

EDN: LDMOSW

DOI: 10.22314/2073-7599-2025-19-2-33-44



Научная статья

УДК 631.17



Прогнозирование урожайности сельскохозяйственных культур: структура данных и методы искусственного интеллекта

Владимир Климентьевич Каличкин¹,

доктор сельскохозяйственных наук,

главный научный сотрудник,

e-mail: vk.kalichkin@gmail.com;

Кирилл Юрьевич Максимович¹,

кандидат биологических наук, научный сотрудник,

e-mail: kiri-maksimovi@mail.ru;

Ольга Александровна Алещенко²,

младший научный сотрудник,

e-mail: oelab2@mail.ru;

Виталий Викторович Алещенко³,

доктор экономических наук, доцент,

e-mail: 564435@mail.ru

¹Сибирский федеральный научный центр агробиотехнологий РАН, Новосибирская область, р.п. Краснообск, Российская Федерация;

²Институт экономики и организации промышленного производства СО РАН, Новосибирск, Российская Федерация;

³Новосибирский государственный аграрный университет, Новосибирск, Российская Федерация

Статья подготовлена по плану НИР ИЭОПП СО РАН, проект «Движущие силы и механизмы развития кооперационных и интеграционных процессов в экономике Сибири», № 121040100279-5.

Реферат. Интеллектуальное (умное) земледелие является современным этапом развития сельскохозяйственной науки и практики. Его характерная особенность заключается в активном применении методов искусственного интеллекта, в частности, машинного и глубокого обучения, при решении частных задач, направленных на устойчивое производство в растениеводстве. (*Цель исследования*) Целью данного исследования стал анализ структуры данных и сравнение алгоритмов машинного и глубокого обучения, используемых в прогнозировании урожайности сельскохозяйственных культур. (*Материалы и методы*) На основе конвергентного подхода с использованием методов когнитивного и семантического анализа авторами рассмотрена предметная область «Применение методов искусственного интеллекта при прогнозировании урожайности сельскохозяйственных культур», а также базовые аспекты, связанные со структурой исходных данных, основные этапы реализации предиктивных моделей и наиболее используемые методы машинного и глубокого обучения. (*Результаты и обсуждение*) По результатам работы представлены основная структура и способы получения данных, а также типовая схема реализации моделей в предиктивной аналитике урожайности сельскохозяйственных культур. Выделены наиболее распространенные методы машинного и глубокого обучения, подробно рассмотрены их функциональные особенности. На основе сравнительного анализа показано, что глубокое обучение и гибридные подходы превосходят традиционные методы машинного обучения по метрикам ошибок (точности прогнозирования). (*Выводы*) По результатам исследований установлено преимущество методов глубокого обучения ($R_{\text{ср}}^2 = 0,85$) и гибридного подхода ($R_{\text{ср}}^2 = 0,87$) в прогнозировании урожайности сельскохозяйственных культур в зависимости от складывающихся условий и управляющего воздействия. Развитием дальнейшей исследовательской работы может быть адаптация современных подходов искусственного интеллекта к пространственным объектам землепользования и культурам с преимущественным использованием данных дистанционного зондирования.

Ключевые слова: прогнозирование урожайности, интеллектуальное земледелие, искусственный интеллект, структура данных, машинное обучение, глубокое обучение.

■ **Для цитирования:** Каличкин В.К., Максимович К.Ю., Алещенко О.А., Алещенко В.В. Прогнозирование урожайности сельскохозяйственных культур: структура данных и методы искусственного интеллекта // *Сельскохозяйственные машины и технологии*. 2025. Т. 19. №2. С. 33-44. DOI: 10.22314/2073-7599-2025-19-2-33-44. EDN: LDMOSW.

Scientific article

Crop Yield Prediction: Data Structure and Ai-Powered Methods

Vladimir K. Kalichkin¹,

Dr.Sc.(Eng.), chief researcher,

e-mail: vk.kalichkin@gmail.com;

Kirill Yu. Maksimovich¹,

Ph.D.(Eng.), researcher,

e-mail: kiri-maksimovi@mail.ru;

Olga A. Aleshchenko²,
junior researcher,
e-mail: oelab2@mail.ru;

Vitaliy V. Aleshchenko³,
Dr.Sc.(Eng.), associate professor,
e-mail: 564435@mail.ru

¹Siberian Federal Scientific Center of Agro-Bio Technology of the Russian Academy of Sciences, Krasnoobsk, Novosibirsk Region, Russian Federation;

²Institute of Economics and Industrial Engineering, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation;

³Novosibirsk State Agrarian University, Novosibirsk, Russian Federation

The article was prepared according to the research plan of the IEIE SB RAS, project «Driving forces and mechanisms for the development of cooperative and integration processes in the Siberian economy», No. 121040100279-5.

Abstract. Smart farming, also known as intelligent agriculture, represents a modern stage in the development of agricultural science and practice. Its defining feature lies in the active application of artificial intelligence methods, particularly machine learning and deep learning, to address specific tasks aimed at ensuring sustainable crop production. (*Research purpose*) The aim of this study is to analyze data structures and compare machine learning and deep learning algorithms used in crop yield prediction. (*Materials and methods*) Using a convergent approach and applying methods of cognitive and semantic analysis, the authors examined the subject area of artificial intelligence applications in crop yield prediction. The study also explores key aspects related to the structure of input data, the main stages of implementing predictive models, and the most widely used machine learning and deep learning methods. (*Results and discussion*) The study presents the core data structure and methods for data acquisition, along with a typical workflow for implementing predictive analytics models for crop yield prediction. The most commonly used machine learning and deep learning methods are identified and their functional characteristics are examined in detail. Comparative analysis demonstrates that deep learning and hybrid approaches outperform traditional machine learning methods in terms of prediction accuracy, as measured by standard error metrics. (*Conclusions*) The findings confirm the advantages of deep learning methods (mean $R^2 = 0.85$) and hybrid approaches (mean $R^2 = 0.87$) in crop yield prediction under varying conditions and management interventions. Future research may focus on adapting modern AI approaches to spatial land use objects and crop types, with an emphasis on remote sensing data.

Keywords: crop yield prediction, smart farming, artificial intelligence, data structure, machine learning, deep learning.

■ **For citation:** Kalichkin V.K., Maksimovich K.Yu., Aleshchenko O.A., Aleshchenko V.V. Crop yield prediction: data structure and Ai-powered methods. *Agricultural Machinery and Technologies*. 2025. Vol. 19. N2. 33-44 (In Russian). DOI: 10.22314/2073-7599-2025-19-2-33-44. EDN: LDMOSW.

Эффективные рекомендации по возделыванию сельскохозяйственных культур необходимы с целью разработки и реализации плана развития растениеводства для повышения доходов товаропроизводителей и проведения импортно-экспортной политики. Методы прогнозирования урожая используются для того, чтобы предложить виды культур для выращивания на конкретном поле и агротехнические приемы управления продуктивностью посевов. Раннее и надежное прогнозирование урожайности сельскохозяйственных культур в масштабе поля может также служить целям страховой оценки.

