

ИНТЕРПРЕТИРУЕМАЯ МОДЕЛЬ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ УРОЖАЙНОСТИ ЯРОВОЙ ПШЕНИЦЫ

© 2025 г. В. К. Каличкин¹, доктор сельскохозяйственных наук, В. И. Усенко², доктор сельскохозяйственных наук, А. А. Гаркуша², кандидат сельскохозяйственных наук, Д. С. Федоров¹, К. Ю. Максимович¹, кандидат биологических наук

¹Сибирский федеральный научный центр агробиотехнологий Российской академии наук, 630501, Новосибирская область, Новосибирский р-он, пос. Краснообск
E-mail: vk.kalichkin@gmail.com

²Федеральный Алтайский научный центр агробиотехнологий, 656910, Алтайский край, Барнаул, пос. Научный городок, 35
E-mail: usenko.001@mail.ru

Исследования проводили с целью создания интерпретируемой модели машинного обучения (объяснимый искусственный интеллект) для прогнозирования урожайности яровой пшеницы. Использовали данные длительного полевого опыта (2001–2024 гг.), проведенного в лесостепи Алтайского Приобья. Схема эксперимента предусматривала изучение роли предшественников, приемов основной обработки выщелоченного чернозема, а также уровней применения минеральных удобрений и химических средств защиты растений в формировании урожайности яровой мягкой пшеницы. Для создания модели использовали экстремальный градиентный бустинг (XGBoost), а для ее интерпретации – аддитивные объяснения Шепли (SHAP), что позволило оценить вклад каждого признака. Построенная модель XGBoost продемонстрировала высокую точность прогнозирования ($R^2=0,95$, $MAE=0,13$ т/га, $RMSE=0,17$ т/га), а интеграция с SHAP-анализом выявила наиболее значимые признаки (5...6 из 18), определяющие урожайность в лесостепи Алтайского Приобья. Наибольший вклад в высокую прогнозную урожайность вносили достаточное увлажнение за сельскохозяйственный год (596,5 мм; 1,19 т/га), использование пара в качестве предшественника (0,58 т/га) и применение азотно-фосфорных удобрений (0,21 т/га). Низкая прогнозная урожайность обусловлена недостатком увлажнения в течение сельскохозяйственного года (317 мм; –0,77 т/га) и с мая по октябрь (246 мм; –0,24 т/га), а также высокими значениями суммы положительных температур (2527,5 °C; –0,13 т/га), низким количеством осадков в период вегетации пшеницы (175 мм; –0,10 т/га) и отсутствием применения средств защиты растений (–0,10 т/га). Разработанная модель расширяет возможности машинного обучения, позволяя получать более надежные и информативные результаты.

INTERPRETABLE MACHINE LEARNING MODEL FOR SPRING WHEAT YIELD FORECASTING

V. K. Kalichkin¹, V. I. Usenko², A. A. Garkusha², D. S. Fedorov¹, K. Yu. Maksimovich¹

¹Siberian Federal Research Center of Agricultural Biotechnology, Russian Academy of Sciences, 630501, Novosibirskaya obl., Novosibirskii r-n, pos. Krasnoobsk
E-mail: vk.kalichkin@gmail.com

²Federal Altai Scientific Center of Agrobiotechnologies, 656910, Barnaul, pos. Nauchnyi gorodok, 35
E-mail: usenko.001@mail.ru

The research was conducted to construct an interpretable machine learning model (explainable artificial intelligence) for spring wheat yield forecasting. The data of a long-term field experiment (2001–2024) carried out in the forest-steppe of the Altai Priobye region were used. The scheme of the experiment provided for the analysis of predecessors' role, methods of basic cultivation of leached chernozem, as well as levels of application of mineral fertilizers and chemical plant protection products in the formation of spring soft wheat yield. Extreme gradient boosting (XGBoost) was used to construct the model, and SHapley Additive exPlanations (SHAP) were used for its interpretation, which allowed estimating the contribution of each feature. The constructed XGBoost model demonstrated high prediction accuracy ($R^2=0.95$, $MAE=0.13$ t/ha, $RMSE=0.17$ t/ha), and integration with SHAP-analysis revealed the most significant features (5...6 out of 18) determining yield in the forest-steppe of the Altai Priobye. The greatest contribution to high predicted yield in the experiment was made by sufficient moisture for the crop year (596.5 mm; 1.19 t/ha), the use of fallow as a predecessor (0.58 t/ha) and the use of nitrogen-phosphorus fertilizers (0.21 t/ha). Low predicted yields are due to lack of moisture during the crop year (317 mm; –0.77 t/ha) and from May to October (246 mm; –0.24 t/ha), as well as high values of the sum of positive temperatures (2527.5 °C; –0.13 t/ha), low precipitation during wheat growing season (175 mm; –0.10 t/ha) and lack of application of plant protection products (–0.10 t/ha). The model expands the possibilities of applying machine learning, allowing us to obtain more reliable and informative results.

Ключевые слова: объяснимый искусственный интеллект, урожайность яровой пшеницы, многофакторные полевые опыты, XGBoost, SHAP.

Keywords: explicable artificial intelligence, spring wheat yield, multifactorial field experiments, XGBoost, SHAP.

Длительные многофакторные полевые опыты, охватывающие различные почвенно-климатические зоны, служат основой для формирования знаний в сельскохозяйственной науке и практике. Осуществление

многолетних экспериментов и наблюдений позволяет получать данные об эффективности возделывания сельскохозяйственных культур в зависимости от уровня применения агрохимических средств [1, 2, 3], параметров

технологических операций и почвенно-климатических условий конкретного пространственного объекта (территории) [4, 5, 6]. В мировой и отечественной практике с использованием аналитики данных, полученных в полевых опытах, и агрономических знаний создается система рекомендаций по выращиванию культур – инструмент поддержки принятия решений для агрономов. Составляющим звеном этого инструмента выступает моделирование (прогноз) возможной урожайности культур в зависимости от условий окружающей среды и управляющих воздействий.

