

ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ И ВЕГЕТАЦИОННЫХ ИНДЕКСОВ ДЛЯ ПРОГНОЗИРОВАНИЯ УРОЖАЙНОСТИ ЯРОВОЙ ПШЕНИЦЫ

✉Гарафутдинова Л.В., Федоров Д.С., Каличкин В.К., Максимович К.Ю., Колбин С.А.

Сибирский федеральный научный центр агробиотехнологий Российской академии наук
Новосибирская область, р.п. Краснообск, Россия

✉e-mail: garafutdinovavl@sfsca.ru

Представлены результаты исследования по разработке и оценке точности предиктивных моделей урожайности яровой пшеницы на основе использования данных дистанционного зондирования Земли и методов машинного обучения. В работе использованы данные урожайности яровой пшеницы сорта Новосибирская 31, полученные в полевом опыте в центральной лесостепи Новосибирской области в 2019–2022 гг. При создании моделей учитывались как качественные предикторы (уровень интенсификации агротехнологий), так и количественные (атмосферные осадки в критические фазы развития растений пшеницы и показатели вегетационных индексов, характеризующих состояние посевов). Использование различных методов интеллектуального анализа данных, а также сочетание параметрических и непараметрических подходов в исследовании обеспечили достаточно высокую точность прогнозирования урожайности яровой пшеницы. Для прогнозирования урожайности яровой пшеницы использовали методы, включающие линейную регрессию, нелинейную модель на основе регрессионного сплайна (Regression Splines), дерево решений (CART), случайный лес (Random Forest), адаптивный бустинг (AdaBoost) и градиентный бустинг (Gradient Boosting). Установлено, что модели на основе алгоритмов случайный лес, градиентный и адаптивный бустинги характеризовались наиболее высокими прогностическими возможностями урожайности культуры в зависимости от складывающихся условий вегетации и управляющего воздействия ($R^2 = 0,74–0,80$). Разработка предиктивных моделей урожайности с использованием дистанционного зондирования Земли и машинного обучения представляют определенную научную новизну и практическую значимость для эффективного управления продуктивностью посевов в изменяющихся почвенно-климатических и хозяйственных условиях. Прогнозное моделирование сталкивается с многоуровневой неопределенностью окружающей среды и высокой вариативностью результирующих показателей на конкретном земельном участке. В связи с этим многоуровневый подход может представлять перспективное решение для эффективного прогнозирования урожайности яровой пшеницы.

Ключевые слова: вегетационные индексы, дистанционное зондирование, машинное обучение, урожайность яровой пшеницы, уровень интенсификации

UTILIZING MACHINE LEARNING AND VEGETATION INDICES FOR SPRING WHEAT YIELD FORECASTING

✉Garafutdinova L.V., Fedorov D.S., Kalichkin V.K., Maksimovich K.Yu., Kolbin S.A.

Siberian Federal Scientific Centre of Agro-BioTechnologies of the Russian Academy of Sciences
Krasnoobsk, Novosibirsk region, Russia

✉e-mail: garafutdinovavl@sfsca.ru

The results of research on the development and assessment of the accuracy of predictive models of spring wheat yield based on the use of remote sensing data and machine learning methods are presented. Yield data of spring wheat variety Novosibirskaya 31 obtained in a field experiment in the central forest-steppe of the Novosibirsk region in 2019–2022 were used in this work. Both qualitative predictors (the level of agrotechnologies intensification) and quantitative predictors (atmospheric precipitation in critical phases of wheat plant development and indicators of vegetation indices characterizing the condition of crops) were taken into account when creating the models. The use of various methods of intellectual data analysis, as well as the combination of parametric and non-parametric approaches in the study provided a sufficiently high accuracy of spring wheat yield forecasting. The methods used

to predict spring wheat yield included linear regression, nonlinear Regression Splines based model, decision tree (CART), Random Forest, Adaptive Boosting (AdaBoost) and Gradient boosting. It was found that the models based on random forest, gradient and adaptive boosting algorithms were characterized by the highest predictive capabilities of crop yield depending on the emerging conditions of vegetation and controlling influence ($R^2 = 0.74\text{--}0.80$). The development of predictive yield models using remote sensing and machine learning represent a certain scientific novelty and practical significance for effective management of crop productivity in changing soil-climatic and economic conditions. Predictive modeling is faced with multilevel environmental uncertainty and high variability of the resulting indicators on a particular land plot. In this regard, the multilevel approach may represent a promising solution for effective forecasting of spring wheat yield.

Keywords: vegetation indices, remote sensing, machine learning, spring wheat yield, level of intensification

Для цитирования: Гарафутдинова Л.В., Федоров Д.С., Каличкин В.К., Максимович К.Ю., Колбин С.А. Использование машинного обучения и вегетационных индексов для прогнозирования урожайности яровой пшеницы // Сибирский вестник сельскохозяйственной науки. 2024. Т. 54. № 1. С. 31–41. <https://doi.org/10.26898/0370-8799-2024-1-4>

For citation: Garafutdinova L.V., Fedorov D.S., Kalichkin V.K., Maksimovich K.Yu., Kolbin S.A. Utilizing machine learning and vegetation indices for spring wheat yield forecasting. *Sibirskii vestnik sel'skokhozyaistvennoi nauki* = *Siberian Herald of Agricultural Science*, 2024, vol. 54, no. 1, pp. 31–41. <https://doi.org/10.26898/0370-8799-2024-1-4>

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Conflict of interest

The authors declare no conflict of interest.