Точное прогнозирование урожайности сельскохозяйственных культур зависит от таких факторов, как географическое положение, погодные условия, применяемые управляющие воздействия, а также от качества и количества исходных данных и выбранного подхода, используемого в предиктивной аналитике. С помощью анализа данных, машинного обучения и агрономических знаний создается

система рекомендаций по выращиванию культур – инструмент поддержки принятия решений для агрономов. Он помогает выбирать наилучшие культуры для выращивания, исходя из условий рельефа, типа почвы, вида агрокультуры, климата, доступных ресурсов и рыночного спроса, для реализации агротехнологий. В предыдущие годы урожайность сельскохозяйственных культур прогнозировалась с помощью статистических и механистических моделей. Модели, основанные на искусственном интеллекте, такие как модели машинного обучения и глубокого обучения, могут стать потенциальной заменой статистическому и механистическому моделированию, поскольку они дают возможность масштабировать результаты прогнозной аналитики и гораздо эффективнее обрабатывать сложные и нелинейные закономерности [1-3].

В России неплохо развита система прогнозирования урожайности 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 метеоусловиями. Затем формируются регрессионные модели прогнозирования на основе совместного использования наземных и спутниковых данных [4, 5]. В работе [6] по данным среднемесячной температуры воздуха, суммарного количества осадков и урожайности построена адаптивная нечетко-логическая модель анализа и прогноза урожайности озимой пшеницы. Разрабатываются также математические модели прогнозирования урожайности культур, основанные на агрохимических показателях почвы (гумус, обменный калий К, доступный фосфор Р и рН солевой вытяжки) [7]. Осуществляются попытки создания методики прогнозирования урожайности на основе использования эконометрических инструментов. Например, показано, что на урожайность сельскохозяйственных культур оказывают влияние социально-экономические и технологические факторы, а также высокая доля освоенных под посевы территорий [8].

В мировой практике при прогнозировании урожайности сельскохозяйственных культур широко применяются методы искусственного интеллекта, в частности, машинное обучение [9]. В отечественной предиктивной аналитике в этой предметной области исследования по применению машинного обучения имеют ограниченное развитие. В качестве примера можно привести исследование по применению 8 алгоритмов машинного обучения [10]. Установлено, что *Random Forest*, *Gradient Boosting* и *Neural Networks* были более эффективными в задачах прогнозирования урожайности по сравнению с алгоритмами *K-Nearest Neighbors*, *Decision Trees*, *Support Vector Machines*. Имеются также работы с использованием моделей глубокого обучения, в частности, искусственная нейронная сеть (ИНС) с алгоритмом обратного распространения ошибки [11]; ИНС в форме многослойного персептрона (*MLP*) [12, 13]; *MLP*, радиальные базисные функции (*RBF*) и обобщенно-регрессионная нейронная сеть (*GRNN*) [14].

В то же время в эпоху цифровых технологий в сельском хозяйстве применение машинного обучения открывает широкие возможности получения более точных прогнозов урожайности культур в условиях растущего количества данных. Выбор метода или подхода машинного обучения при прогнозировании зависит от многих сопутствующих условий. Прежде всего, это доступность и структура данных, методов и источников их получения, профессиональная подготовка исследователя, доступность программного обеспечения и др. В связи с этим целесообразно кратко рассмотреть мировые тенденции в отношении научного обеспечения проблемы.

Цель исследования – анализ структуры данных и сравнение алгоритмов машинного и глубокого обучения, используемых в прогнозировании урожайности сельскохозяйственных культур.

Материалы и методы. Общеметодологической основой исследований служил метод обобщения – способ познания посредством определения общих существенных признаков определенной предметной области (ПрО). В нашем случае ПрО формулировалась как «применение машинного обучения при прогнозировании урожайности сельскохозяйственных культур». Обобщение базировалось на анализе и синтезе научных публикаций отечественных и зарубежных авторов, направленных на установление и сравнение существенных признаков ПрО. Также использован конвергентный подход с применением методов когнитивного и семантического анализов. В этом исследовании использовались базы данных *Web of Science*, *Scopus*, *Science Direct*, *Springer Link*, *Google Scholar*, *elibrary.ru*, *cyberleninka.ru* и др. Применялось индуктивное обобщение (от единичного достоверного к общему вероятностному), которое предполагает определение общих существенных признаков ПрО и фиксирование их в форме суждений.

Также выполнено сравнение эффективности различных методов машинного обучения, глубокого обучения и гибридных моделей на основе коэффициента детерминации (R^2). График оценки точности моделей (рис. 4) выполнен в формате «ящик с усами» для визуализации распределения значений R^2 . «Ящик» показывает межквартильный размах (от 25- до 75-го перцентиля), где горизонтальная линия внутри «ящика» представляет медиану. «Усы» указывают на минимальные и максимальные значения, исключая выбросы. Отдельные точки за пределами «усов» представляют статистические выбросы.

Для обобщения точности прогнозных моделей машинного обучения использованы вариационно-статистические методы средствами языка программирования для статистической обработки данных в интегрированной среде разработки *R-Studio*. Дополнительные графические построения выполнялись в программном пакете *Draw.IO* (*app.diagrams.net*).

Результаты и обсуждение. Участие в формировании урожая взаимосвязанных факторов и условий окружающей среды, чаще всего недоопределенных, существование возмущений в реальных системах с сильно нелинейными явлениями делает задачу прогнозирования урожайности сельскохозяйственных культур очень сложной и может вызывать ошибки моделирования. Погодные условия влияют на рост растений на различных стадиях, приводя к значительным годовым и межгодовым

колебаниям урожайности. Кроме того, пространственная изменчивость свойств почвы, а также чередование культур (севообороты), системы обработки почвы и удобрений, развитие вредных организмов и применяемые агротехнологии увеличивают сложность прогнозирования урожайности.

Разработка точного и эффективного метода прогноза требует правильной оценки этих условий и факторов, что возможно путем их мониторинга и проведения экспериментов по выращиванию сельскохозяйственных культур. То есть, анализ длительных временных рядов полевых опытов научных и учебных учреждений дает наиболее правдоподобный прогноз, но который не может быть реализован без дополнительного формирования параметров признаков географических объектов (например, видов земель различной территориальной принадлежности) для использования в прогнозном моделировании в пространственном масштабе. Для внутрихозяйственного использования методов прогнозирования урожайности большее значение имеют также данные по ее историческим записям на каждом поле.

В мировой предиктивной аналитике существует два основных подхода для прогнозирования урожайности сельскохозяйственных культур (рис. 1). В первом подходе используются математические модели, называемые моделями процессов или моделями роста сельскохозяйственных культур (механистические модели). Эти модели могут использоваться для представления разнообразных взаимодействий физиологических процессов растений с окружающей средой для оценки потенциала производства биомассы и обеспечивают абстрактное представление о реализации динамического поведения физиологических стадий развития растений.



