

## Моделирование урожайности зерновых культур сельскохозяйственных регионов с использованием технологий компьютерного зрения<sup>1</sup>

В статье рассматриваются новые направления моделирования урожайности зерновых культур в сельскохозяйственных регионах России на основе использования дистанционных возможностей получения информации о состоянии полей. Предлагаемый подход позволяет найти новые решения в разработке системы показателей, обосновании методологических платформ и моделей для получения более точных прогнозных оценок по сравнению с традиционными регрессионными моделями за счет использования системы компьютерного зрения в качестве дополнительного источника информации. Статистическая значимость спутниковых фотоснимков полей для повышения точности моделей прогнозирования урожайности сельскохозяйственных культур подтверждается проверкой соответствующей статистической гипотезы. Значительный интерес в исследовании представляет сравнение классических эконометрических инструментов с различными нейросетевыми моделями с точки зрения нахождения оптимальной модели, позволяющей повысить точность прогнозных оценок. Апробация предлагаемого инструментария проводилась на данных по 100 сельскохозяйственным полям, расположенным в муниципальных образованиях 43 регионов России, которые были выбраны пропорционально объему продукции растениеводства данного региона. Проведенное исследование показало преимущество нейросетевой модели по смешанным данным по сравнению с другими нейросетевыми моделями (многослойный перцептрон и свёрточная нейронная сеть), а также с традиционными регрессионными моделями. Нейросетевая модель по смешанным данным в условиях неопределенности и большого количества данных различной природы позволила получить более точные прогнозные оценки по сравнению с другими классами моделей. Также было показано, что несмотря на то, что экологические факторы оказывают разное влияние на урожайность сельскохозяйственных культур, их необходимо учитывать наряду с социально-экономическими характеристиками. Использование новых моделей и типов данных, отличных от классической табличной информации, может давать существенное преимущество в точности прогнозирования и объяснении решающих факторов. Результаты проведенного анализа могут использоваться в исследованиях и мониторинге развития сельскохозяйственного производства региональных муниципальных образований, определения потребностей в ресурсах, необходимых для успешного ведения хозяйства, а также при разработке отраслевых и комплексных проектов и программ развития агрокомплекса.

**Ключевые слова:** прогнозирование урожайности, компьютерное зрение, нейронные сети, эконометрические модели, сельскохозяйственный статистический анализ, эконометрическое моделирование, космические фотоснимки

### Благодарности

Исследование выполнено при поддержке Российского научного фонда, грант № 22-28-20360 «Трансформация образа жизни людей в цифровой среде современного мегаполиса».

Автор выражает благодарность студенту НИУ ВШЭ А. И. Смирнову за помощь в подготовке статьи.

**Для цитирования:** Архипова М. Ю. Моделирование урожайности зерновых культур сельскохозяйственных регионов с использованием технологий компьютерного зрения // Экономика региона. 2022. Т. 18, вып. 2. С. 581-594. <https://doi.org/10.17059/ekon.reg.2022-2-20>.

<sup>1</sup> © Архипова М. Ю. Текст. 2022.

Marina Yu. Arkhipova

National Research University "Higher School of Economics", Moscow, Russian Federation  
<https://orcid.org/0000-0002-9022-7385>, e-mail: [marhipova@hse.ru](mailto:marhipova@hse.ru)

## Modelling Crop Yield in Agricultural Regions Using Computer Vision Technology

*The article examines new methodologies for modelling crop yield in agricultural regions of Russia based on the use of remote capabilities to get information on the field state. The proposed approach can be applied to develop indicator systems and create methodological platforms and models necessary to obtain more accurate estimates. In comparison with the traditional regression model, this method uses computer vision technology to gather additional data. Statistical hypothesis testing confirmed the significance of satellite photographs of fields for improving the accuracy of crop yield forecasting models. Traditional econometric tools were compared with various neural networks in order to discover the optimal model. The proposed tools were tested using data from 100 agricultural fields located in municipalities of 43 Russian regions, selected in proportion to the volume of crop production in this region. The conducted analysis showed the advantage of the mixed data neural network in comparison with other neural (multilayer perceptron and convolutional neural network) and regression models. In conditions of uncertainty and a large amount of data, the mixed data neural network can help obtain more accurate estimates. Additionally, while environmental factors have different effects on crop yields, they must be considered along with socio-economic characteristics. The use of new models and data types differing from table information can significantly improve the forecasting accuracy and interpretation. The analysis results can be used for examining and monitoring agricultural production in regional municipalities, determining farm resource requirements, as well as for creating sectoral and comprehensive projects and programmes for the development of the agricultural industry.*