ВВЕДЕНИЕ

Урожайность сельскохозяйственных культур является одним из наиболее важных показателей сельскохозяйственного производства. Прогнозирование урожайности позволяет оптимизировать процессы возделывания сельскохозяйственных культур и принимать обоснованные решения при планировании агротехнологий, применяемых в земледелии. В настоящее время для пространственной оценки состояния растительности активно применяют дистанционное зондирование Земли. Использование спектров электромагнитного излучения в видимой и инфракрасной областях успешно применяется для мониторинга растительного покрова, влажности почвы, недостатка азота и урожайности культур [1, 2].

В частности, для оценки состояния посевов сельскохозяйственных культур широко используются вегетационные индексы, такие как нормализованный разностный ве-

гетационный индекс (Normalized Difference Vegetation Index – NDVI) и нормализованный разностный индекс влажности (Normalized Difference Moisture Index – NDMI)^{1, 2} [3–7]. NDVI представляет собой комбинацию коэффициента отражения в видимой и ближней инфракрасной областях (VIS-NIR), связанного с поглощением специфического солнечного излучения во время фотосинтеза. Главным фактором, определяющим его величину, является тип подстилающей поверхности, а изменения индекса позволяют судить о влиянии неблагоприятных погодных условий, повреждении растений вредителями и болезнями или недостаточном обеспечении питанием [8]. Измерение NDVI имеет решающее значение для мониторинга и управления продуктивностью посевов, так как изменение данного показателя тесно связано с процессом роста растений, их урожайностью и позволяет оценить размеры областей с аномальным развитием культур. Важная особенность NDVI заключается в том, что измене-

¹Сахарова Е.Ю., Сладких Л.А., Кулик Е.Н. Спутниковый мониторинг состояния посевов зерновых культур с использованием индекса вегетации // Интерэкспо ГЕО-Сибирь. 2015. Т. 4. № 1. С. 47–52.

²Тельнова Н.О. Выявление и картографирование многолетних трендов NDVI для оценки вклада изменений климата в динамику биологической продуктивности агроэкосистем лесостепной и степной зон Северной Евразии // Современные проблемы дистанционного зондирования Земли из космоса. 2017. Т. 14. № 6. С. 97–107. DOI: 10.21046/2070-7401-2017-14-6-97-107.

ния значений индекса становятся более заметными в засушливых условиях вегетации из-за тесной связи с влагообеспеченностью посевов³ [9, 10]. В дополнение к NDVI для более точной оценки общей влагообеспеченности посевов используется NDMI. Этот индекс представляет собой показатель, который дает возможность оценить степень водного стресса растительности^{4, 5}.

Недостатком вегетационных индексов является то, что они предоставляют косвенную информацию о фактической урожайности культур. Для повышения точности прогнозирования урожайности сельскохозяйственных культур необходимо использовать не только вегетационные индексы, но и дополнительную информацию об окружающей среде и управляющих воздействиях. В этом случае целесообразно при формировании предиктивных моделей урожайности культур использовать возможности машинного обучения, одного из разделов искусственного интеллекта. Среди наиболее известных и применяемых алгоритмов машинного обучения можно назвать нейронные сети (NN), случайный лес (RF), метод опорных векторов (SVM) и экстремальный градиентный бустинг (XGB). Эти модели используют метеорологические данные, методы управления и вегетационные индексы [11–13]. В целом сравнительный анализ предиктивных моделей урожайности яровой пшеницы с использованием вегетационных индексов и машинного обучения может быть полезным для определения наиболее эффективного подхода к обоснованию оптимальных управляющих воздействий.

Цель исследований – разработка моделей прогнозирования урожайности яровой пшеницы на основе использования данных полевых опытов, дистанционного зондирования Земли и применения методов машинного обучения.

МАТЕРИАЛ И МЕТОДЫ

Для построения моделей машинного обучения использовали данные урожайности яровой пшеницы сорта Новосибирская 31, полученные в 2019–2022 гг. в полевом стационарном опыте СФНЦА РАН. Опыт расположен в ОС «Элитная» СФНЦА РАН в центральной лесостепи Новосибирской области (54°55'27.45" с. ш., 82°57'14.05" в. д.). Почва опытного участка – чернозем выщелоченный среднесуглинистый, среднегумусный. Опыт представлял собой бессменный посев яровой пшеницы, возделываемой на двух уровнях интенсификации, которые можно условно охарактеризовать как малоинтенсивный (ежегодное внесение стартовой дозы N30) и интенсивный (внесение N60 на планируемую урожайность). В опыте использовали комплекс средств защиты растений от сорняков, вредителей и болезней (гербициды, фунгициды и инсектициды). Основная обработка почвы проводилась стойками СибИМЭ на глубину 25–27 см.

Данные дистанционного зондирования посевов яровой пшеницы, находящиеся в открытом доступе, получены с использованием sentinel-hub (<https://www.sentinel-hub.com>) со спутника Sentinel-2 с высоким пространственным разрешением (10 м). Многозональные космические снимки состоят из 13 спектральных каналов с разрешением 10, 30 и 60 м. Для анализа использовали снимки, полученные в годы наблюдений за июнь и июль, со следующими спектральными диапазонами: красный – 650,0–680,0 нм, инфракрасный – 784,5–899,5 нм и коротковолновый инфракрасный – 1564,5–1654,5 нм. Обработку полученных материалов (36 безоблачных изображений) проводили с помощью QGIS с открытой архитектурой (<https://qgis.org/ru/site/>). Для отображения состояния расти-

³De Keersmaecker W., Lhermitte S., Hill M.J., Tits L., Coppin P., Somers B. Assessment of regional vegetation response to climate anomalies: A case study for Australia using GIMMS NDVI time series between 1982 and 2006 // Remote Sensing. 2017. Vol. 9. N 1. P. 34. DOI: 10.3390/rs9010034.