Прогнозирование урожайности сельскохозяйственных культур становится одним из ключевых элементов в процессе цифрового управления посевами, оценке рентабельности производства и производственных рисков, а также распределении химико-техногенных ресурсов в конкретном землепользовании [7, 8]. В предыдущие годы урожайность сельскохозяйственных культур прогнозировали посредством статистических моделей с использованием различных данных. Например, неплохо развита система прогнозирования урожайности сельскохозяйственных культур в масштабе регионов (субъектов РФ) с использованием данных по метеопараметрам и дистанционному зондированию Земли. Обычно анализируется временной ход NDVI по фазам вегетации, определяется теснота связи NDVI с урожайностью по этим периодам в комплексе с метеорологическими условиями. Затем формируются регрессионные модели прогнозирования урожайности на основе совместного использования наземных и спутниковых данных [9]. Разрабатываются также математические модели прогнозирования урожайности сельскохозяйственных культур, основанные на агрохимических показателях почвы, – гумус, обменный калий (K), доступный фосфор (P) и pH солевой вытяжки [10, 11]. Однако прогнозирование урожайности культур с использованием традиционных регрессионных моделей представляет собой ряд проблем, в первую очередь из-за их неспособности обрабатывать большие наборы данных и эффективно улавливать влияние условий окружающей среды.

Модели искусственного интеллекта (ИИ) открывают новый путь для обработки обширных наборов данных и извлечения информации из полевых агрономических экспериментов [12, 13]. Эти вычислительно эффективные модели способны анализировать сложные взаимодействия между признаками и условиями окружающей среды, тем самым улучшая понимание моделей и повышая точность прогнозов [14, 15, 16]. Применение технологий ИИ, преимущественно методов машинного обучения (МО), для анализа сельскохозяйственных данных делает более доступным и научно-обоснованным их преобразование в прикладные инструменты для поддержки принятия решений при реализации агротехнологий. В то же время модели прогнозирования урожайности сельскохозяйственных культур, основанные на сложных алгоритмах МО, например ансамблевых методах, несмотря на высокую точность прогнозирования, часто непрозрачны и не поддаются объяснению. Такие модели получили название «черного ящика». Поэтому возникла потребность в их интерпретируемости, что может быть решено с использованием объяснимого ИИ [17, 18]. Интерпретируемая модель имеет решающее значение, поскольку она позволяет понять важные факторы и условия, влияющие на урожайность сельскохозяйственных культур, и обосновать конкретные прогнозы. В дополнение к прогнозам с использованием МО для получения объяснений и интерпретаций поведения моделей применяют аддитивные объяснения

Шепли (SHapley Additive ExPlanations – SHAP) [19]. Использование значений SHAP позволяет оценить вклад каждого признака в модели МО и понять, какие признаки влияют на результаты ее работы [20]. В российском научном дискурсе SHAP используется в электроэнергетике [21, 22], телекоммуникации [23], медицине [24] и др. Публикаций по использованию метода SHAP для интерпретации прогнозных моделей МО в сельском хозяйстве нами не обнаружено.

Урожайность яровой пшеницы зависит от условий окружающей среды и различных управляющих воздействий, обычно называемых в моделях МО признаками. Использование интерпретируемых алгоритмов имеет важное значение для понимания сложных взаимодействий между различными признаками и их влияния на урожайность культуры. Применение новых подходов интеллектуального анализа данных полевых экспериментов существенно увеличивает их информативность и имеет важное значение для обеспечения надежности и согласованности прогнозов.

Цель исследования – разработать интерпретируемую модель машинного обучения для прогнозирования урожайности яровой пшеницы с использованием данных длительного многофакторного полевого опыта.

Методика. Машинное обучение выполняли с использованием данных длительного полевого опыта, проведенного Федеральным Алтайским научным центром агробиотехнологий в 2001–2024 гг. в лесостепи Алтайского Приобья. Почва опытного участка – чернозем выщелоченный среднесуглинистый малогумусный среднесуглинистый на склоне юго-восточной экспозиции крутизной 1...2°. Содержание гумуса в пахотном слое почвы составляло 3,8 %.

Набор данных состоял из 3456 записей, содержащих информацию об урожайности пшеницы и факторах, влияющих на нее (всего 18 признаков). Признаки включали категориальные переменные: предшественники (горох, овес, пар, бессменная пшеница); приемы основной обработки почвы (без обработки – БО, глубокая (на 25...27 см) плоскорезная – ГПО, мелкая (на 14...16 см) плоскорезная – МПО); удобрения (без удобрений – удобрение₀, P₂₅, N₄₆P₂₅); средства защиты растений (без средств защиты – СЗР₀, гербицид против двудольных сорняков – Г-1, гербициды против двудольных и однодольных сорняков – Г-2, гербициды, инсектициды, фунгициды – ГИФ). Числовые признаки представлены урожайностью пшеницы (т/га), осадками сельскохозяйственного года (O_{схг}, мм), осадками за апрель-октябрь (O_{4...10}, мм), осадками за май-август (O_{5...8}, мм) и суммой положительных температур за вегетационный период (СПТ, °C). Временные ряды осадков и положительных температур воздуха получены на основании анализа данных агрометеорологической станции (АМС) Барнаул Алтайского ЦГМС (источник данных: web-ресурс «<http://www.pogodaiklimat.ru>»).

Для формирования модели прогнозирования урожайности культуры использовали экстремальный градиентный бустинг (XGBoost) – усовершенствованный алгоритм МО, представляющий собой оптимизированную распределенную реализацию градиентного бустинга. XGBoost эффективно применяет принципы параллельного построения деревьев решений (известного также как GBDT или GBM) в рамках парадигмы градиентного бустинга. Общая формула модели градиентного бустинга выглядит следующим образом:

$$f(x) = \sum_{m=1}^M \gamma_m h_m(x), \quad (1)$$

где $h_m(x)$ – базовый предиктор (дерево решений), γ_m – коэффициент усадки, регулирующий вклад каждого дерева, M – число итераций бустинга [25].

Оценку производительности модели XGBoost проводили с использованием трех статистических параметров – средняя абсолютная ошибка (MAE), корень из среднеквадратичной ошибки (RMSE) и коэффициент детерминации (R^2).

