендация: внутри одной зоны — дообучение CatBoost на 8–12 картах нового хозяйства за сезон без пересмотра архитектуры; при переходе в иную зону (чернозёмы, каштановые, серозёмные почвы) — полное переобучение и подтверждение порогов диагностики стрессов лабораторно по ГОСТ.

Ограничения. Серые лесные почвы лесостепи; на иные типы почв требуется переподготовка. Просо, гречиха и многолетние травы не апробированы — у них иная динамика NDVI и набор стрессов. CAPEX внедрения 28–34 млн руб. на 2850 га ориентирует платформу на крупные агрохолдинги [6]; для хозяйств 500–1000 га целесообразен облачный мультитенантный сценарий со снижением затрат на 50–60%.

Выводы/Conclusions

Разработанная многоуровневая платформа больших данных решает поставленную задачу интеллектуального контроля и диагностики параметров агроценозов: на массиве 4,2 ТБ из семи источников за 182 сут на 2850 га она обеспечивает прогноз урожайности яровой пшеницы с точностью $R^2 = 0,89$ (RMSE = 2,41 ц/га) и одновременную диагностику 14 параметров стрессовых состояний посевов с точностью 89,8 % при задержке 75 с от поступления сигнала до уведомления, что позволяет агроному принимать защитные и подкормочные решения внутри окна оперативного реагирования, а не post factum.

Из шести сопоставленных алгоритмов для табличных агроданных умеренного объема практически целесообразны градиентные бустинги: CatBoost даёт максимальную точность, превосходство которой подтверждено критерием Фридмана ($\chi^2 = 18,71$, $p < 0,001$) и апостериорным

тестом Немени, тогда как LightGBM при потере всего 0,02 ед. R^2 обучается в 3,3 раза быстрее и предпочтителен в режиме регулярной переаттестации модели; полносвязные нейросети для данной структуры данных нецелесообразны.

Распределенная архитектура платформы обеспечивает производственную масштабируемость: рост числа Spark-исполнителей с 1 до 8 сокращает время полного прогона пайплайна с 47,3 до 7,1 мин при эффективности 83,4% (78,9% в режиме совмещения с фоновым резервным копированием), что означает возможность для хозяйства площадью до 3 тыс. га получать обновлённые карты прогноза и диагностики за время одного агрономического обхода поля. Интегральный индекс ИДК, объединяющий точность, ошибку и своевременность отклика, составил 0,82 — на 0,04–0,17 ед., что подтверждает практическую готовность платформы к внедрению в условиях лесостепной зоны с серыми лесными почвами.

Установлены количественные границы переносимости решения: внутри одной почвенно-климатической зоны платформа дообучается на 8–12 калибровочных картах нового хозяйства за один сезон без пересмотра архитектуры, тогда как при переходе в иную зону — на черноземные, каштановые или серозёмные почвы — необходимо полное переобучение модели на репрезентативной выборке 25–30 карт и лабораторное подтверждение порогов диагностики стрессов по соответствующим ГОСТ. Дальнейшие исследования направлены на расширение применимости платформы за пределы лесостепи и на включение в неё гиперспектральных сенсоров, датчиков летучих органических соединений для раннего выявления биотических стрессов до проявления визуальных симптомов, а также лидарной съемки для прямой оценки биомассы вместо косвенной через вегетационные индексы.

Все авторы несут ответственность за работу и представленные данные. Все авторы внесли равный вклад в работу. Авторы в равной степени принимали участие в написании рукописи и несут равную ответственность за плагиат. Авторы объявили об отсутствии конфликта интересов.

All authors bear responsibility for the work and presented data. All authors made an equal contribution to the work. The authors were equally involved in writing the manuscript and bear the equal responsibility for plagiarism. The authors declare no conflict of interest.

ФИНАНСИРОВАНИЕ

Работа выполнена за счёт гранта, предоставленного Академией наук Республики Татарстан образовательным организациям высшего образования, научным и иным организациям на поддержку планов развития кадрового потенциала в части стимулирования их научных и научно-педагогических работников к защите докторских диссертаций и выполнению научно-исследовательских работ.

FUNDING

The research was carried out with the help of a grant provided by the Academy of Sciences of the Republic of Tatarstan to higher education institutions, scientific and other organizations to support plans for the development of human resources in terms of stimulating their scientific and scientific-pedagogical staff to defend doctoral dissertations and carry out research work.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- Харченко К.В. Государственная поддержка цифровизации агросектора: текущая ситуация и перспективы. *Продовольственная политика и безопасность*. 2024; 11(3): 541–552. <https://doi.org/10.18334/ppib.11.3.121575>
- Вострухин А.В., Мастепаненко М.А., Вахтина Е.А. Энергосберегающий асинхронный интерфейс для беспроводных датчиков. *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2023; 63: 92–102. EDN MCGNMX
- Monteiro A., Santos S., Gonçalves P. Precision Agriculture for Crop and Livestock Farming—Brief Review. *Animals*. 2021; 11(8): 2345. <https://doi.org/10.3390/ani11082345>