Рис. 1. Подходы к прогнозированию урожайности сельскохозяйственных культур

Fig. 1. Approaches to crop yield prediction

Фактические данные и различные предположения о типах почв, солнечной радиации, изменении осадков и температуры, различных управляющих воздействий, служат исходным материалом для моделей роста растений. В основном механистические модели ориентированы на конкретные культуры. В качестве примера можно привести APSIM, DSSAT, STICS, WOFOST, CropSyst, AGROTOOL и другие модели. Признается, что эти модели спо-

собны к высокоточному прогнозированию урожайности, но требуют обширных входных данных. Несмотря на эффективность, такого рода модели обычно оказываются слишком дорогостоящими и трудозатратными, поэтому непрактичны для массовых приложений в растениеводстве.

Во втором подходе используются модели, управляемые данными. Этот подход, иногда называемый эмпирическим, сравнительно более прост в использовании, чем модели роста сельскохозяйственных культур. При этом подходе рассматриваются данные об урожайности культур за несколько лет, полученных по историческим данным хозяйства или полевым опытам, либо с помощью дистанционного зондирования. На основе этих данных определяется набор признаков, наиболее способствующих изменению урожайности. Принимая эти эффективные признаки как независимые, а урожайность как зависимую переменную, применяются различные варианты моделирования прогноза. Такие методы относительно менее затратные и простые в применении, а также для них не требуется никакой предварительной информации о различных физиологических процессах, участвующих в росте растения.

Этот подход также делится на два направления: статистическое моделирование и машинное обучение (МО). Наиболее распространенными были статистические модели, основанные на уравнениях регрессии между урожайностью и агрометеорологическими параметрами, такими как температура, количество осадков, влажность почвы и др. В свое время они наиболее широко использовались для формулирования соответствующих рекомендаций. Каждая статистическая модель определяет один набор параметров.

Однако были и модели, основанные на других подходах. Примерами могут служить модели: на основе составных погодных индексов; дискриминантных функций; метода водного баланса; индексов роста; вероятности модели цепи Маркова; логистических моделей; на основе детерминированной урожайности с использованием ARIMA и др. Однако сложность с этими моделями прогнозирования урожайности сельскохозяйственных культур заключается в требовании большого объема полевых данных для их параметризации и калибровки.

Машинное обучение рассматривается как подмножество искусственного интеллекта (ИИ). Оно создает самообучающуюся математическую модель на базе данных, известных как «обучающие данные», для прогнозирования. Алгоритмы МО, в отличие от традиционных статистических моделей, интерпретируют выходную переменную как неявную функцию входных параметров, которые могут быть сложными и выполняют решение зада-

чи без явного программирования. При этом всегда возникает необходимость более точного и надежного выделения пространственно-временных признаков, которые обычно извлекаются при участии человека (в моделях глубокого обучения – ГО участие человека ограничивается выбором архитектуры нейронной сети и подготовкой данных).

Однако, в отличие от традиционных статистических методов, МО не делает предположений о правильной структуре модели данных, такой как функциональная форма и распределение вероятностей. Вместо этого методы МО позволяют изучать взаимосвязь между зависимыми и независимыми переменными на основе данных и базируются на полупараметрических и непараметрических структурах. В контексте прогнозирования урожайности сельскохозяйственных культур с несколькими исходными данными применение таких методов позволило создать модели прогноза более высокой точности.

В свою очередь, эти модели делятся на модели МО и ГО.

В моделях МО обычно используются алгоритмы: дерево решений (*Decision Trees – DT*), метод опорных векторов (*Support Vector Machines – SVM*), цепи Маркова (*Markov Chain Correlation – MCC*), к-ближайший сосед (*K-Nearest Neighbors – KNN*), случайный лес (*Random Forest – RF*), наивный байесовский классификатор (*Naïve Bayes classifier – NB*), градиентный бустинг (*Gradient Boosting – GB*), адаптивный бустинг (*Adaptive Boosting – AdaBoost*), экстремальный градиентный бустинг (*eXtreme Gradient Boosting – XGBoost*), искусственная нейронная сеть (*Artificial Neural Network – ANN*), классические регрессионные модели (*Regression Models – RM*) и др.

В моделях ГО используются: сверточная нейронная сеть (*Convolutional Neural Network – CNN*), рекуррентная нейронная сеть (*Recurrent Neural Network – RNN*), глубокая нейронная сеть (*Deep Neural Networks – DNN*), спайковая нейронная сеть (*Spiking Neural Network – SNN*), долговременная и кратковременная память (*Long-Short Term Memory – LSTM*) и др.

Сбор данных имеет решающее значение как по качеству, так и по размеру для получения точных результатов при моделировании прогнозов. Выбор подходящих признаков и их переменных является очень важной частью любых алгоритмов МО. Чаще всего в моделях как МО, так и ГО используются признаки, отражающие агрометеорологические ресурсы местности в течение вегетационного периода, характеристики почвы, состояние посевов по стадиям роста растений и применяемые управляющие воздействия на агрофитоценоз (рис. 2).

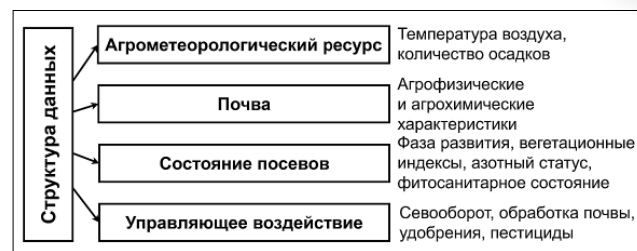


Рис. 2. Основные признаки, используемые в машинном обучении при прогнозировании урожайности сельскохозяйственных культур

Fig. 2. Main features used in machine learning for crop yield prediction

Группа признаков «агрометеорологический ресурс» включает информацию о погоде, которая собирается с помощью полевых метеостанций, и данные о температуре воздуха и почвы, влажности воздуха и почвы, скорости и направлении ветра, количестве осадков и о сумме активных температур. Данные о погоде могут быть использованы также из различных источников, например, цифровых глобальных баз данных по агрометеорологическим параметрам. В качестве примеров можно привести наиболее распространенные базы с открытым доступом: ERA Interim (<https://www.ecmwf.int>), NCEP/NCAR (<http://www.esrl.noaa.gov>) – открытые мировые базы данных реанализов; World Clim (<https://www.worldclim.org>) – мировая база данных климата; Climate Data Online (<https://www.ncei.noaa.gov>) – архив глобальных исторических данных о погоде и климате; Погода и климат (<http://www.pogodaiklimat.ru>) – отечественная база данных климата, содержащая многолетний архив погоды по всему миру.