**Keywords:** yield forecasting, computer vision, neural networks, econometric models, agricultural statistical analysis, econometric modelling, satellite photos

### Acknowledgements

*The article has been prepared with the support of the Russian Science Foundation, grant No. 22-28-20360 "Transformation of people's lifestyle in the digital environment of a modern metropolis".*

*The author would like to thank the HSE student A. I. Smirnov for his help in preparing the article.*

**For citation:** Arkhipova, M. Yu. (2022). Modelling Crop Yield in Agricultural Regions Using Computer Vision Technology. *Ekonomika regiona [Economy of regions]*, 18(2), 581-594, <https://doi.org/10.17059/ekon.reg.2022-2-20>.

### Введение

Сельское хозяйство играет ключевую роль в обеспечении продовольственной безопасности страны и ее жителей продуктами питания, развитии сельских территорий, а также в вопросах защиты окружающей среды и противодействия глобальному потеплению. Вместе с тем существует достаточно широкий пласт задач, которые требуют первоочередного решения для развития сельскохозяйственных производств. Среди таких задач следует отметить разработку системы показателей, необходимых для исследования урожайности сельскохозяйственных культур, выявление факторов оказывающих стимулирующее и тормозящее воздействие на развитие сельскохозяйственных производств, повышение точности прогнозных моделей и др. Вместе с тем развитие вычислительной техники, систем компьютерного зрения, совершенствование методологической базы и исследовательского инструментария позволяют решать ряд обозначенных задач

и оперативно реагировать на происходящие изменения. Развитие аналитического инструментария за счет использования новых цифровых возможностей позволяет значительно улучшить аналитическую составляющую при планировании и прогнозировании основных индикаторов развития сельского хозяйства. Оптимизация операционной деятельности агрокомплекса поможет решить проблемы окружающей среды с помощью выявления неплодородных, пересушенных или заболоченных, обедневших минеральными веществами, малоэффективных для ведения растениеводства полей, на которых возможно выращивание искусственных лесов и восстановление изначальной экосистемы. Также повышение урожайности потенциально решает экологические проблемы, связанные с использованием нефтепродуктов в качестве основного сырья для топлива. При высокой рентабельности агрокомплекса в будущем возможно замещение вредных при производстве соединений на ве-

щества растительного происхождения, например, биоэтанол и биодизель.

В России существуют отдельные коммерческие и некоммерческие проекты по частичному сбору статистических данных, включая использование спутниковых фотоснимков. Среди них можно отметить проект «Vega-PRO»<sup>1</sup>, в основу которого легли разработки Института космических исследований Российской академии наук, а также Систему дистанционного мониторинга земель сельскохозяйственного назначения (СДМЗ)<sup>2</sup>, которая начала разрабатываться по заказу Министерства сельского хозяйства РФ. Тем не менее, на сегодняшний день нет единых баз данных со спутниковыми изображениями и табличной информацией о полях всех регионов России и, соответственно, существует недостаток данных, на основании которых могли бы приниматься управленческие решения в агрокомплексе. По этой причине представленные результаты и используемые методы могут быть интересны специалистам и лицам, принимающим управленческие решения для улучшения анализа урожайности и повышения уровня рентабельности агрокультурного сектора России.

### Основные подходы к оценке урожайности

При проведении обзора литературы нас, прежде всего, интересовало, какие статистические показатели могут быть положены в основу проведения анализа урожайности и как можно получить дополнительную информацию, позволяющую повысить точность прогнозных моделей, и инструментарий, который лучше использовать для этих целей.

Отметим, что бурное развитие современных методов машинной обработки данных (MachineLearning, DataCulture) в начале двухтысячных годов послужило сигналом и дало возможность новому развитию современного многомерного статистического и эконометрического инструментария на основе использования таких направлений анализа данных, как случайные леса и нейросети. Данный инструментарий, согласно ряду исследований, позволяет в условиях неопределенности и большого количества неоднородных данных получить более устойчивые точные результаты по сравнению с традиционными методами.