⁴Wang S., Baig M.H.A., Zhang L., Jiang H., Ji Y., Zhao H., Tian J. A simple enhanced water index (EWI) for percent surface water estimation Landsat data // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2015. Vol. 8. P. 90–97. DOI: 10.1109/JSTARS.2014.2387196.

⁵McCall D.S., Zhang X., Sullivan D.G., Askew S.D., Ervin E.H. Enhanced soil moisture assessment using narrowband reflectance vegetation indices in creeping bentgrass // Crop Science. 2017. Vol. 57. N S1. P. 161–168. DOI: 10.2135/cropsci2016.06.0471.

тельности и влажности растений в посевах использовали нормализованный разностный индекс растительности (NDVI) и нормализованный разностный индекс влажности (NDMI).

NDVI – это относительная величина, представляющая собой отношение разности спектральных отражающих способностей земной поверхности в ближнем инфракрасном (NIR) и красном (Red) диапазонах к их сумме:

$$NDVI = \frac{NIR - Red}{NIR + Red}.$$

NDVI является показателем радиационного типа теплообмена, однако благодаря высокой корреляции с количеством зеленой фитомассы он используется как показатель состояния растительного покрова. В то же время он является показателем общего количества и состояния фитомассы агроценоза, а не урожайности сельскохозяйственных культур. Исследования показывают, что большая фитомасса не всегда соответствует высокой урожайности.

NDMI показывает уровень содержания влаги в растениях и определяется как отношение между разностью и суммой преломленных излучений в области ближнего ИК-диапазона (NIR) и коротковолнового инфракрасного диапазона (SWIR). Он является надежным индикатором уровня водного стресса сельскохозяйственных культур и вычисляется по формуле

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR}.$$

Значение показателя содержания влаги в листьях растений является ключевым индикатором для оценки уровня увлажнения. Интерпретация абсолютного значения NDMI позволяет идентифицировать сельскохозяйственные культуры с водным стрессом. Одним из факторов, значительно влияющих на точность определения содержания влаги в растениях, является наличие отклонений, вызванных внутренней структурой листа и содержанием сухого вещества. Однако комбинация NIR и SWIR позволяет устранить эти отклонения и повысить точность опре-

деления влажности [14]. Количество влаги, находящейся во внутренней структуре листа, играет ключевую роль при контроле спектральной отражательной способности образца. Поэтому был использован NDMI для оценки уровня влажности листьев в посевах пшеницы. Значения приведенных индексов варьируют в диапазоне от –1 до 1.

Для учета атмосферного увлажнения использовали данные метеорологических наблюдений АМС «Огурцово», расположенной на расстоянии 1 км от опытного участка (<http://www.pogodaiklimat.ru/monitor.php>). Условия увлажнения лет наблюдений учитывали по трем временным периодам, соответствующим фазам развития яровой пшеницы: 1-й период – всходы – кущение (III декада мая – II декада июня), 2-й период – кущение – стеблевание (III декада июня – I декада июля) и 3-й период – стеблевание – цветение (II и III декада июля). Атмосферное увлажнение по этим периодам в годы исследований различалось. Так, в 2019 и 2020 гг. наблюдался высокий уровень увлажнения (осадков выпало больше среднеемноголетнего количества) во все периоды. В 2021 и 2022 гг. засушливым был 1-й период, относительно увлажненным – 2-й, засушливым – 3-й период вегетации культуры (см. табл. 1).

Выбор методов статистического анализа данных был обусловлен сравнительно небольшим объемом выборки, несоответствием данных закону нормального распределения и наличием как качественных, так и количественных факторов. Для подтверждения предположения о наличии связи урожайности яровой пшеницы с атмосферным увлажнением, вегетационными индексами и фоновом интенсификации и определения степени отклонения распределения вероятностей от нормального провели ряд статистических тестов: тесты Шапиро – Уилка, Лиллиефорса и Q-Q графики. Для парных сравнений данных в вариантах опыта использовали непараметрический критерий (Манна – Уитни). Для оценки связи урожайности яровой пшеницы с количественными характеристиками применяли ранговый коэффициент корреляции Спирмена, дисперсионный анализ (ANOVA)

Табл. 1. Условия атмосферного увлажнения по периодам вегетации яровой пшеницы, мм
Table 1. Conditions of atmospheric moistening by periods of spring wheat vegetation, mm

Год	Период вегетации яровой пшеницы			Сумма осадков (III декада мая – июль)
	1-й	2-й	3-й	
2019	63	36	66	165
2020	55	32	54	141
2021	32	66	4	102
2022	39	49	21	109
Среднемноголетнее количество	47	36	46	129

Примечание. Здесь и в табл. 2. 1-й период – сумма осадков за III декаду мая и I–II декады июня; 2-й период – сумма осадков за III декаду июня и I декаду июля; 3-й период – сумма осадков за II и III декады июля.

и метод главных компонент (РСА). Уровень значимости (*p*-значение) оценивали на уровне 0,05. Для визуализации структуры исходных данных и результатов анализа были построены диаграммы размаха.