Группа признаков «почва» состоит из переменных: почвенные карты, тип почвы, значение pH, питательные вещества (азот, фосфор, калий, магний, сера, цинк, бор, кальций, марганец), влажность почвы, структурное состояние, залегание в рельефе, местоположение в качестве географического объекта и др. Информацию о почве получают путем оцифровки почвенных карт на бумажных носителях, по материалам дистанционного зондирования спутниками и беспилотными летательными аппаратами (БПЛА), а также с помощью нативных датчиков. Валидация этих переменных осуществляется обычно путем полевого обследования и проведения агрохимических и агрофизических лабораторных анализов.

Группа признаков «состояние посевов» относится к информации о культуре, такой как биомасса, прирост в процессе выращивания, густота травостоя, проективное покрытие и др., которые определяются с помощью полевых обследований и вегетационных индексов. Обычно используется NDVI, расчет которого базируется на двух наиболее стабильных (не зависящих от прочих факторов) участ-

ках спектральной кривой отражения растительностью. В красной области спектра (0,6-0,7 мкм) лежит максимум поглощения солнечной радиации хлорофиллом растений, а в инфракрасной области (0,7-1,0 мкм) находится область максимального отражения клеточных структур листа. Отношение этих показателей друг к другу позволяет четко отделять растительность от прочих природных объектов и анализировать ее состояние. В эту группу также включаются другие показатели, указывающие на рост, например, индекс площади листьев, азотный статус посевов и развитие вредителей и болезней.

В группе признаков «управляющие воздействия» могут быть предыдущие данные об урожайности (временные ряды), годе, площади возделывания. Также используются данные фитосанитарного и агрохимического обследования, чередования культур во времени на поле, применения удобрений и средств защиты растений (пестицидов), приемов обработки почвы, сроков посева и других агротехнических показателей.

Использование в анализе более одного набора данных из разных источников как правило обеспечивает более высокий процент успеха прогнозирования.

Типовая схема реализации моделей машинного обучения при прогнозировании урожайности сельскохозяйственных культур состоит из четырех основных этапов (рис. 3).

Первый этап – сбор данных описан выше.

Второй этап – предварительная обработка данных существенно влияет на эффективность работы моделей МО. Этот этап включает серию операций, направленных на устранение шума и генерацию высококачественных данных, подходящих для обучения модели. Ключевые цели этапа включают: достижение баланса данных для уменьшения смещения, выполнение преобразований данных и устранение нерелевантных данных. Этап предварительной обработки данных связан также с выбором соответствующих признаков для использования в последующем анализе.

Значимые признаки и их переменные выбираются из обучающих данных с использованием различных статистических критериев, например, оценки прироста информации, коэффициента прироста информации, снижения коэффициента Джини, дисперсионного анализа, хи-квадрат, алгоритма *ReliefF* (алгоритм выбора признаков ранговой важности) и *FCBF* (быстрый фильтр на основе корреляции для выбора признаков). Ведутся также поиски новых методов по выбору релевантных признаков, особенно в моделях ГО, где количество признаков и переменных может насчитываться десятками при анализе больших данных. Например, описан новый подход оптимального выбора признаков, в котором извлекаются функции более высокого порядка,



Рис. 3. Типовая схема реализации моделей машинного обучения при прогнозировании урожайности сельскохозяйственных культур

Fig. 3. Typical architecture for implementing machine learning models in crop yield prediction

улучшенные энтропийные функции и функции, основанные на корреляции в работе [15]. Затем выбираются оптимальные функции из этого набора с помощью алгоритма *IBS-BOA* (*IBS* – эволюционный алгоритм с использованием роевого интеллекта стаи птиц, а *BOA* – алгоритм оптимизации «бабочка»). *CNN* обучается с помощью алгоритма *IBS-BOA*. Попарные положительные корреляции между различными признаками помогают удалять высокоррелированные признаки, поскольку они снижают прогностический потенциал модели. Для уменьшения размерности данных может быть использован анализ главных компонент (*principal component analysis, PCA*).

Управление значениями осуществляется путем исключения строки или выбора определенного постоянного значения для заполнения строки, или путем добавления известного среднего значения, которое определяется на основе выбранного признака. В некоторых случаях категориальные значения, которые представлены в виде строковых констант, преобразуются в целочисленные значения, чтобы модель могла их легко интерпретировать. Разделение данных на обучение и тестирование осуществляется в стандартном соотношении 80:20, где 80% данных учитывается для обучения модели и 20% для тестирования (могут быть варианты 70:30%).

Третий этап состоит из выбора алгоритма МО, обучения и тестирования модели с использованием обучающих и тестовых данных.

Четвертый этап – оценка точности модели.

Для оценки точности моделей используют набор метрик, охватывающий различные аспекты производительности моделей. Среднеквадратичная ошибка ($RMSE$) и средняя абсолютная ошибка (MSE) позволяют оценить общую точность прогнозов, при этом $RMSE$ выражается в тех же единицах, что и целевая переменная. Средняя абсолютная ошибка (MAE) и среднее абсолютное отклонение (MAD) предоставляют информацию о средней величине ошибок прогноза.

Для оценки относительной точности прогноза используют среднюю абсолютную ошибку в процентах ($MAPE$), среднюю относительную ошибку в процентах ($MRPE$), среднюю процентную ошибку (MPE), а также специализированные метрики $AMAPE$ и $SMAPE$, отражающие отклонения средних и суммарных значений, соответственно. Нормализованная среднеквадратичная ошибка ($NRMSE$) и среднеквадратичная логарифмическая ошибка ($RMSLE$) дополнительно учитывают масштаб и распределение данных. Для оценки монотонности связи между прогнозируемыми и фактическими значениями применяют коэффициенты ранговой корреляции Кендалла ($Kendall\ Tau$) и Спирмена ($Spearman\ Corr$). Коэффициент детерминации (R^2) используют для оценки доли дисперсии зависимой переменной, объясняемой моделью.

Большинство алгоритмов машинного обучения решают задачи как классификации, так и регрессии. При использовании метода опорных векторов (SVM) обычно применяются исторические данные, где выбранные характеристики служат входными данными, а целевой переменной выступает урожайность. Целью SVM является установление оптимальной границы принятия решения, которая эффективно различает уровни урожайности.

Цепи Маркова (*Markov chain correlation – MCC*) – это математическая модель, представляющая собой последовательность расчетов, где вероятность наступления каждого события зависит только от состояния, достигнутого в предыдущем событии. Характеризуется тем, что, при текущем состоянии системы, ее будущее состояние не зависит от прошлого, а опирается на происходящее сейчас.

Алгоритм К-ближайшего соседа (KNN) представляет собой простой метод контролируемого машинного обучения. Прогнозирование урожайности сельскохозяйственных культур, неизвестного значения признака, может быть выполнено путем использования значений ближайших известных соседей. Это достигается путем вычисления евклидова расстояния между точками данных.

Алгоритм дерево решений (DT) также представляет собой метод контролируемого машинного обучения. Структура этого классификатора напоминает дерево, где каждый конечный узел пред-

ставляет результат, ветви демонстрируют правила принятия решений, а основные узлы отражают свойства набора данных.