Для интерпретации модели МО использовали аддитивные объяснения Шепли (SHAP). SHAP использует теорию игр (значения Шепли) для объяснения вклада каждого признака в предсказание модели. Хотя значения Шепли выступают более общей концепцией, SHAP специально фокусируется на интерпретации и объяснении моделей МО [25, 26]. Значение SHAP определяли следующим уравнением:

$$S_i(x) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(x_S \cup \{i\}) - f(x_S)], \quad (2)$$

где $S_i(x)$ – значение SHAP для i -го признака и наблюдения x ; N – множество признаков, $|N|$ – размер множества признаков; S – подмножество признаков, исключая i -й признак; $|S|$ – размер подмножества S ; x_S – наблюдение x с учетом только признаков из множества S ; $f(x_S \cup \{i\})$ – прогноз модели для наблюдения x , включая i -й признак; $f(x_S)$ – прогноз для наблюдения x_S .

Для построения прогнозной модели данные разделили в соотношении 80:20, где 80 % использовали для обучения, а оставшиеся 20 % – для тестирования качества модели. Настройку гиперпараметров проводили с использованием методов перекрестной проверки, что позволило оптимизировать производительность модели и избежать переобучения. Среди ключевых гиперпараметров, которые были скорректированы, выделяются количество деревьев в ансамбле, максимальная глубина деревьев и минимальное количество наблюдений, необходимое для разделения узла дерева.

Статистические расчеты и графические построения выполнены средствами языка программирования Python в интерактивном блокноте Jupyter. Модель прогнозирования урожайности яровой пшеницы реализована в авторской программе Crop Yield Analysis & Forecast (CYAF) [27, 28].

Результаты и обсуждение. Предварительная обработка выступает важным этапом в уточнении исходного набора данных, создавая среду, подходящую для эффективного анализа и моделирования методами МО. Этот ключевой шаг охватывает различные процессы, направленные на повышение качества и релевантности данных. Была проведена очистка данных с целью выявления и исправления отсутствующих или несоответствующих значений. Также были проведены обнаружение и удаление выбросов для предотвращения искажений в обучении модели и нормализация данных.

Основные статистические показатели категориальных признаков были проанализированы и использованы для интерпретации влияния различных признаков на урожайность пшеницы. Анализ диаграмм размахов или графиков «ящик с усами» (рис. 1) позволяет оценить влияние каждого категориального признака на урожайность.

Из графика видно, что пар был наиболее благоприятным предшественником, обеспечивая максимальную среднюю урожайность (2,31 т/га) и медиану (2,31 т/га). В то же время пшеница как предшественник при бесменном возделывании демонстрировала наименьшие

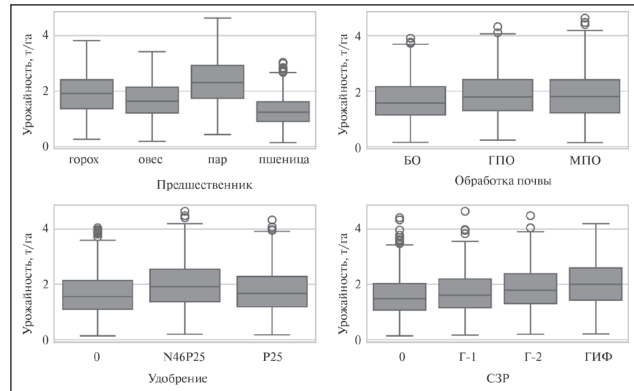


Рис. 1. Диаграммы размахов данных по урожайности яровой пшеницы.

в опыте значения: средняя урожайность составила 1,29 т/га, а медиана – 1,24 т/га. Горох и овес занимали промежуточное положение по влиянию на урожайность.

При сравнении способов обработки почвы наивысшая в эксперименте урожайность выявлена при глубокой (1,87 т/га) и мелкой плоскорезной (1,85 т/га) обработках, при этом медианные значения близки к средним (1,79 и 1,80 т/га соответственно). Наименьшие значения урожайности зафиксированы при отсутствии основной обработки почвы.

Наибольшую урожайность яровой пшеницы обеспечивала доза минеральных удобрений $N_{46}P_{25}$. Без внесения удобрений или при использовании только фосфорных удобрений (P_{25}) урожайность снижалась. Это подтверждает важность сбалансированного внесения азота и фосфора для повышения урожайности зерна пшеницы.

Применение средств защиты растений (СЗР) также положительно влияло на урожайность культуры. Наибольший в эксперименте эффект достигался при комплексном использовании гербицидов, инсектицида и фунгицида, где средняя урожайность составляла 2,04 т/га.

Кроме того, анализ числовых признаков показал, что урожайность культуры изменялась от 0,14 до 4,64 т/га, с медианой в 1,72 т/га. Осадки за сельскохозяйственный год варьировали от 317 до 596,5 мм, с медианой в 441,2 мм. Осадки с апреля по октябрь составляли 213,8...436,3 мм, с медианой в 298,3 мм. Осадки с мая по август находились в диапазоне от 135,6 до 288,8 мм, с медианой в 206,4 мм. Сумма положительных температур изменялась от 2043,3 до 2536,5 °C. Эти данные позволили провести анализ диапазонов значений климатических признаков и их связи с урожайностью культуры.

Корреляционный анализ методом Пирсона выявил статистически значимые связи: положительную корреляцию между количеством осадков и урожайностью ($p < 0,05$) и отрицательную – между суммой положительных температур и урожайностью ($p < 0,05$). Эти результаты подчеркивают важность учета климатических факторов при прогнозировании урожайности культуры.

Результаты дисперсионного анализа методом многовариантного ANOVA для категориальных признаков указывают на их значимое влияние на урожайность ($p < 0,05$): предшественник – $F = 333,96$ ($p < 0,001$), обработка почвы – $F = 22,51$ ($p < 0,001$), удобрение – $F = 29,56$ ($p < 0,001$), СЗР – $F = 56,52$ ($p < 0,001$). Эти результаты подтверждают, что все перечисленные признаки (18 шт.) в той или иной степени влияли на урожайность пшеницы, что важно учитывать при разработке прогностических моделей.