В ходе исследования построена группа моделей, включающих линейную регрессию, нелинейную модель на основе регрессионного сплайна (Regression Splines), дерево решений (алгоритм CART), случайного леса (алгоритм RF), адаптивного бустинга (алгоритм AdaBoost) и градиентного бустинга (алгоритм Gradient Boosting). Исходная база данных включала 448 наблюдений, которые были разделены на обучающую и тестовую выборки в соотношении 80 и 20% соответственно.

Статистические расчеты и графические построения выполнены средствами языка программирования Python в интерактивном блокноте Jupyter, а также средствами языка программирования для статистической обработки данных и работы с графикой R в интегрированной среде разработки R-Studio.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Урожайность яровой пшеницы в опытах изменялась в зависимости от года исследования и уровня интенсификации технологий возделывания (см. рис. 1).

В среднем за годы наблюдений урожайность культуры на малоинтенсивном уровне составила 2,1 т/га, на интенсивном – 3,2 т/га. Самые низкие и самые высокие значения урожайности были зафиксированы в 2019 и

2020 гг. и составили 1,2 т/га на малоинтенсивном уровне и 4,2 т/га – на интенсивном соответственно. Наименьшая изменчивость урожайности яровой пшеницы наблюдалась на малоинтенсивном уровне в 2021 и 2022 гг., когда минимальное значение урожайности составило 1,4 т/га, а максимальное – 2,6 т/га. На интенсивном уровне наименьшая изменчивость наблюдалась в 2020 и 2022 гг., когда минимальная урожайность зерна составила 3,0 т/га, а максимальная – 3,7 т/га. Такие колебания в урожайности яровой пшеницы объясняются природными факторами, прежде всего количественного изменения атмосферного увлажнения. Например, низкая урожайность в 2019 г. обусловлена проявлением атмосферной засухи в июне, когда осадки составляли 47% от среднемноголетнего количества, и низкими значениями среднесуточной температуры воздуха – 15,4 и 16,9 °С. Уменьшение урожайности в 2021 г. вне зависимости от уровня интенсификации, возможно, было обусловлено условиями атмосферного увлажнения ранневесеннего периода.

Среднее значение вегетационных индексов за годы исследований в июне изменялось в период вегетации от 0,4 единицы (ед.) NDVI и 0,1 ед. NDMI. В июле значения NDVI увеличивались во все годы исследований из-за накопления биомассы и составляли 0,7 ед. Значения NDMI в годы исследований варьировали от 0,2 до 0,6 ед., что обусловлено изменениями в содержании влаги в растениях.

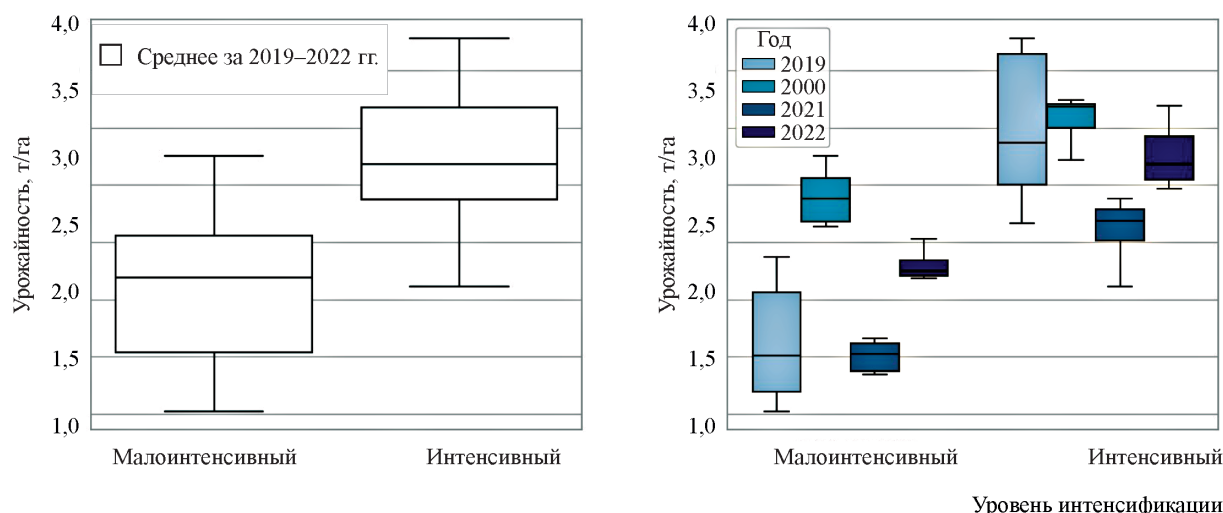


Рис. 1. Диаграмма размахов значений урожайности яровой пшеницы

На рисунке центральная линия показывает значение медианы, границы прямоугольника соответствуют значениям нижней и верхней квартилям (25 и 75 процентилем), ширина прямоугольника – размер выборочной совокупности, высота прямоугольника – интерквартильный размах (ИКР), длина верхних (нижних) «усов» диаграммы показывают границы распространения наибольшего (наименьшего) значения, находящегося в пределах $1,5 \times \text{ИКР}$.

Fig. 1. Box-plot of spring wheat yield values

In the figure, the center line shows the median value, the borders of the rectangle correspond to the values of the lower and upper quartiles (25th and 75th percentiles), the width of the rectangle is the size of the total sample, the height of the rectangle is the interquartile range (IQR), the length of the upper (lower) "whiskers" of the diagram show the boundaries of the distribution of the largest (smallest) value, which is within $1.5 \times \text{IQR}$.