На этапе обучения алгоритм DT определяет оптимальный признак для сегментации данных, используя метрику, такую как энтропия, или индекс Джини (примесь). Эти метрики измеряют уровень беспорядка или разнообразия в подмножествах данных. Цель состоит в том, чтобы точно определить особенности, которые дают максимальный прирост информации или минимальные примеси после разделения. Значения корневых признаков сравниваются с атрибутами записи. На основе этой оценки выполняется соответствующая ветвь, связанная с этими данными, направляя переход к следующему узлу.

Работа алгоритма случайный лес (RF) основана на создании деревьев решений и заданных данных. Каждое дерево распространяет прогнозы, и алгоритм принимает оптимальное решение в процессе голосования. Поскольку результаты здесь усредняются перед принятием решения, это всегда эффективнее, чем вывод отдельного дерева решений.

Наивный байесовский классификатор (NB) вычисляет вероятность того, что событие произойдет, учитывая истинность другого события. Мультиномиальные алгоритмы, а также Гаусса и Бернулли составляют три наивных метода Байеса.

Искусственная нейронная сеть (ANN) – это вычислительную модель с N количеством узлов (или нейронов), соединенных друг с другом. Каждый узел представляет собой определенную выходную функцию, называемую функцией активации. Связь между каждыми двумя узлами представляет собой вес для сигнала, проходящего через связь, который эквивалентен памяти ANN . Выход сети будет варьироваться в зависимости от того, как связана сеть, значения веса и функции стимулирования. На обучающем множестве происходит обучение нейронной сети. На тестовом и контрольных множествах осуществляется проверка построенной модели. При обучении и тестировании синаптические веса и смещения выбираются таким образом, чтобы минимизировать ошибку обучения.

Ансамблевые алгоритмы МО, такие как градиентный бустинг (GB), адаптивный бустинг ($AdaBoost$) и экстремальный градиентный бустинг ($XGBoost$) используются с целью объединения прогнозов нескольких моделей МО для повышения точности и надежности.

В градиентном бустинге каждый последующий метод обучается уменьшать функцию потерь. Метод вычисляет градиент процедуры потерь по отношению к расчетам, выполненным текущим ансамблем для каждого процесса. Впоследствии он готовит новую слабую модель для уменьшения это-

го градиента. Вычисления, произведенные с помощью этой методики, включаются в совокупность, и эта итерационная процедура продолжается до достижения заранее определенного конечного условия.

Адаптивный бустинг принадлежит к группе ускоряющих классификаторов. Его основная цель – повысить точность классификатора за счет интеграции нескольких слабых классификаторов. Фундаментальный принцип *AdaBoost* заключается в корректировке весов классификаторов и итеративном обучении выборок данных.

Экстремальный градиентный бустинг – это масштабируемая и точная реализация градиентного бустинга, специально разработанная для использования потенциала вычислительных мощностей для расширенных древовидных алгоритмов. Алгоритм революционизирует построение деревьев, используя параллельную обработку вместо последовательных методов. Используя стратегию по уровням, он эффективно оценивает качество потенциальных различий в обучающей выборке путем сканирования значений градиента и использования частичных сумм.

Модели глубокого обучения представляют собой значительный прогресс в подотрасли МО. Наиболее существенными аспектами методов ГО является автоматическое извлечение признаков и высокая производительность. Они представляют собой производную от архитектур искусственных нейронных сетей, отличающуюся своими иерархическими структурами обучения с множеством взаимосвязанных уровней. Каждый уровень в этой структуре отвечает за извлечение различных фрагментов информации из анализируемых данных. Эти уровни могут напрямую обрабатывать необработанные или слегка обработанные исходные данные (включая обработку данных дистанционного зондирования высокого разрешения, особенно снимков с БПЛА), автономно определяя необходимые функции для поставленных задач. Эта возможность устраняет необходимость в ручном извлечении признаков, что обычно требуется в традиционных методах МО, оптимизируя таким образом рабочий процесс обработки данных и повышая квалификацию модели в распознавании сложных закономерностей.

Сверточная нейронная сеть (*CNN*) обрабатывает данные в формате нескольких массивов, таких как одномерные данные (сигналы и последовательности), двумерные данные (изображения) и трехмерные данные (видео) и позволяет автоматически извлекать признаки из каждого обработанного пикселя изображения во время обучения. Каждая единица на карте объектов имеет одинаковый вес и занимает лишь небольшую часть выборки изображений, что делает *CNN* более практичными и подробными, чем обычные нейронные сети. Мо-

дель *CNN* обычно состоит из сверточных и объединяющих слоев, за которыми следует несколько полностью связанных слоев. *CNN* имеют некоторые конструктивные параметры, включая количество фильтров, размер фильтра, тип заполнения и шаг.

Рекуррентная нейронная сеть (*RNN*) используется для задач, связанных с последовательными данными, чтобы зафиксировать их временные зависимости. *RNN* хранит историю всех прошлых элементов последовательности в своих скрытых единицах, называемых вектором состояния, и использует эту информацию при обработке входной последовательности по одному элементу за раз. *RNN* – это очень мощные модели для моделирования последовательности, но их обучение бывает очень сложной задачей из-за проблем с исчезновением или резким увеличением градиента. Для решения этой проблемы *RNN* улучшаются с помощью ячеек долговременной и кратковременной памяти (*LSTM*), которые представляют собой тщательно спроектированные рекуррентные нейроны, обеспечивающие хорошую производительность в широком спектре приложений моделирования последовательностей. Ячейки *LSTM* используют специальный блок, называемый ячейкой памяти, для длительного запоминания входных данных и предотвращения проблемы исчезающего градиента.

Глубокая нейронная сеть (*DNN*) – это модель, которая связывает веса для входных данных (нейронов), а выходной информацией обычно является функция активации (*ReLU*, *Softmax*). На каждом уровне входные данные преобразуются, и между слоями, которые не являются последовательными, не может быть обнаружено прямой связи. Для оптимизации параметров сети используется обратное распространение. Каждое обновление сети выполняется для полного обучающего набора данных, называемого эпохой. Используемое количество эпох может существенно колебаться. Сеть обучается до тех пор, пока точность проверки не достигнет максимума или не начнет падать, чтобы избежать переобучения.

Спайковая нейронная сеть (*SNN*) – это третье поколение моделей нейронных сетей, которая относится к подтипу *ANN* с существенными отличиями в самой архитектуре. Спайковые нейронные сети не имеют общей линейной структуры, а сама сеть строится с использованием определенных топологических цепочек. Вместо прямых слоев использует более сложные структуры, такие как петли или разнонаправленные связи, для передачи и обработки данных между ключевыми элементами структуры сети. Для *SNN* характерно инкрементное обучение – возможность оперативно учитывать поступающую новую информацию в режиме реального времени и тут же использовать ее в процессе обучения. Асинхронность в *SNN* имитирует

поведение биологических нейронных сетей, где существует большая степень параллелизма и случайности в обработке входящего потока информации.