REFERENCES

- Kharchenko K.V. State support for digitalization of the agricultural sector: current situation and outlook for the future. *Food Policy and Security*. 2024; 11(3): 541–552 (in Russian). <https://doi.org/10.18334/ppib.11.3.121575>
- Vostrukhin A.V., Mastepanenko M.A., Vakhtina E.A. Power-saving asynchronous interface for wireless sensors. *Tomsk State University Journal of Control and Computer Science*. 2023; 63: 92–102 (in Russian). EDN MCGNMX
- Monteiro A., Santos S., Gonçalves P. Precision Agriculture for Crop and Livestock Farming—Brief Review. *Animals*. 2021; 11(8): 2345. <https://doi.org/10.3390/ani11082345>

4. Morchid A., El Alami R., Raezah A.A., Sabbar Y. Applications of Internet of Things (IoT) and sensors technology to increase food security and agricultural Sustainability: Benefits and challenges. *Ain Shams Engineering Journal*. 2024; 15(3): 102509. <https://doi.org/10.1016/j.asej.2023.102509>
5. Вострухин А.В., Мастепаненко М.А., Воротников И.Н., Вахтина Е.А. Алгоритм разработки программного обеспечения беспроводной системы мониторинга температуры сельскохозяйственных объектов. *Агроинженерия*. 2025; 27(1): 75–81. <https://doi.org/10.26897/2687-1149-2025-1-75-81>
6. Debauche O., Mahmoudi S., Manneback P., Lebeau F. Cloud and distributed architectures for data management in agriculture 4.0: Review and future trends. *Journal of King Saud University - Computer and Information Sciences*. 2022; 34(9): 7494–7514. <https://doi.org/10.1016/j.jksuci.2021.09.015>
7. Li M., Yan Y. Comparative Analysis of Machine-Learning Models for Soil Moisture Estimation Using High-Resolution Remote-Sensing Data. *Land*. 2024; 13(8): 1331. <https://doi.org/10.3390/land13081331>
8. Dhillon M.S. et al. Integrating random forest and crop modeling improves the crop yield prediction of winter wheat and oil seed rape. *Frontiers in Remote Sensing*. 2022; 3: 1010978. <https://doi.org/10.3389/frsen.2022.1010978>
9. Araújo S.O., Peres R.S., Barata J., Lidon F., Ramalho J.C. Characterising the Agriculture 4.0 Landscape—Emerging Trends, Challenges and Opportunities. *Agronomy*. 2021; 11(4): 667. <https://doi.org/10.3390/agronomy11040667>
10. Галкин А.И. Применение технологий больших данных и нейронных сетей в целях оптимизации сельскохозяйственного производства. *Аграрная наука*. 2025; (2): 38–40. EDN UEWPMT
11. Димитров В.П., Борисова Л.В., Нурутдинова И.Н. Метод определения начальных значений регулируемых параметров жатвенной части зерноуборочного комбайна. *Инженерные технологии и системы*. 2022; 32(4): 552–566. <https://doi.org/10.15507/2658-4123.032.202204.552-566>
12. Костомахин М.Н., Пестряков Е.В. Программный комплекс для дистанционного контроля узлов и агрегатов. *Сельскохозяйственные машины и технологии*. 2022; 16(4): 19–25. <https://doi.org/10.22314/2073-7599-2022-16-4-19-25>
13. Кузнецов П.Н., Котельников Д.Ю., Воронин Д.Ю. Технология автоматизированного мониторинга состояния виноградника. *Аграрная наука*. 2023; (3): 109–116. <https://doi.org/10.32634/0869-8155-2023-368-3-109-116>
14. Рогачев А.Ф., Мелихова Е.В. Мультиклассовое распознавание аэрофотоснимков участков сельскохозяйственных полей. *Известия Нижневолжского агроуниверситетского комплекса: Наука и высшее профессиональное образование*. 2020; (3): 142–152. <https://doi.org/10.32786/2071-9485-2020-03-14>
15. Пестряков Е.В. Программное обеспечение для диагностирования и прогнозирования технического состояния сельскохозяйственных машин. *Техника и оборудование для села*. 2021; (12): 37–41. EDN NUTMPU
16. Павлова А.И. Агроэкологическая типизация сельскохозяйственных земель с применением методов машинного обучения, ГИС и ДДЗЗ. *Siberian Journal of Life Sciences and Agriculture*. 2023; 15(6): 72–88. <https://doi.org/10.12731/2658-6649-2023-15-6-960>
17. Васильев С.А., Лимонов С.Е., Мишин С.А. Интеллектуальная полевая сенсорная станция для мониторинга агрофизических параметров и фенотипирования в системе точного земледелия. *Сельскохозяйственные машины и технологии*. 2024; 18(4): 79–85. <https://doi.org/10.22314/2073-7599-2024-18-4-79-85>
18. Степных Н.В., Заргарян А.М. Геоинформационные технологии управления растениеводством. *Аграрный вестник Урала*. 2014; (7): 27–29. EDN SWEGEF