Для создания моделей прогнозирования урожайности культуры необходимо учитывать различные факторы, которые оказывают наибольшее влияние на точность моделей. Определение этих факторов и выбор наиболее информативных предикторов является ключевым шагом в увеличении точности прогнозных моделей. Для этих целей применяется несколько методов. Например, метод ANOVA (анализ дисперсии), используемый для предварительной обработки данных, позволяет определить статистическую значимость различий между группами факторов. Однако помимо ANOVA существуют и другие методы, которые могут быть использованы для предварительной обработки данных. Например, метод главных компонент (PCA) позволяет снизить размерность исходного набора данных путем создания новых переменных, которые являются линейными комбинациями исходных переменных⁶. Такой подход позволяет сохранить максимальное

количество информации и в то же время сократить ее объем, упростив анализ и обработку исходных данных.

Для определения наиболее значимых факторов, влияющих на урожайность культуры в анализируемых условиях, и включения их для формирования моделей урожайности яровой пшеницы были использованы различные методы. Результаты предварительной обработки выборки данных приведены в табл. 2.

Существенный вклад, определяющий значимость воздействующих факторов на урожайность культуры, вносил уровень интенсификации технологий возделывания. При увеличении уровня интенсификации, в том числе за счет повышенных доз минеральных удобрений, установлено некоторое нивелирование влияния условий вегетации. В том числе уменьшилась до определенных пределов значимость фактора атмосферного увлажнения при интенсивном уровне агротехнологии. Так, при анализе полевых данных урожайно-

⁶Ahmad S., Pasha I., Saeed M., Shahid M. Principal component analysis and correlation studies of spring wheats in relation to cookie making quality // International Journal of Food Properties. 2017. Vol. 20. N 10. P. 2299–2313. DOI: 10.1080/10942912.2016.1236273.

Табл. 2. Перечень статистически значимых предикторов, влияющих на урожайность яровой пшеницы

Table 2. Statistically significant predictors affecting spring wheat yields

Метод	Выявленные значимые факторы
ANOVA	NDVI (июль): $p = 0,0087$ Осадки (1-й период): $p = 0,0231$ Осадки (2-й период): $p = 0,0004$ Осадки (3-й период): $p = 0,0001$
Корреляционный анализ	NDMI (июль): корреляция 0,67 NDVI (июнь): корреляция 0,60 NDVI (июль): корреляция 0,61
Метод главных компонент	PC1: Осадки (3-й период), Осадки (2-й период), Осадки (1-й период) PC2: Осадки (2-й период), Осадки (1-й период), Осадки (3-й период) PC3: Осадки (1-й период), Осадки (3-й период), NDMI (июнь) PC4: NDMI (июль), NDVI (июль), Осадки (1-й период) PC5: NDMI (июнь), NDMI (июль), NDVI (июнь)

сти культуры, полученных на интенсивном уровне возделывания, методом ANOVA ни один из факторов не показал статистически значимого влияния (p -значения $> 0,05$) на урожайность. На малоинтенсивном уровне при анализе этим же методом «осадки 2-го периода» имели статистически значимую связь (p -значение $< 0,05$). По результатам корреляционного анализа установлено, что при интенсивном уровне «осадки 1-го и 3-го периодов» имели наибольшую корреляционную связь с урожайностью яровой пшеницы (p -значение $< 0,05$). На малоинтенсивном уровне высокая корреляционная связь с урожайностью культуры установлена с «NDMI в июле» и «NDVI в июне» (p -значение $< 0,05$). Анализ данных методом главных компонент (PCA) с последующим регрессионным анализом выявил статистически значимые компоненты для каждого варианта набора данных по урожайности в зависимости от уровня интенсификации агротехнологии. На интенсивном уровне возделывания пшеницы – это компоненты PC2 и PC5 (p -значение $< 0,05$).

Необходимо отметить, что атмосферные осадки за наблюдаемые периоды развития культуры имели статистически доказанную значимость во всех проведенных непараметрических методах анализа (p -значение $< 0,05$). Установлено, что индексы NDMI и NDVI, отражающие состояние растительности, являются важными предикторами урожайности, особенно в различные временные периоды. Эти индексы позволяют более точно оценить состояние посевов и предсказать их будущую продуктивность. Выявленные наиболее значимые и информативные предикторы урожайности яровой пшеницы использованы для создания моделей машинного обучения.

Для прогнозирования урожайности яровой пшеницы применены модели: линейная регрессия, нелинейная регрессия с использованием сплайнов, дерево решений (CART), случайный лес (Random Forest), градиентный бустинг (Gradient Boosting) и адаптивный бустинг (AdaBoost). Линейная регрессия (Ridge-регрессия) была оптимизирована путем перебора коэффициентов регуляризации, что позволило контролировать степень переобучения моделей, учитывая взаимосвязи между переменными. Регрессионный сплайн был оптимизирован по степени сплайна (выбрана 2-я степень), что позволило гибко подходить к моделированию сложных нелинейных связей в данных. Определение оптимального значения степени выполнено на основе метрик, таких как MAE (средняя абсолютная ошибка), RMSE (среднеквадратическая ошибка) и R^2 (коэффициент детерминации). Использование дерева принятия решений (CART) позволило разбить данные на различные группы на основе значений предикторов, определить наиболее важные факторы, влияющие на урожайность пшеницы, и осуществить ее прогноз. Случайный лес (RF) был использован для создания ансамбля деревьев решений, что позволило учесть различные вариации в данных и получить более устойчивые и точные прогнозы. Градиентный бустинг (Gradient Boosting) и адаптивный бустинг (AdaBoost) были использованы для создания ансамблей моделей, которые

комбинируют несколько слабых моделей для получения наиболее сильного варианта.