Долговременная и кратковременная память (*LSTM*) – это расширение *RNN*, разработанное для уменьшения проблемы долгосрочной зависимости. В отличие от *RNN*, *LSTM* может запоминать данные в течение длительных периодов. В архитектуре *RNN* скрытые слои имеют простую структуру (например, один слой *tanh*), в то время как архитектура *LSTM* более сложная. Она состоит из четырех скрытых слоев. Основным компонентом *LSTM* является состояние ячейки. Чтобы добавить или удалить информацию из состояния ячейки, используются вентили для ее защиты с использованием сигмоидальной функции.

В последнее время получило развитие также использование для прогнозирования урожайности сельскохозяйственных культур гибридных моделей. Например, комбинации: сверточные нейронные сети (*CNN*) – глубокие нейронные сети (*DNN*), *CNN* – *XGBoost*, *CNN* – рекуррентные нейронные сети (*RNN*) и *CNN* – долгосрочная краткосрочная память (*LSTM*) [15-17].

Для определения эффективности модели используют различные метрики ошибок, перечень которых показан выше, однако наиболее распространенными являются *RMSE*, *MAE* и R^2 . Сравнение двух первых метрик ошибок осложняется доступностью исходного набора данных и различными единицами измерения, которые выражены в абсолютных или относительных величинах (% , кг/га, ц/га, т/га и др.), поэтому для сравнительной стандартизации предиктивной эффективности моделей

нами был использован коэффициент детерминации R^2 . С помощью обобщения современных литературных источников [1-5, 9, 10, 15-20], проведена оценка точности моделей (предиктивные характеристики) на основе R^2 (рис. 4).

Традиционные методы машинного обучения (например, *SVM*, *KNN*, *DT*) демонстрируют умеренную эффективность (значения R^2 находилось преимущественно в диапазоне от 0,50 до 0,70). Методы на основе регрессии имели схожие предиктивные характеристики (средний $R^2 = 0,58$) и наибольший разброс значений, что говорит о недостаточной предиктивной способности данного подхода. Модели градиентный бустинг (*GB*), адаптивный бустинг (*AdaBoost*) и экстремальный градиентный бустинг (*XGBoost*) обладали самой высокой точностью среди традиционных методов МО (значения R^2 находились преимущественно в диапазоне 0,70-0,85).

Методы глубокого обучения демонстрировали более высокую эффективность относительно традиционных методов МО (значения R^2 находились преимущественно в диапазоне 0,75-0,95). Значения R^2 в моделях *RNN*, *SNN* и *DNN* находились преимущественно в диапазоне от 0,80 до 0,90 ($R_{ср}^2 = 0,85$). Наивысшие показатели точности достигались при реализации *SNN* и *DNN* (максимальные значения R^2 составляли 0,95 и 0,94 соответственно).

Гибридные модели также демонстрировали высокую предиктивную эффективность ($R_{ср}^2 = 0,87$), например, в моделях *CNN-DNN* значения R^2 находились преимущественно в диапазоне от 0,79 до 0,92. Однако среди гибридных подходов максимальные значения R^2 (0,95) достигались с помощью модели *CNN-XGBoost*.

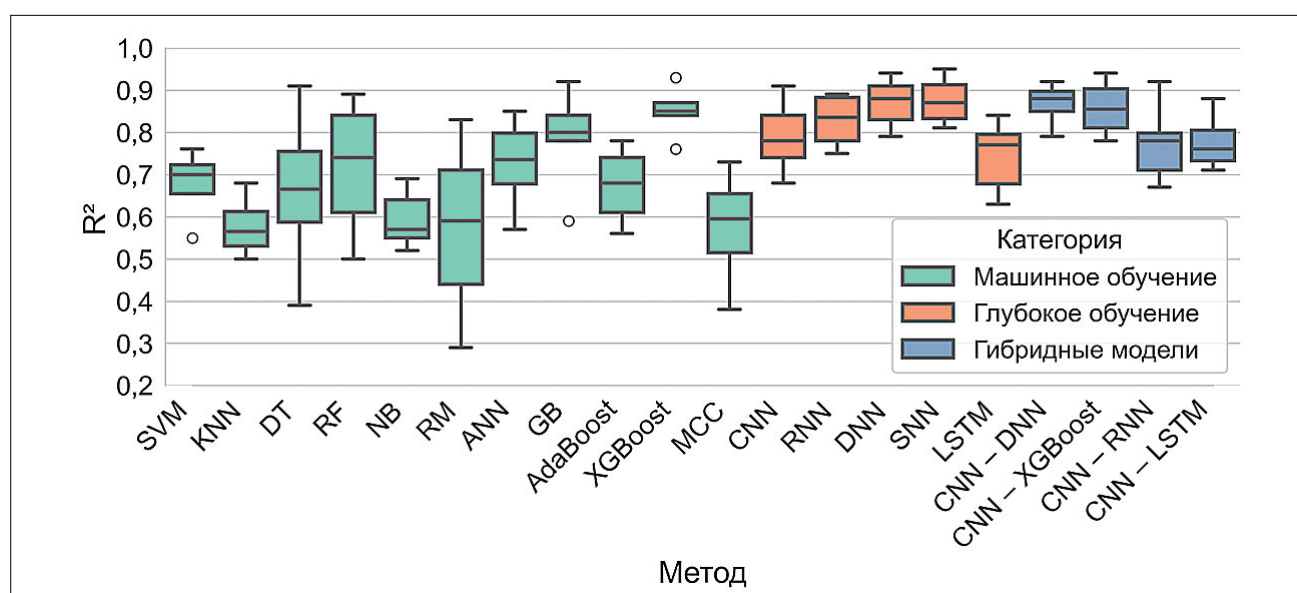


Рис. 4. Сравнение эффективности различных моделей прогнозирования урожайности сельскохозяйственных культур на основе R^2

Fig. 4. R^2 -based comparison of crop yield prediction models

Необходимо учесть, что каждый случай построения предиктивной модели уникален и характеризуется определенным набором параметров и доступных данных. Доступность данных, их качество и выбор метода с учетом имеющихся данных имеет решающее значение в предиктивной эффективности моделей. Так, при использовании данных онлайн-платформы *Kaggle* и интернета вещей (*IoT*) точность прогнозирования урожайности различных культур достигала 99,59% с использованием алгоритма сети Байеса (*Bayes Net*) и 99,46% с использованием алгоритмов наивного байесовского классификатора (*NB*) и дерева Хеффдинга (*Hoeffding Tree*) [20].

Развитием исследовательской работы в этой предметной области может быть адаптация современных подходов искусственного интеллекта к анализу специфики пространственных объектов землепользования в географическом контексте и районированных культур на основе преимущественного использования данных дистанционного зондирования и интернета вещей. Актуальными остаются также исследования, направленные на решение проблемы «ограниченности данных» и интерпретации результатов глубокого обучения для преодоления эффекта «черного ящика» нейронных сетей с помощью объяснимого искусственного интеллекта.