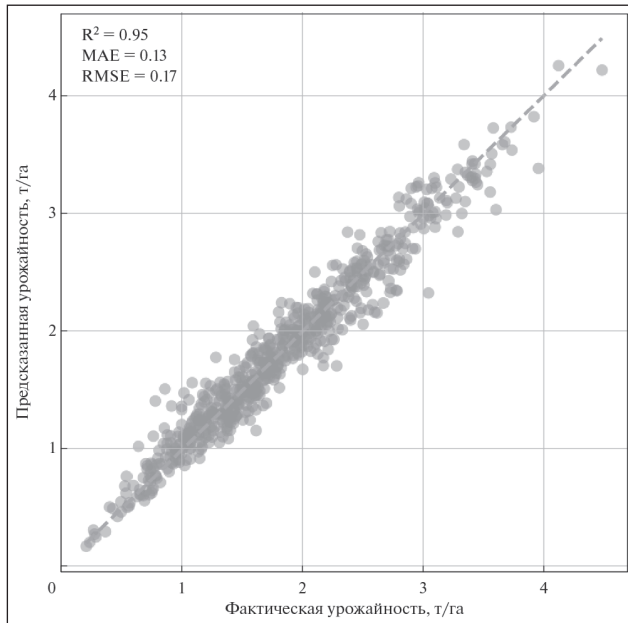


Рис. 2. Сравнение фактических и предсказанных значений урожайности яровой пшеницы.

Сравнение результатов расчета вклада признаков в целевую переменную традиционными статистическими методами (дисперсионный, регрессионный и корреляционный анализы) и оценки важности признаков алгоритмом XGBoost показало схожие закономерности: осадки за сельскохозяйственный год (важность рассчитанного с использованием алгоритма XGBoost признака – 0,2037), предшественник – пар (0,1953), предшественник – пшеница (0,1783), сумма положительных температур (0,1453), осадки с мая по август (0,0949), осадки с мая по октябрь (0,0808), обработка почвы – БО (0,0788), СЗР – ГИФ (0,0712), СЗР – 0 (0,0608), а важность остальных признаков не превышала значение 0,04.

Для построения модели МО были использованы все признаки, полученные с применением временных рядов данных полевого эксперимента. Построенная модель XGBoost продемонстрировала высокую точность прогнозирования урожайности яровой пшеницы. Основные метрики качества модели позволяют судить о ее эффективности: коэффициент детерминации (R^2) составил 0,95, средняя абсолютная ошибка (MAE) – 0,13 т/га, среднеквадратичная ошибка (RMSE) – 0,17 т/га. Коэффициент детерминации свидетельствует о том, что модель объясняет 95 % вариаций в урожайности пшеницы.

Для визуальной оценки качества модели был построен график сравнения фактических и предсказанных значений урожайности (рис. 2). На представленном графике видно, что предсказанные значения располагаются вдоль линии идеального предсказания, что подтверждает высокую точность модели. Важно отметить, что значения равномерно распределены вдоль всей линии без значительных выбросов (при урожайности от 1,0 до 3,0 т/га), что свидетельствует о стабильности модели на различных уровнях урожайности культуры.

Расчет значений SHAP предполагает оценку вклада каждого признака в модель МО с учетом всех возможных комбинаций признаков. Для визуализации значений SHAP строятся соответствующие графики, которые помогают понять вклад входных признаков в прогноз модели для конкретного наблюдения. Значения SHAP

показывают, как каждый признак влияет на итоговый прогноз – положительно или отрицательно, а также с какой силой. График SHAP отображает признаки на оси x и соответствующие им значения SHAP на оси y (рис. 3). Каждая точка на графике представляет значение SHAP для конкретного признака в конкретном наблюдении. Анализируя эти графики, можно получить представление о внутренней работе сложных моделей МО и больше доверять их прогнозам.

SHAP-анализ выявил отчетливую положительную зависимость между количеством осадков за сельскохозяйственный год и урожайностью пшеницы. Высокие значения осадков (отмеченные серым цветом) преимущественно смещены в правую часть графика, что свидетельствует об увеличении связи с прогнозируемой урожайностью. Это наблюдение подтверждает, что для условий Алтайского Приобья атмосферное увлажнение выступает ключевым лимитирующим признаком, и увеличение количества осадков практически всегда положительно сказывается на урожайности культуры. Количеством этот признак имеет наибольшее среднее абсолютное значение SHAP (0,204 т/га) и максимальный диапазон влияния от –0,927 до 1,077 т/га, что подчеркивает его доминирующую роль в модели.

График также наглядно продемонстрировал влияние пара в качестве предшественника на урожайность пшеницы. Серые точки на графике соответствуют случаям использования пара в качестве предшественника (значение 1), а черные – остальным предшественникам (значение 0). На графике отчетливо видно, что серые точки преимущественно расположены в правой части, что указывает на положительное влияние пара в качестве предшественника на урожайность. Со средним абсолютным значением SHAP 0,195 т/га, этот признак выступает вторым по вкладу в прогнозную способность

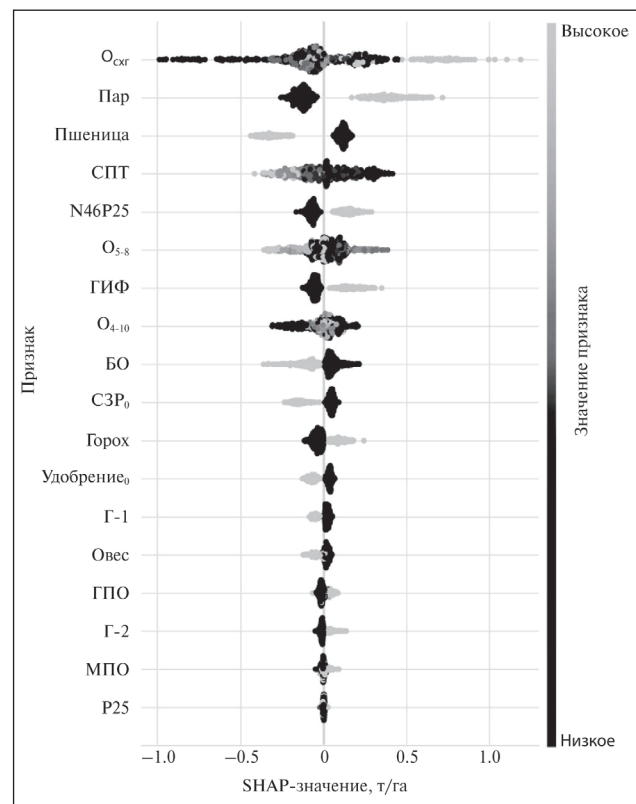


Рис. 3. Вклад различных признаков в прогнозирование урожайности пшеницы.