Статистические показатели оценки точности моделей прогнозирования урожайности яровой пшеницы с применением алгоритмов машинного обучения представлены в табл. 3.

Ниже представлен график, демонстрирующий как фактические значения урожайности, так и значения, предсказанные каждой из рассмотренных моделей (см. рис. 2).

Анализ результатов моделирования и оценки их предиктивной точности показал,

что алгоритмы случайного леса (RF), градиентного (Gradient Boosting) и адаптивного бустингов (AdaBoost) имели наибольшие показатели точности прогнозов ($R^2 = 0,74–0,80$). Наиболее высокий коэффициент детерминации получен путем использования алгоритма градиентного бустинга (GB), что свидетельствует об его эффективности в прогнозировании урожайности культуры в данном исследовании.

Высокая предиктивная способность алгоритма бустинга заключается в том, что

Табл. 3. Статистические показатели точности моделей машинного обучения

Table 3. Statistical measures of accuracy of machine learning models

№	Метод	Категории предикторов	Метрика		
			MAE (средняя абсолютная ошибка)	RMSE (среднеквадратическая ошибка)	R^2 (коэффициент детерминации)
1	Линейная регрессия	Количественные	0,54	0,74	0,13
2	Регрессионный сплайн	»	0,35	0,51	0,64
3	Дерево решений	Количественные и качественные	0,33	0,48	0,67
4	Случайный лес	То же	0,29	0,44	0,74
5	Градиентный бустинг	»	0,29	0,40	0,80
6	Адаптивный бустинг	»	0,28	0,42	0,78

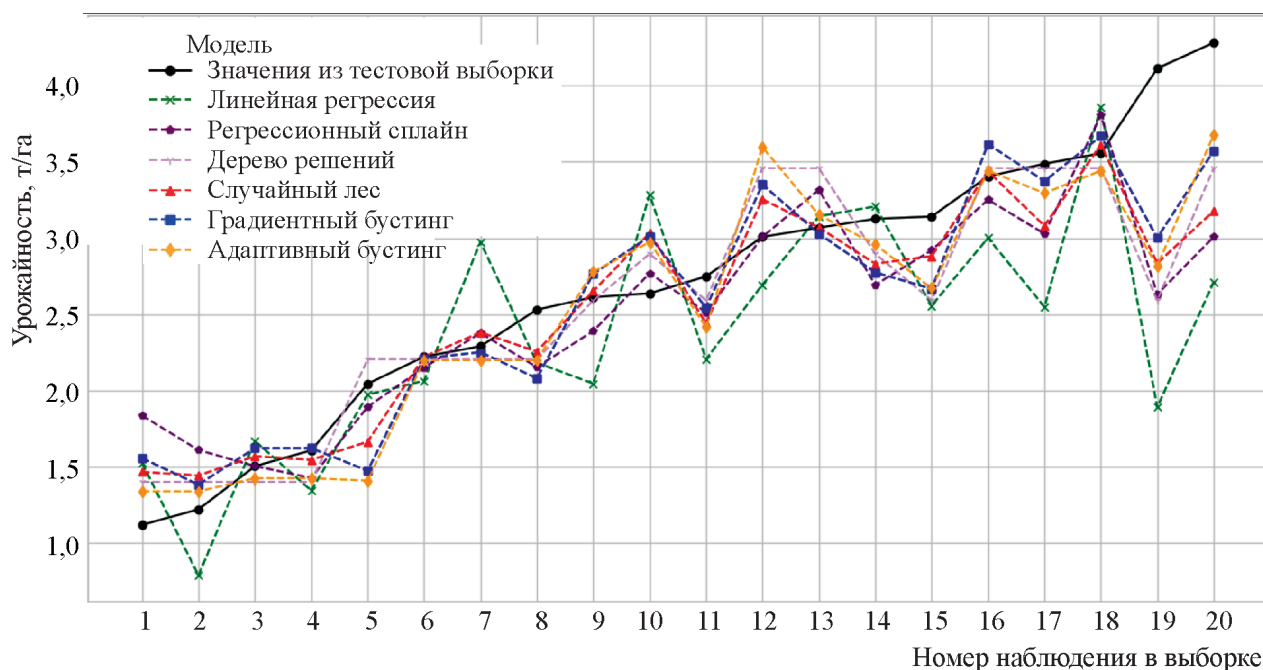


Рис. 2. Прогнозные значения урожайности яровой пшеницы и соответствующие значения из тестовой выборки

Fig. 2. Predicted values of spring wheat yield and corresponding values from the test sample

при его использовании каждая следующая модель в ансамбле строится с учетом минимизации ошибки предыдущих моделей. Ансамблевый принцип построения моделей позволяет значительно повысить точность прогнозов в условиях относительно небольшой выборки данных, подверженных высокой вариативности.

Линейная регрессионная модель имела наименьшие прогностические характеристики, что подчеркивает сложность биологических систем, отсутствие линейных связей, многоуровневую вариативность условий окружающей среды и управляющих воздействий, влияющих на урожайность яровой пшеницы.

ЗАКЛЮЧЕНИЕ

Прогнозирование урожайности яровой пшеницы является одним из важных факторов в сельскохозяйственной деятельности. Этот процесс является основой для принятия решений, направленных на оптимизацию производства и эффективное распределение ресурсов. Для осуществления прогноза урожайности культуры необходимо учитывать ряд факторов. Исходя из полученных результатов, осадки в различные временные периоды, NDMI (июль) и NDVI (июнь и июль) являются наиболее важными факторами прогнозирования урожайности яровой пшеницы.