Выводы. Прогнозирование урожайности сельскохозяйственных культур может быть достигнуто с использованием широкого спектра методов искусственного интеллекта. Способность этих методов обрабатывать историческую и актуальную информацию о погоде, плодородии почв и фитосанитарном состоянии посевов повышает точность прогнозов урожайности. В большинстве случаев в

предиктивных моделях используются признаки, отражающие агрометеорологические ресурсы местности в течение вегетационного периода, характеристики почвы, состояние посевов по стадиям роста растений и применяемые управляющие воздействия на агрофитоценоз.

Реализация моделей машинного обучения (МО) при прогнозировании урожайности сельскохозяйственных культур состоит из четырех этапов: сбор данных; предварительная обработка и анализ данных; построение предиктивных моделей; оценка точности моделей. Ключевые этапы (предварительная обработка и анализ данных, а также построение предиктивных моделей) связаны с будущей эффективностью прогнозирования. Предварительная обработка данных, которая заключается в выполнении преобразований данных и устранении нерелевантных признаков, существенно влияет на эффективность работы предиктивных моделей. После определения релевантных признаков происходит процесс выбора алгоритма с последующим обучением и оптимизацией модели, в том числе за счет настройки гиперпараметров. Выбор модели машинного обучения для успешного прогнозирования урожайности существенно зависит от выбора релевантных признаков и их параметров (данных), а также от функциональных возможностей алгоритмов.

На основе сравнительного анализа установлено, что методы глубокого обучения демонстрируют высокую эффективность прогнозирования ($R_{cp}^2 = 0,85$) относительно традиционных методов машинного обучения ($R_{cp}^2 = 0,58$). Гибридные подходы превосходят все традиционные методы машинного обучения и сравнимы с методами глубокого обучения по величине R^2 и разбросов его значений ($R_{cp}^2 = 0,87$).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. 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. N29. 95-112. DOI: 10.1007/s11831-021-09569-8.
2. Shingade S.D., Mudhalwadkar R.P. Analysis of crop prediction models using data analytics and ML techniques: a review. *Multimedia Tools and Applications*. 2024. N83. 37813-37838. DOI: 10.1007/s11042-023-17038-6.
3. Meghraoui K., Sebari I., Pilz J. et al. Applied deep learning-based crop yield prediction: a systematic analysis of current developments and potential challenges. *Technologies*. 2024. 12. N4. 43. DOI: 10.3390/technologies12040043.
4. Степанов А.С., Асеева Т.А., Дубровин К.Н. Построение и оценка точности регрессионных моделей для определения урожайности зерновых и зернобобовых культур на основе данных дистанционного Земли и климатических характеристик // *ИнтерКарто. ИнтерГИС*. 2020. 26. N3. С. 159-169. DOI: 10.35595/2414-9179-2020-3-26-159-169.
5. 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. N2 (380). С. 111-137. DOI: 10.37162/2618-9631-2021-2-111-137.
6. Бисчоков Р.М., Ахматов М.М. Анализ и прогноз урожайности сельскохозяйственных культур методами нечеткой логики // *Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета*. 2021. N168. С. 274-287. DOI: 10.21515/1990-4665-168-020.
7. Благов Д.А., Митрофанов С.В., Никитин В.С. и др. Математические модели и программный комплекс по прогнозированию урожайности сельскохозяйственных культур // *Агротехника и энергообеспечение*. 2019. N3 (24). С. 182-188. EDN: OMWNXT.

8. Архипова М.Ю., Смирнов А.И. Современные направления прогнозирования урожайности сельскохозяйственных культур на основе использования эконометрических моделей // *Вопросы статистики*. 2020. 27. N5. 65-75. DOI: 10.34023/2313-6383-2020-27-5-65-75.
9. Momenpour S.E., Bazgeer S., Moghbel M. A bibliometric analysis of the literature on crop yield prediction: insights from previous findings and prospects for future research. *International Journal of Biometeorology*. 2024. 68. 829-842. DOI: 10.1007/s00484-024-02628-2.
10. Макеев К.А., Греченева А.В., Котов Я.С. и др. Сравнение эффективности алгоритмов машинного обучения в задачах прогнозирования урожайности сельскохозяйственных культур // *Известия Тульского государственного университета. Технические науки*. 2023. N2. С. 203-206. DOI: 10.24412/2071-6168-2023-2-203-206.
11. Дудко Ю.В., Маринкин Е.Б., Владимирова Т.М. Применение нейросетевых технологий в предсказании плодородия почв с зернобобовыми культурами и прогнозировании их урожайности // *Вопросы науки: инноватика, техника и технологии*. 2019. N1. С. 67-72. EDN: YXPXCX.
12. Рогачев А.Ф. Системный анализ и прогнозирование временных рядов урожайности на основе автокорреляционных функций и нейросетевых технологий // *Известия Нижневолжского агроуниверситетского комплекса*. 2018. N3 (51). С. 309-316. EDN: VOQSWI.
13. Рогачев А.Ф., Мелихова Е.В. Адаптация алгоритмов и обоснование инструментария для нейросетевого прогнозирования урожайности агрокультур с использованием ретроспективных данных // *Известия Нижневолжского агроуниверситетского комплекса*. 2020. N1(57). С. 290-302. DOI: 10.32786/2071-9485-2020-01-29.
14. Бисчоков Р.М. Анализ, моделирование и прогноз урожайности сельскохозяйственных культур средствами искусственных нейронных сетей // *Вестник Российского университета дружбы народов. Серия: Агрономия и животноводство*. 2022. 17. N2. С. 146-157. DOI: 10.22363/2312-797X-2022-17-2-146-157.
15. Boppudi S. J. S. Deep ensemble model with hybrid intelligence technique for crop yield prediction. *Multimedia Tools and Applications*. 2024. 1-21. DOI: 10.1007/s11042-024-18354-1.
16. Oikonomidis A., Catal C., Kassahun A. Hybrid deep learning-based models for crop yield prediction. *Applied Artificial Intelligence*. 2022. 36. N1. 2031822. DOI: 10.1080/08839514.2022.2031823.
17. Cao J., Wang H., Li J. et al. Improving the forecasting of winter wheat yields in Northern China with machine learning—dynamical hybrid subseasonal-to-seasonal ensemble prediction. *Remote Sensing*. 2022. 14(7). 1707. DOI: 10.3390/rs14071707.
18. Cedric L.S., Adoni W.Y.H., Aworka R. et al. Crops yield prediction based on machine learning models: Case of West African countries. *Smart Agricultural Technology*. 2022. N2. 100049. DOI: 10.1016/j.atech.2022.100049.
19. Elavarasan D., Vincent D.R., Sharma V. et al. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electronics in Agriculture*. 2018. N155. 257-282. DOI: 10.1016/j.compag.2018.10.024.
20. Meghraoui K., Sebari I., Pilz J. et al. Applied deep learning-based crop yield prediction: a systematic analysis of current developments and potential challenges. *Technologies*. 2024. N12(4). 43. DOI: 10.3390/technologies12040043.