модели. Выявлено негативное влияние бессменного выращивания пшеницы. На графике преобладают серые точки в левой части, что свидетельствует о снижении урожайности культуры. Среднее абсолютное значение SHAP для этого признака составляет 0,178 т/га, что делает его третьим по величине вклада среди всех признаков модели. Показательно, что максимальное в эксперименте отрицательное влияние (–0,501 т/га) по модулю бессменной пшеницы превышает максимальное положительное (0,182 т/га), что подчеркивает преимущественно негативное воздействие на урожайность культуры.

Суммы положительных температур воздуха имеют нелинейную связь с урожайностью пшеницы. Как низкие, так и высокие значения температур негативно влияют на урожайность, в то время как средние значения (точки смешанного цвета в правой части графика) способствуют увеличению урожайности. Диапазон влияния изменяется от –0,396 до 0,425 т/га, а среднее абсолютное значение SHAP (0,145 т/га) подтверждает значимость этого признака для формирования урожайности. Осадки с мая по август показали положительную связь с урожайностью культуры, особенно заметную при высоких значениях осадков. Среднее абсолютное значение SHAP составляет 0,095 т/га, что подтверждает существенный, хотя и меньший по сравнению с осадками сельскохозяйственного года, вклад этого признака. Схожую тенденцию демонстрировали осадки апреля–октября (0,081 т/га), хотя их среднее значение SHAP слегка отрицательно (–0,008 т/га), что указывает на более сложный характер влияния осадков в течение всего вегетационного периода.

Анализ показал, что отсутствие основной обработки почвы преимущественно связано со снижением урожайности. Это объясняется тем, что без обработки в условиях Алтайского края происходит худшее накопление влаги в осенне-зимний период и более позднее прогревание почвы весной. Среднее абсолютное значение SHAP для этого признака составило 0,079 т/га, а диапазон влияния варьировал от –0,384 до 0,222 т/га, что указывает на значительную вариативность эффекта в зависимости от сочетания с другими признаками.

Применение комплекса средств защиты растений (ГИФ) в большинстве случаев положительно влияло на урожайность. Это объясняется тем, что эффективная защита от вредителей, болезней и сорняков позволяет растениям реализовать свой генетический потенциал продуктивности. При значении SHAP в 0,071 т/га влияние этого признака может достигать значительного положительного эффекта (максимум 0,273 т/га).

Анализ показал вариативное влияние уровня минерального питания растений на урожайность культуры в вариантах без внесения минеральных удобрений, преимущественно негативное, особенно в сочетании с другими неблагоприятными признаками. Это подтверждает необходимость сбалансированного минерального питания для формирования урожайности пшеницы. Значение SHAP этого признака составляет 0,061 т/га, а диапазон влияния от –0,191 до 0,089 т/га указывает на преобладающее потенциально негативное воздействие.

Остальные признаки, включающие различные дозы удобрений ($N_{46}P_{25}$, P_{25}), средства защиты растений (Г-1, Г-2) и приемы обработки почвы (глубокая и мелкая плоскорезная), имеют меньшее среднее абсолютное значение SHAP (0,039 т/га и ниже). Это свидетельствует о том, что их вклад в предсказания модели менее выражен, по сравнению с ключевыми климатическими факторами. Однако их присутствие в модели и положительные максимумы значений SHAP свидетель-

ствуют о том, что при определенных условиях они могут оказывать существенное положительное влияние на урожайность пшеницы. Например, удобрение $N_{46}P_{25}$ с SHAP-значением 0,0975 т/га может давать положительный эффект до 0,288 т/га, что сопоставимо с влиянием более важных признаков.

Для более детального выяснения связи признаков с различным уровнем урожайности яровой пшеницы были построены каскадные графики на основе данных SHAP-анализа для высокой (рис. 4) и низкой (рис. 5) урожайности.

В результате этого анализа установлено, что базовое значение урожайности пшеницы составило 1,793 т/га, $\sum SHAP$ равно 2,418 т/га, прогноз: $1,793 + 2,418 = 4,211$ т/га. SHAP-анализ показал, что высокая урожайность обусловлена следующими признаками с наибольшим положительным влиянием (рис. 4):

$O_{ср}$ – высокое значение осадков (596,5 мм) дало наибольший положительный вклад (SHAP=1,19 т/га). Среднее абсолютное значение SHAP по всей модели составляет 0,23 т/га, что делает этот признак самым важным;

пар – использование пара в качестве предшественника внесло значительный вклад в урожайность культуры (SHAP=0,58 т/га), что выше среднего абсолютного значения SHAP по модели (0,21 т/га);

$N_{46}P_{25}$ – применение азотно-фосфорных удобрений внесло существенный положительный вклад (SHAP=0,21 т/га). Среднее абсолютное значение SHAP по модели для этого признака составляет 0,10 т/га, что подтверждает его значимость в этом наблюдении;

бессменная пшеница – этот предшественник внес положительный вклад (SHAP=0,13 т/га). Однако в среднем по модели он имеет смешанное влияние. Среднее абсолютное значение SHAP составляет 0,17 т/га, но максимальное отрицательное значение SHAP (–0,44 т/га) превышает максимальное положительное (0,17 т/га). Это указывает на то, что эффект бессменного возделывания пшеницы зависит от сочетания с другими признаками. В этой модели повышенное количество атмосферного увлажнения и применение удобрений могли компенсировать потенциально негативное влияние ее бессменного возделывания;

$O_{4...10}$ – высокое значение осадков в период с апреля по октябрь (436,3 мм) способствовало увеличению урожайности (SHAP=0,12 т/га). Хотя среднее абсолютное значение SHAP по модели ниже (0,08 т/га), в нашем наблюдении этот признак оказался более значимым;

БО – прием отказа от обработки почвы внес положительный вклад (SHAP=0,12 т/га). Однако в среднем по модели его влияние более вариативное: среднее абсолютное значение SHAP составляет 0,07 т/га с диапазоном от –0,36 до 0,21 т/га. По-видимому, в условиях повышенного атмосферного увлажнения и применения удобрений оставление почвы без обработки не оказывает отрицательного влияния на урожайность пшеницы.