Среди таких пионерных работ в области моделирования и прогнозирования урожайности

сельскохозяйственных культур можно отметить работу Луки Салвати (Salvati et al., 2010), который одним из первых включил в прогнозные модели экономические и социальные данные о муниципалитетах. Используемый им набор переменных, среди которых, например, показатели, отвечающие за благосостояние населения, тип почвы, количество фотосинтетически активной биомассы (NDVI), температуру и осадки, во многом опирался на рекомендации, представленные в статьях других ученых (см, например (Basso et al., 2000; Pantazi et al., 2016)). Интерес с точки зрения расширения числа факторов, используемых для анализа, представляют и работы Рэнджан, Чандел, Кот, Бэллол, Жу и др. исследователей, которые предложили для прогнозирования урожайности использовать спутниковые фотографии (Ranjan et al., 2019). Фотографии были подробными и сняты крупным планом, полученная сверточная нейронная сеть позволила достаточно точно предсказать урожайность. Необходимо отметить и ряд других работ, посвященных сельскохозяйственному статистическому анализу (Pöldaru et al., 2005; Zhang et al., 2010; Zhang et al., 2019; Dharmadhikari, 2018), которые были использованы в работе при разработке системы статистических показателей.

Что касается методической стороны вопроса, то здесь пока не выработано единого мнения о том, какие модели лучше ведут себя при прогнозировании основных сельскохозяйственных показателей. Помимо эконометрического инструментария в научном сообществе все чаще стали использоваться различные типы искусственных нейронных сетей, которые позволяли получать достаточно точные результаты и хорошо зарекомендовали себя в различных сферах деятельности.

Среди базовых работ по нейросетям следует отметить работы С. Осовского и В.И. Ширяева (Осовский, 2004; Ширяев, 2013). С. Осовский описал ряд методов распознавания и классификации изображений, В.И. Ширяев представил алгоритм обработки нейронными сетями входных сигналов в соответствии с архитектурой модели, количеством слоев и нейронов, рассмотрел методы оптимизации, которые сводятся к подходу оптимизации с помощью градиентного спуска. Среди исследований, проведенных в последние несколько лет, можно отметить работу (Kung et al., 2016), в которой ее авторы Кунг, Куо, Чен и Цай использовали ансамблевую нейронную сеть (ENN), содержащую несколько моделей с разным количеством скрытых слоев и нейронов. Исследователи про-

<sup>1</sup> См. Спутниковый сервис анализа вегетации (<http://pro-vega.ru/>).

<sup>2</sup> См. <https://sovzond.ru/projects/2072/>.

демонстрировали на реальных данных, что метод ENN дает более точные результаты по сравнению с простыми нейронными сетями с обратным распространением ошибки и регрессионным анализом.

Совершенствованию методов кодирования изображений полей посвящена работа Пантази и соавторов (Pantazi et al., 2016). Авторы статьи использовали искусственные нейронные сети для прогнозирования урожайности в Бедфордшире, Великобритания. В анализ были включены некоторые параметры почвы и нормализованный разностный индекс растительности (NDVI), полученный с помощью анализа спутниковых изображений. Уровень NDVI может быть средним или достаточно высоким, однако у поля могут быть проблемные участки с гораздо более низкой вегетацией, которые препятствуют успешному земледелию и высокой урожайности. Именно по этой причине, космические фотографии полей целесообразно включать в анализ.

С точки зрения необходимости учета нелинейных связей при прогнозировании урожайности зерновых культур интерес представляет работа Де ла Каса и соавторов (Dela Casa et al., 2018), в которой авторы показали высокую предсказательную способность нелинейных моделей в течение нескольких лет. Выводы авторов о нелинейном характере связей между рядом рассматриваемых переменных интересны и с точки зрения выбора нейросетевых моделей. Используемая в работе многослойная нейронная сеть позволила учесть нелинейные связи между признаками.

Таким образом, проведенный анализ современных исследований показал, что, с одной стороны, не существует устоявшегося мнения об использовании того или иного инструментария для прогнозирования урожайности зерновых культур, с другой стороны, проведенный обзор позволил определить базовый набор показателей, которые могут быть использованы при проведении исследований такого уровня.

### Исходные данные и методы исследования

При проведении исследования нас интересовали следующие вопросы, которые были сформулированы в виде приведенных ниже гипотез.

$H_1$ : Спутниковые фотоснимки полей являются статистически значимым фактором и повышают точность прогнозных моделей урожайности сельскохозяйственных культур (полей).