REFERENCES

1. 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. 29. 95-112 (In English). DOI: 10.1007/s11831-021-09569-8.
2. Shingade S.D., Mudhalwadkar R.P. Analysis of crop prediction models using data analytics and ML techniques: a review. *Multimedia Tools and Applications*. 2024. 83. 37813-37838 (In English). DOI: 10.1007/s11042-023-17038-6.
3. Meghraoui K., Sebari I., Pilz J. et al. Applied Deep Learning-Based Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges. *Technologies*. 2024. 12. N4. 43 (In English). DOI: 10.3390/technologies12040043.
4. Stepanov A.S., Aseeva T.A., Dubrovina K.N. Building and precision assessment of regression models for determining of cereals' and legumes' crop yield based on Earth remote sensing data and climatic characteristics. *InterCarto. InterGIS*. 2020. 26. N3. 159-169 (In Russian). DOI: 10.35595/2414-9179-2020-3-26-159-169.
5. Strashnaya A.I., Bereza O.V., Klang P.S. Forecasting grain crop yield based on the integration of ground and satellite data in the subjects of the Southern Federal District. *Hydrometeorological Research and Forecasting*. 2021. N2 (380). 111-137 (In Russian). DOI: 10.37162/2618-9631-2021-2-111-137.
6. Bischokov R.M., Akhmatov M.M. Agricultural crops analysis and yield forecast by fuzzy logic methods. *Polythematic online Scientific Journal of Kuban State Agrarian University*. 2021. N168. 274-287 (In Russian). DOI: 10.21515/1990-4665-168-020.
7. Blagov D.A., Mitrofanov S.V., Nikitin V.S. et al. Mathematical models and software package for forecasting crop yields. *Agrotehnika i energoobespechenie*. 2019. N3 (24). 182-188 (In Russian). EDN: OMWNXT.
8. Arkhipova M.Yu., Smirnov A.I. Current trends in crop yield forecasting based on the use of econometric models. *Voprosy statistiki*. 2020. 27. N5. 65-75 (In Russian). DOI: 10.34023/2313-6383-2020-27-5-65-75.
9. Momenpour S.E., Bazgeer S., Moghbel M. A bibliometric

- analysis of the literature on crop yield prediction: insights from previous findings and prospects for future research. *International Journal of Biometeorology*. 2024. N68. 829-842 (In English). DOI: 10.1007/s00484-024-02628-2.
10. Makeev K.A., Grecheneva A.V., Kotov Ya.S. et al. Comparison of the effectiveness of machine learning algorithms in crop yield forecasting tasks. *Izvestiya Tula State University*. 2023. N2. 203-206 (In Russian). DOI: 10.24412/2071-6168-2023-2-203-206.
 11. Dudko Yu.V., Marinkin E.B., Vladimirova T.M. Application of neural network technologies in predicting soil fertility with leguminous crops and predicting their yields. *Voprosy Nauki: Innovatika, Tekhnika i Tekhnologii*. 2019. N1. 67-72 (In Russian). EDN: YXPXCX.
 12. Rogachev A.F. System analysis and forecasting of time series of crop production based on autocorrelation function and neural network technologies. *Proceedings of Lower Volga Agro-University Complex*. 2018. N3 (51). 309-316 (In Russian). EDN: VOQSWI.
 13. Rogachev A.F., Melikhova E.V. Justification of algorithms and tools for neural network forecasting of agricultural productivity using retrospective data. *Proceedings of Lower Volga Agro-University Complex*. 2020. N1(57). 290-302 (In Russian). DOI: 10.32786/2071-9485-2020-01-29.
 14. Bischokov R.M. Analysis, modelling and forecasting of crop yields using artificial neural networks. *RUDN Journal of Agronomy and Animal Industries*. 2022. 17. N2. 146-157 (In Russian). DOI: 10.22363/2312-797X-2022-17-2-146-157.
 15. Boppudi S.J.S. Deep ensemble model with hybrid intelligence technique for crop yield prediction. *Multimedia Tools and Applications*. 2024. 1-21 (In English). DOI: 10.1007/s11042-024-18354-1.
 16. Oikonomidis A., Catal C., Kassahun A. Hybrid deep learning-based models for crop yield prediction. *Applied Artificial Intelligence*. 2022. 36. N1. 2031822. (In English). DOI: 10.1080/08839514.2022.2031823.
 17. Cao J., Wang H., Li J. et al. Improving the forecasting of winter wheat yields in Northern China with machine learning—dynamical hybrid subseasonal-to-seasonal ensemble prediction. *Remote Sensing*. 2022. N14(7). 1707 (In English). DOI: 10.3390/rs14071707.
 18. Cedric L.S., Adoni W.Y.H., Aworka R. et al. Crops yield prediction based on machine learning models: Case of West African countries. *Smart Agricultural Technology*. 2022. 2. 100049 (In English). DOI: 10.1016/j.atech.2022.100049.
 19. Elavarasan D., Vincent D.R., Sharma V. et al. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electronics in Agriculture*. 2018. N155. 257-282 (In English). DOI: 10.1016/j.compag.2018.10.024.
 20. Meghraoui K., Sebari I., Pilz J. et al. Applied deep learning-based crop yield prediction: a systematic analysis of current developments and potential challenges. *Technologies*. 2024. N12(4). 43. (In English). DOI: 10.3390/technologies12040043.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Conflict of interest

The authors declare no conflict of interest.

Заявленный вклад соавторов:

Каличкин В.К. – научное руководство, постановка проблемы, формулирование основной концепции работы, целей и задач исследования, написание текста, формирование общих выводов, итоговая переработка статьи;
 Максимович К.Ю. – формулирование основных целей и задач исследования, анализ литературных источников, доработка основного текста и разработка графического материала, участие в формировании общих выводов;
 Аleshchenko O.A. – поиск и первичный анализ литературных источников, работа с базами данных, техническое редактирование;
 Аleshchenko V.V. – методология, доработка окончательного варианта рукописи, верификация выводов.
 Авторы прочитали и одобрили окончательный вариант рукописи.

Coauthors' contribution:

Kalichkin V.K. – scientific supervision, problem statement, formulation of the main concept, definition of research objectives and tasks, drafting of the manuscript, formulation of general conclusions, and final revision of the manuscript;
 Maksimovich K.Yu. – definition of key research objectives and tasks, literature review, revision of the manuscript, development of graphical materials, and contribution to the formulation of general conclusions;
 Aleshchenko O.A. – identification and preliminary analysis of literature sources, database management, and technical editing;
 Aleshchenko V.V. – development of methodology, revision of the final manuscript version, and validation of the conclusions.
 The authors read and approved the final manuscript.

Статья поступила в редакцию
 Статья принята к публикации

The paper was submitted to the Editorial Office on
 The paper was accepted for publication on

03.04.2025
 27.05.2025