Использование моделей машинного обучения позволяет автоматически обрабатывать большие объемы данных и находить скрытые закономерности, что выявляет неочевидные факторы, влияющие на урожайность яровой пшеницы, и улучшает точность прогнозов. Установлено, что применение методов машинного обучения является эффективным инструментом прогнозирования урожайности. Особенно результативными оказались ансамблевые модели на основе алгоритмов случайного леса, градиентного и адаптивного бустингов, которые обладали наиболее высокими прогностическими способностями ($R^2 = 0,74-0,80$).

Реализация ансамблевого подхода (комплекса предиктивных моделей) с использо-

ванием вегетационных индексов позволила значительно повысить точность прогнозирования возможной продуктивности посевов. Данный подход является перспективным для решения задач прогнозирования урожайности яровой пшеницы с учетом многоуровневой неопределенности окружающей среды и высокой вариативности результирующих показателей на конкретном земельном участке.

СПИСОК ЛИТЕРАТУРЫ

1. Berger K., Verrelst J., Féret J., Wang Z., Woche M., Strathmann M., Danner M., Mauser W., Hank T. Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions // *Remote Sensing of Environment*. 2020. Vol. 242. P. 111758. DOI: <https://doi.org/10.1016/j.rse.2020.111758>.
2. Karthikeyan L., Chawla I., Mishra A.K. A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses // *Journal of Hydrology*. 2020. Vol. 586. P. 124905. DOI: <https://doi.org/10.1016/j.jhydrol.2020.124905>.
3. Железный О.М., Тутубалина О.В., Кравцова В.И. Оценка изменений растительности Норильского промышленного района по дистанционным данным на основе анализа трендов спектральных индексов // *Современные проблемы дистанционного зондирования Земли из космоса*. 2022. Т. 19. №. 1. С. 170–178. DOI: 10.21046/2070-7401-2022-19-1-170-178.
4. Ochtyra A., Marcinkowska-Ochtyra A., Raczyko E. Threshold-and trend-based vegetation change monitoring algorithm based on the inter-annual multi-temporal normalized difference moisture index series: A case study of the Tatra Mountains // *Remote Sensing of Environment*. 2020. Vol. 249. P. 112026. DOI: 10.1016/j.rse.2020.112026.
5. Терехин Э.А. Оценка пространственно-временных изменений в зеленой фитомассе аграрной растительности с использованием спектрально-отражательных признаков // *Современные проблемы дистанционного зондирования Земли из космоса*. 2021. Т. 18. № 1. С. 138–148. DOI: 10.21046/2070-7401-2021-18-1-138-148.
6. Huang S., Tang L., Hupy J.P., Wang Y., Shao G. A commentary review on the use of normalized

- difference vegetation index (NDVI) in the era of popular remote sensing // *Journal of Forestry Research*. 2021. Vol. 32. N 1. P. 1–6. DOI: 10.1007/s11676-020-01155-1.
7. Li S., Xu L., Jing Y., Jing Y., Yin H., Li X., Guan X. High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques // *International Journal of Applied Earth Observation and Geoinformation*. 2021. Vol. 105. P. 102640. DOI: 10.1016/j.jag.2021.102640.
 8. Beyer M., Ahmad R., Yang B., Rodríguez-Bocca P. Deep Spatial-Temporal Graph Modeling for Efficient NDVI Forecasting // *Smart Agricultural Technology*. 2023. Vol. 4. P. 100172. DOI: 10.1016/j.atech.2023.100172.
 9. Olmos-Trujillo E., González-Trinidad J., Júnez-Ferreira H., Pacheco-Guerrero A., Bautista-Capetillo C., Avila-Sandoval C., Galván-Tejada E. Spatial-temporal response of vegetation indices to rainfall and temperature in a semiarid region // *Sustainability*. 2020. Vol. 12. № 5. P. 1939. DOI: 10.3390/su12051939.
 10. Li W., Migliavacca M., Forkel M., Denissen J.M.C., Reichstein M., Yang H., Duveiller G., Weber U., Orth R. Widespread increasing vegetation sensitivity to soil moisture // *Nature Communications*. 2022. Vol. 13. № 1. P. 3959. DOI: 10.1038/s41467-022-31667-9.
 11. Chen Y., Tao F. Potential of remote sensing data-crop model assimilation and seasonal weather forecasts for early-season crop yield forecasting over a large area // *Field Crops Research*. 2022. Vol. 276. P. 108398. DOI: 10.1016/j.fcr.2021.108398.
 12. Kamir E., Waldner F., Hochman Z. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2020. Vol 160. P. 124–135. DOI: 10.1016/j.fcr.2021.108398.
 13. Bazrafshan O., Ehteram M., Moshizi Z.G., Jamshidi S. Evaluation and uncertainty assessment of wheat yield prediction by multilayer perceptron model with Bayesian and copula Bayesian approaches // *Agricultural Water Management*. 2022. Vol. 273. P. 107881. DOI: 10.1016/j.agwat.2022.107881.
 14. Ibrahim G.R.F., Rasul A., Abdullah H. Assessing how irrigation practices and soil moisture affect crop growth through monitoring Sentinel-1 and Sentinel-2 data // *Environmental monitoring and assessment*. 2023. Vol. 195. N 11. P. 1262. DOI: 10.1007/s10661-023-11871-w.