Низкая прогнозная урожайность пшеницы представляется собой противоположный сценарий, где агрометеорологические условия и агротехнические факторы оказывают негативное влияние на продуктивность.

Базовое значение урожайности пшеницы составило 1,793 т/га, $\sum SHAP$ равна –1,63 т/га, прогноз: $1,793 - 1,63 = 0,161$ т/га. SHAP-анализ показал, что низкая урожайность культуры обусловлена следующими признаками с наибольшим отрицательным влиянием (рис. 5):

$O_{ср}$ – низкое значение осадков за сельскохозяйственный год (317 мм) оказало наибольшее отрицательное

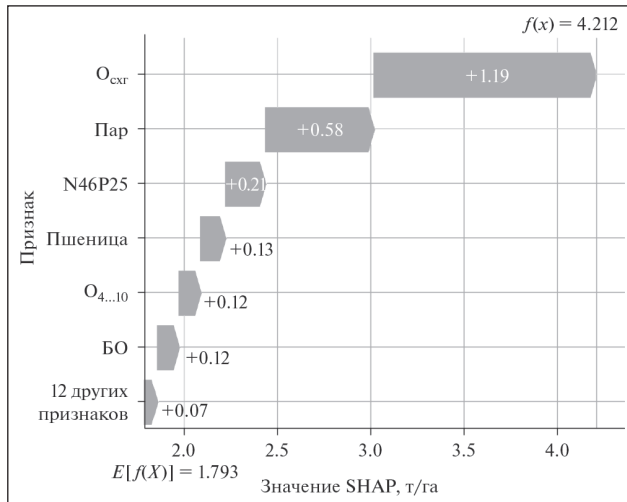


Рис. 4. Каскадная диаграмма значений SHAP для прогнозной модели высокой урожайности пшеницы.

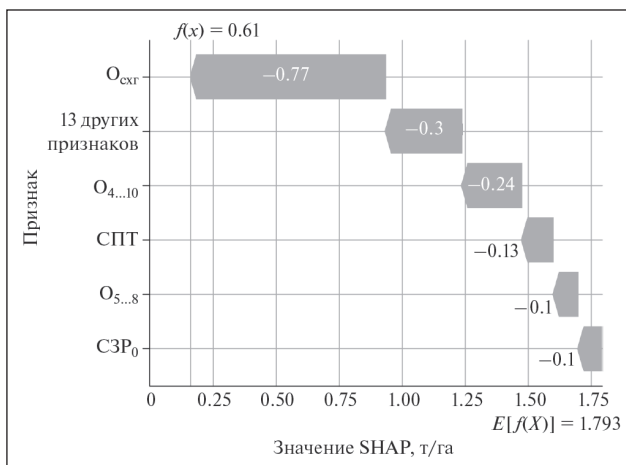


Рис. 5. Каскадная диаграмма значений SHAP для прогнозной модели низкой урожайности пшеницы.

влияние (SHAP=-0,77 т/га). Для сравнения, среднее абсолютное значение SHAP по всей модели составляет 0,23 т/га, что подчеркивает роль влагообеспеченности в варьировании урожайности культуры;

O_{4...10} – недостаток атмосферных осадков в апреле-октябре (246 мм) также снизил урожайность (SHAP=-0,24 т/га). Среднее абсолютное значение SHAP по модели ниже (0,08 т/га), что указывает на значимость этого периода для урожайности пшеницы;

СПТ – высокое значение суммы положительных температур (2527,5 °C) оказало умеренно отрицательное влияние (SHAP=-0,13 т/га). Для сравнения, среднее абсолютное значение SHAP по всей модели составляет 0,14 т/га;

O_{5...8} – низкое количество осадков в период вегетации пшеницы (175 мм) внесло отрицательный вклад (SHAP=-0,10 т/га). Среднее абсолютное значение SHAP для этого признака по всей модели составляет 0,09 т/га;

СЗР₀ – отсутствие применения средств защиты растений также внесло отрицательный вклад (SHAP=-0,10 т/га). Среднее абсолютное значение SHAP по модели составляет 0,07 т/га.

Разница в значениях SHAP между общим анализом построенной модели (рис. 3) и моделями с высокой (рис. 4) и низкой урожайностью культуры (рис. 5) объяс-

няется несколькими ключевыми моментами. Во-первых, значения SHAP следует считать контекстуальными, то есть их рассчитывают для каждого конкретного наблюдения, и они зависят от взаимодействия признаков. В общем анализе значения SHAP усредняют по всей выборке, что может не отражать индивидуальные особенности каждого наблюдения. Например, в случае высокой урожайности благоприятные климатические условия (годы с достаточным и повышенным увлажнением) и применение удобрений могут нивелировать негативное влияние повторного размещения пшеницы, что не всегда отражается в средних значениях по всей модели.

Во-вторых, взаимодействие признаков играет значительную роль в формировании значений SHAP. В конкретных случаях благоприятные или неблагоприятные условия могут усиливать или компенсировать влияние отдельных признаков. Например, при высокой урожайности осадки за сельскохозяйственный год (O_{схг}) имели SHAP 1,19 т/га, что значительно выше среднего абсолютного значения SHAP по всей модели (0,23 т/га). Это подчеркивает, что в конкретных наблюдениях определенные признаки могут оказывать более выраженное влияние на урожайность.

Кроме того, различия в значимости признаков между общим анализом и конкретными наблюдениями также объясняются изменчивостью данных. Значения SHAP могут варьировать из-за случайности и вариативности данных, что в конкретных случаях может привести к более выраженному влиянию определенных признаков. Например, в случае прогнозирования низкой урожайности культуры недостаток осадков с апреля по октябрь (O_{4...10}) оказывает значительное отрицательное влияние (SHAP=-0,24 т/га), хотя среднее абсолютное значение SHAP этого признака в целом по модели меньше (0,08 т/га).