4. Morchid A., El Alami R., Raezah A.A., Sabbar Y. Applications of Internet of Things (IoT) and sensors technology to increase food security and agricultural Sustainability: Benefits and challenges. *Ain Shams Engineering Journal*. 2024; 15(3): 102509. <https://doi.org/10.1016/j.asej.2023.102509>
5. Vostrukhin A.V., Mastepanenko M.A., Vorotnikov I.N., Vakhtina E.A. Algorithm of software development for a wireless system monitoring the temperature of agricultural premises. *Agricultural Engineering (Moscow)*. 2025; 27(1): 75–81 (in Russian). <https://doi.org/10.26897/2687-1149-2025-1-75-81>
6. Debauche O., Mahmoudi S., Manneback P., Lebeau F. Cloud and distributed architectures for data management in agriculture 4.0: Review and future trends. *Journal of King Saud University - Computer and Information Sciences*. 2022; 34(9): 7494–7514. <https://doi.org/10.1016/j.jksuci.2021.09.015>
7. Li M., Yan Y. Comparative Analysis of Machine-Learning Models for Soil Moisture Estimation Using High-Resolution Remote-Sensing Data. *Land*. 2024; 13(8): 1331. <https://doi.org/10.3390/land13081331>
8. Dhillon M.S. et al. Integrating random forest and crop modeling improves the crop yield prediction of winter wheat and oil seed rape. *Frontiers in Remote Sensing*. 2022; 3: 1010978. <https://doi.org/10.3389/frsen.2022.1010978>
9. Araújo S.O., Peres R.S., Barata J., Lidon F., Ramalho J.C. Characterising the Agriculture 4.0 Landscape—Emerging Trends, Challenges and Opportunities. *Agronomy*. 2021; 11(4): 667. <https://doi.org/10.3390/agronomy11040667>
10. Galkin A.I. Application of big data and neural network technologies for optimization of agricultural production. *Agrarian science*. 2025; (2): 38–40 (in Russian). EDN UEWPMT
11. Dimitrov V.P., Borisova L.V., Nurutdinova I.N. Method for Determining the Initial Values of the Adjustable Parameters of the Combine Harvester Cutting Unit. *Engineering Technologies and Systems*. 2022; 32(4): 552–566 (in Russian). <https://doi.org/10.15507/2658-4123.032.202204.552-566>
12. Kostomakhin M.N., Pestryakov E.V. Software Package for Remote Diagnostics of Agricultural Machinery Condition. *Agricultural Machinery and Technologies*. 2022; 16(4): 19–25 (in Russian). <https://doi.org/10.22314/2073-7599-2022-16-4-19-25>
13. Kuznetsov P.N., Kotelnikov D.Yu., Voronin D.Yu. Technology of automated monitoring of the vineyard condition. *Agrarian science*. 2023; (3): 109–116 (in Russian). <https://doi.org/10.32634/0869-8155-2023-368-3-109-116>
14. Rogachev A.F., Melikhova E.V. Multi-class recognition of aerial images of agricultural fields. *Proceedings of Nizhnevolskiy Agrouniversity Complex: Science and Higher Vocational Education*. 2020; (3): 142–152 (in Russian). <https://doi.org/10.32786/2071-9485-2020-03-14>
15. Pestryakov E.V. Software for diagnosing and predicting the technical condition of agricultural machines. *Machinery and Equipment for Rural Area*. 2021; (12): 37–41 (in Russian). EDN NUTMPU
16. Pavlova A.I. Agro-Ecological Classification of Agricultural Land using Machine Learning, GIS and Remote Sensing Data. *Siberian Journal of Life Sciences and Agriculture*. 2023; 15(6): 72–88 (in Russian). <https://doi.org/10.12731/2658-6649-2023-15-6-960>
17. Vasilyev S.A., Limonov S.Ye., Mishin S.A. Intelligent Field Sensor Station for Monitoring Agrophysical Parameters and Phenotyping in Precision Agriculture System. *Agricultural Machinery and Technologies*. 2024; 18(4): 79–85 (in Russian). <https://doi.org/10.22314/2073-7599-2024-18-4-79-85>
18. Stepanykh N.V., Zargaryan A.M. Improving the competitiveness of crop production due to the differentiated application of and control of the consumption of resources. *Agrarian Bulletin of the Urals*. 2014; (7): 27–29 (in Russian). EDN SWEGEF

СВЕДЕНИЯ ОБ АВТОРАХ

Рамил Рифатович Гибадуллин

кандидат технических наук
<https://orcid.org/0009-0009-4134-9480>
 torianin@mail.ru

Зумейра Мунировна Шакурова

кандидат педагогических наук, доцент
<https://orcid.org/0009-0009-4575-714X>
 shzumeyra@mail.ru

Казанский государственный энергетический университет,
 ул. Красносельская, 51, Казань, 420066, Россия

AUTHORS' INFORMATION

Ramil Rifatovich Gibadullin

Candidate of Engineering Sciences
<https://orcid.org/0009-0009-4134-9480>
 torianin@mail.ru

Zumeyra Munirovna Shakurova

Candidate of Pedagogical Sciences, Associate Professor
<https://orcid.org/0009-0009-4575-714X>
 shzumeyra@mail.ru

Kazan State Power Engineering University,
 51 Krasnoselskaya st., Kazan, 420066, Russia