REFERENCES

1. Berger K., Verrelst J., Féret J., Wang Z., Woche M., Strathmann M., Danner M., Mauser W., Hank T. Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment*, 2020, vol. 242. p. 111758. DOI: 10.1016/j.rse.2020.111758.
2. Karthikeyan L., Chawla I., Mishra A.K. A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *Journal of Hydrology*, 2020, vol. 586, p. 124905. DOI: 10.1016/j.jhydrol.2020.124905.
3. Zhelezny O.M., Tutubalina O.V., Kravtsova V.I. Assessment of changes in vegetation of the Norilsk industrial region with remote sensing data based on the trend analysis of spectral indices. *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa = Current problems in remote sensing of the Earth from space*, 2022, vol. 19, no. 1, pp. 170–178. (In Russian). DOI: 10.21046/2070-7401-2022-19-1-170-178.
4. Ochtyra A., Marcinkowska-Ochtyra A., Raczkó E. Threshold-and trend-based vegetation change monitoring algorithm based on the inter-annual multi-temporal normalized difference moisture index series: A case study of the Tatra Mountains. *Remote Sensing of Environment*, 2020, vol. 249, p. 112026. DOI: 10.1016/j.rse.2020.112026.
5. Terekhin E.A. Assessment of spatial-temporal changes in green phytomass of agricultural vegetation using spectral response. *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa = Current problems in remote sensing of the Earth from space*, 2021, vol. 18, no. 1, pp. 138–148. (In Russian). DOI: 10.21046/2070-7401-2021-18-1-138-148.
6. Huang S., Tang L., Hupy J.P., Wang Y., Shao G. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, 2021, vol. 32, no. 1, pp. 1–6. DOI: 10.1007/s11676-020-01155-1.
7. Li S., Xu L., Jing Y., Jing Y., Yin H., Li X., Guan X. High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques. *International Journal*

- of *Applied Earth Observation and Geoinformation*, 2021, vol. 105, p. 102640. DOI: <https://doi.org/10.1016/j.jag.2021.102640>.
8. Beyer M., Ahmad R., Yang B., Rodríguez-Bocca P. Deep Spatial-Temporal Graph Modeling for Efficient NDVI Forecasting. *Smart Agricultural Technology*. 2023, vol. 4, p. 100172. DOI: 10.1016/j.atech.2023.100172.
 9. Olmos-Trujillo E., González-Trinidad J., Júnez-Ferreira H., Pacheco-Guerrero A., Bautista-Capetillo C., Avila-Sandoval C., Galván-Tejada E. Spatial-temporal response of vegetation indices to rainfall and temperature in a semiarid region. *Sustainability*, 2020, vol. 12, no. 5, p. 1939. DOI: 10.3390/su12051939.
 10. Li W., Migliavacca M., Forkel M., Denissen J.M.C., Reichstein M., Yang H., Duveiller G., Weber U., Orth R. Widespread increasing vegetation sensitivity to soil moisture. *Nature Communications*, 2022, vol. 13, no. 1, p. 3959. DOI: 10.1038/s41467-022-31667-9.
 11. Chen Y., Tao F. Potential of remote sensing data-crop model assimilation and seasonal weather forecasts for early-season crop yield forecasting over a large area. *Field Crops Research*, 2022, vol. 276, p. 108398. DOI: 10.1016/j.fcr.2021.108398.
 12. Kamir E., Waldner F., Hochman Z. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, vol. 160, pp. 124–135. DOI: 10.1016/j.fcr.2021.108398.
 13. Bazrafshan O., Ehteram M., Moshizi Z.G., Jamshidi S. Evaluation and uncertainty assessment of wheat yield prediction by multilayer perceptron model with Bayesian and copula Bayesian approaches. *Agricultural Water Management*, 2022, vol. 273, p. 107881. DOI: 10.1016/j.agwat.2022.107881.
 14. Ibrahim G.R.F., Rasul A., Abdullah H. Assessing how irrigation practices and soil moisture affect crop growth through monitoring Sentinel-1 and Sentinel-2 data. *Environmental monitoring and assessment*, 2023, vol. 195, no. 11, p. 1262. DOI: 10.1007/s10661-023-11871-w.

ИНФОРМАЦИЯ ОБ АВТОРАХ

✉ **Гарафутдинова Л.В.**, аспирант, младший научный сотрудник; **адрес для переписки:** Россия, 630501, Новосибирская область, р.п. Краснообск, а/я 463; e-mail: garafutdinovalv@sfsc.ru

Федоров Д.С., инженер-программист

Каличкин В.К., доктор сельскохозяйственных наук, профессор, главный научный сотрудник

Максимович К.Ю., научный сотрудник

Колбин С.А., кандидат сельскохозяйственных наук, ведущий научный сотрудник

AUTHOR INFORMATION

✉ **Lyudmila V. Garafutdinova**, Post-graduate Student, Junior Researcher; **address:** PO Box 463, Krasnoobsk, Novosibirsk Region, 630501, Russia; e-mail: garafutdinovalv@sfsc.ru

Dmitry S. Fedorov, Software Engineer

Vladimir K. Kalichkin, Doctor of Science in Agriculture, Professor, Senior Researcher

Kirill Yu. Maximovich, Researcher

Sergey A. Kolbin, Candidate of Science in Agriculture, Lead Researcher

Дата поступления статьи / Received by the editors 02.11.2023
Дата принятия к публикации / Accepted for publication 27.12.2023
Дата публикации / Published 20.02.2024