SHAP-анализ позволяет детально оценить вклад каждого признака в предсказание модели, но его результаты зависят от контекста и взаимодействия признаков. Конкретные наблюдения могут демонстрировать более выраженное влияние определенных условий и факторов, по сравнению с общим анализом.

В целом, интеграция модели XGBoost с SHAP-анализом показала, что урожайность яровой пшеницы в условиях Алтайского Приобья определяется сложным взаимодействием климатических и агротехнических признаков. Агротехнические приемы могут усиливать положительное влияние благоприятных погодных условий, а также нивелировать негативное воздействие неблагоприятных. Это наблюдение имеет практическое значение для гибкой адаптации технологии возделывания яровой пшеницы к конкретным условиям региона и текущей агрометеорологической ситуации.

Выводы. Интеграция SHAP в структуру предиктивного моделирования значительно повысила интерпретируемость модели прогнозирования урожайности яровой пшеницы, созданной с использованием алгоритма XGBoost. Анализ значений SHAP способствовал более глубокому пониманию предиктивной модели и существенно расширил возможности по проверке эффективности МО в реальных сценариях. Интерпретируемая модель МО позволяет также дифференцировать признаки и выделять наиболее важные из них.

В условиях лесостепи Алтайского Приобья общий SHAP-анализ выявил ключевые признаки, определяющие прогнозную урожайность яровой пшеницы. Десять наиболее значимых признаков, с вкладом более 0,07 т/га, можно расположить в следующий ряд в порядке убывания влияния: атмосферное увлажнение за сельскохозяйственный год, пар и пшеница в качестве

предшественников, сумма положительных температур, удобрение $N_{46}P_{25}$, атмосферное увлажнение за период с мая по август, средства защиты растений ГИФ, осадки за период с апреля по октябрь, отсутствие обработки почвы и защиты растений.

Высокая прогнозная урожайность культуры в условиях полевого опыта достигалась при достаточном атмосферном увлажнении за сельскохозяйственный год (596,5 мм; 1,19 т/га), использовании пара в качестве предшественника (0,58 т/га) и применении азотно-фосфорных удобрений (0,21 т/га). Умеренное положительное влияние оказывали варианты опыта с пшеницей в качестве предшественника (0,13 т/га), отсутствие обработки почвы (0,12 т/га) и атмосферное увлажнение апреля–октября (436,3 мм; 0,12 т/га). Низкая прогнозная урожайность яровой пшеницы обусловлена недостатком увлажнения в течение сельскохозяйственного года (317 мм; –0,77 т/га) и с апреля по октябрь (246 мм; –0,24 т/га). Также умеренное отрицательное влияние оказали высокая сумма положительных температур (2527,5 °С; –0,13 т/га), низкое количество осадков в период вегетации пшеницы (175 мм; –0,10 т/га) и отсутствие средств защиты растений – 0,10 т/га).

Разработанная модель на основе данных временных рядов длительных полевых опытов позволяет проводить детальный анализ того, как различные факторы окружающей среды и управляющие воздействия влияют на урожайность яровой пшеницы. Понимая поведение модели, можно принимать решения на основе ее выводов, а также использовать результаты моделирования при разработке агрономических рекомендаций.

ФИНАНСИРОВАНИЕ РАБОТЫ.

Работа финансировалась за счет средств бюджета Сибирского федерального научного центра агробиотехнологий РАН и Федерального Алтайского научного центра агробиотехнологий в рамках Государственных заданий Министерства науки и высшего образования РФ. Никаких дополнительных грантов на проведение или руководство этим конкретным исследованием получено не было.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ.

В работе отсутствуют исследования человека или животных.

КОНФЛИКТ ИНТЕРЕСОВ.

Авторы работы заявляют, что у них нет конфликта интересов.

Литература.

1. Влияние длительного применения органических и минеральных удобрений на урожайность и качество яровой пшеницы / И. В. Понкратенкова, А. Ю. Гаврилова, Г. Е. Мерзлая и др. // *Аграрный вестник Урала*. 2019. № 7(186). С. 39–44.
2. Эффективность использования азота в длительных и краткосрочных опытах агрохимслужбы и Геосети Российской Федерации / В. А. Романенков, М. В. Беличенко, О. В. Рухович и др. // *Агрохимия*. 2020. № 12. С. 28–37. doi: 10.31857/S0002188120120091.
3. Зерновая продуктивность свекловичных севооборотов в зависимости от степени биологизации в условиях Центрального Черноземья / А. С. Акименко, В. И. Свиридов, Т. А. Дудкина и др. // *Земледелие*. 2022. № 3. С. 12–18.
4. Сычев В. Г., Беличенко М. В., Романенков В. А. Результаты мониторинга урожайности сельскохозяйственных культур, продуктивности севооборотов и изменения свойств почв в длительных опытах Географической сети // *Плодородие*. 2017. № 6(99). С. 2–5.
5. Рублюк М. В., Иванов Д. А. Изменение свойств дерново-подзолистого почвенного слоя в зависимости от условий осушаемого агроландшафта при возделывании овса в фитоценозе с травами // *Земледелие*. 2023. № 3. С. 8–12.
6. Подлесных И. В., Тарасов С. А., Рубаник Ю. О. Динамика органического углерода почвы в пахотном слое и продуктивность культур почвозащитного агропесопандишафтного комплекса в ЦЧР // *Земледелие*. 2023. № 5. С. 37–41.
7. Bali N., Singla A. Emerging Trends in Machine Learning to Predict Crop Yield and Study Its Influential Factors: A Survey // *Archives of Computational Methods in Engineering*. 2022. Vol. 29. P. 95–112. doi: 10.1007/s11831-021-09569-8.
8. Иванов Д. А., Рублюк М. В., Анциферова О. Н. Прогнозирование размещения посевов льна на основе данных мониторинга и ГИС-технологий // *Земледелие*. 2023. № 7. С. 3–6.
9. Страшная А. И., Береза О. В., Кланг П. С. Прогнозирование урожайности зерновых культур на основе комплексирования наземных и спутниковых данных в субъектах Южного федерального округа // *Гидрометеорологические исследования и прогнозы*. 2021. № 2. С. 111–137. doi: 10.37162/2618-9631-2021-2-111-137.
10. Математические модели и программный комплекс по прогнозированию урожайности сельскохозяйственных культур / Д. А. Благов, С. В. Митрофанов, В. С. Никитин и др. // *Агротехника и энергообеспечение*. 2019. № 3(24). С. 182–188.
11. Прогнозирование урожайности яровой пшеницы по агрохимическим свойствам в условиях Тетюшского района Республики Татарстан / А. А. Лукманов, Ю. П. Переведенцев, А. Б. Мустафина и др. // *Вестник Казанского государственного аграрного университета*. 2023. Т. 18. № 3(71). С. 39–45.
12. Panigrahi B., Kathala K. C. R., Sujatha M. A machine learning-based comparative approach to predict the crop yield using supervised learning with regression models // *Procedia Computer Science*. 2023. Vol. 218. P. 2684–2693. doi: 10.1016/j.procs.2023.01.241.
13. Boppudi S. Deep ensemble model with hybrid intelligence technique for crop yield prediction // *Multimedia Tools and Applications*. 2024. Vol. 83. No. 31. P. 75709–75729. doi: 10.1007/s11042-024-18354-1.
14. Applied Deep Learning-Based Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges / K. Meghraoui, I. Sebari, J. Pilz, et al. // *Technologies*. 2024. Vol. 12. No. 4. P. 43. URL: <https://www.mdpi.com/2227-7080/12/4/43> (дата обращения: 03.03.2025). doi: 10.3390/technologies12040043.
15. Shingade S. D., Mudhalwadkar R. P. Analysis of crop prediction models using data analytics and ML techniques: a review // *Multimedia Tools and Applications*. 2024. Vol. 83. No. 13. P. 37813–37838. doi: 10.1007/s11042-023-17038-6.
16. Ramesh V., Kumaresan P. Stacked Ensemble Model for Accurate Crop Yield Prediction Using Machine Learning Techniques // *Environmental Research Communications*. 2025. Vol. 7. No. 3. URL: <https://iopscience.iop.org/article/10.1088/2515-7620/adb9c0/pdf> (дата обращения: 03.03.2025). doi: 10.1088/2515-7620/adb9c0.
17. Ryo M. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis //

- Artificial Intelligence in Agriculture*. 2022. Vol. 6. P. 257–265. doi: 10.1016/j.iaia.2022.11.003.
18. Cartolano A., Cuzzocrea A., Pilato G. Analyzing and assessing explainable AI models for smart agriculture environments // *Multimedia Tools and Applications*. 2024. Vol. 83. P. 37225–37246. doi: 10.1007/s11042-023-17978-z.
 19. Bifarin O. O. Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification // *Plos one*. 2023. Vol. 18. No. 5. P. e0284315. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0284315> (дата обращения: 03.03.2025). doi: 10.1371/journal.pone.0284315.
 20. Interpretable Machine Learning Techniques for an Advanced Crop Recommendation Model / M. Bouni, B. Hssina, K. Douzi, et al. // *Journal of Electrical and Computer Engineering*. 2024. Vol. 2024. No. 1. P. 7405217. URL: <https://onlinelibrary.wiley.com/doi/full/10.1155/2024/7405217> (дата обращения: 03.03.2025). doi: 10.1155/2024/7405217.
 21. Степанова А. И., Хальясмаа А. И., Матренин П. В. Краткосрочное прогнозирование нагрузки предприятия нефтегазовой промышленности с использованием технологических факторов и аддитивного объяснения Шепли // *Известия высших учебных заведений. Проблемы энергетики*. 2024. Т. 26. № 4. С. 75–88. doi: 10.30724/1998-9903-2024-26-4-75-88.
 22. Горшенин А. Ю., Грицай А. С., Денисова Л. А. Применение машинного обучения деревьев решений для краткосрочного прогнозирования электропотребления // *Известия Тульского государственного университета. Технические науки*. 2023. № 11. С. 226–231. doi: 10.24412/2071-6168-2023-11-226-227.
 23. Бадыкова И. Р., Биктимирова К. Р. Выявление факторов воздействия на сектор связи и телекоммуникаций с применением ансамблевых методов машинного обучения // *π-Econoty*. 2024. Т. 17. № 6. С. 61–78. doi: 10.18721/JE.17604.
 24. Интерпретируемые модели машинного обучения как инструмент объяснения прогнозных оценок в кардиологии / К. И. Шахгельдян, В. Ю. Рублев, Н. С. Куксин и др. // *Вестник современной клинической медицины*. 2025. Т. 18. № 1. С. 98–106. doi: 10.20969/VSKM.2025.18(1).98-106.
 25. Friedman J. H. Greedy function approximation: a gradient boosting machine // *Annals of statistics*. 2001. Vol. 29. No. 5. P. 1189–1232. URL: <https://www.jstor.org/stable/2699986> (дата обращения: 03.03.2025).
 26. From Local Explanations to Global Understanding With Explainable AI for Trees / S. M. Lundberg, G. Erion, H. Chen, et al. // *Nature Machine Intelligence*. 2020. Vol. 2. No. 1. P. 56–67. doi: 10.1038/s42256-019-0138-9.
 27. Программа анализа и прогнозирования урожайности сельскохозяйственных культур (CYAF – Crop Yield Analysis & Forecast) / Д. С. Федоров, О. К. Альсова, В. К. Каличкин и др. // *Свидетельство о государственной регистрации программы для ЭВМ № 2021777894*, 08.11.2021.
 28. Разработка программы анализа и прогнозирования урожайности сельскохозяйственных культур / В. К. Каличкин, Д. С. Федоров, О. К. Альсова и др. // *Достижения науки и техники АПК*. 2022. Т. 36. № 1. С. 51–56. doi: 10.53859/02352451_2022_36_1_51.

Поступила в редакцию 12.03.2025
После доработки 31.03.2025
Принята к публикации 15.04.2025