Космические фотоснимки вегетации полей имеют специфическую структуру: различные элементы снимков соответствуют речкам, пересушенным участкам, постройкам, незасеянным участкам и т. д., что может быть использовано при моделировании.

$H_2$ : нейронные сети позволяют получить более, по сравнению с классическим эконометрическим инструментарием, точные результаты и могут быть использованы для прогнозирования урожайности полей муниципальных образований.

Выбор нейросетей в работе не случаен. В последние годы их использование набирает все большую популярность и признание среди исследователей и аналитиков в связи с высокой предсказательной способностью и возможностью интерпретации промежуточных результатов.

$H_3$ : экологические факторы (загрязнение воздуха, почвы, воды) в равной степени оказывают отрицательное воздействие на показатели урожайности полей.

Для решения поставленных задач использовались методы статистической проверки гипотез. Согласно данным методам выдвигается и статистически проверяется нулевая гипотеза, которая по смыслу противоположна проверяемому утверждению. Если на основе реальных выборочных статистических данных удастся отклонить нулевую гипотезу, то можно с высокой достоверностью говорить о том, что выдвинутое утверждение доказано (в работе вероятность ошибки полученных выводов не превышала  $\alpha = 0,05$ )

Для анализа были отобраны данные о 100 муниципальных образований, расположенных в 43 регионах России. Отобранные регионы были обследованы по 12 переменным за 2018 г., описывающим экономические и природные характеристики района, в котором находится поле. Выбор регионов осуществлялся пропорционально объему продукции растениеводства в данном регионе, что позволило отобрать регионы с традиционно развитым сельским хозяйством, характеризующиеся благоприятными климатическими условиями и типами почв для выращивания продукции растениеводства. Это преимущественно степные области Центрально-Чернозёмного района России, лесостепные районы Северного Кавказа, Поволжья, южные районы Урала и Западной Сибири. Наибольшее число отобранных для анализа полей расположено в Ростовской, Воронежской областях, Ставропольском, Краснодарском, Алтайском



Рис. 1. Распределение полей, отобранных для анализа по регионам России (источник: Федеральная служба государственной статистики (Росстат))

Fig. 1. Distribution of fields selected for the analysis by Russian regions

краях, республиках Башкортостан и Татарстан.

Распределение полей по регионам представлено на рисунке 1. Темно-синим цветом отмечены регионы, которые представлены пятью-шестью полями, синим — от трех до четырех и светло-синим — от одного до двух полей. Отметим, что исходные данные взяты из отчета Росстата по удельному весу продукции растениеводства в продукции сельского хозяйства по категориям хозяйств по субъектам РФ за 2018 г.

В качестве экзогенных переменных были использованы количественные и качественные переменные:

$X_1$  — доля растениеводства в сельском хозяйстве муниципального образования (%). Предполагается, что чем выше данный показатель, тем выгоднее вести растениеводство в регионе, выше урожайность.

$X_2$  — доля посевной площади от площади МО. Считается, что чем больше территорий района засеяно зерновыми культурами, тем выше урожайность.

$X_3$  — среднемесячная заработная плата в сельскохозяйственном секторе (Раздел ОКВЭД А). Предполагается, что высокая среднемесячная заработная плата повышает мотивацию работников в эффективном ведении хозяйства.

$X_4$  — количество тракторов на гектар засеянной территории (единиц). Данный показатель отвечает за оснащенность агрокультурного комплекса современной техникой и позволяет

проследить за его влиянием на урожайность зерновых культур.

$X_5$  — внесение минеральных удобрений (тонн на гектар). Показатель характеризует уровень питательной среды почвы для выращивания сельскохозяйственных культур.

$X_6$  — численность сельского населения на 1 января 2018 г. (чел.) Предполагается, что высокая численность населения положительно сказывается на специализации рынка труда, следовательно, можно ожидать более высокие показатели урожайности.

$X_7$  — наличие (отсутствие) загрязнения воздуха по критерию превышения предельно допустимой концентрации вредных веществ, рассчитанное Росгидрометом (фиктивная переменная). Согласно российскому законодательству установлено ПДК для целого ряда частиц, которые могут как вредить здоровью людей, быть опасными для экологии, так и серьезно мешать эффективному земледелию.

$X_8$  — количество серы на снежном покрове, измеряемое в граммах на квадратный километр. Считается, что недостаток (переизбыток) серы в почве может сильно влиять на злаковые, одно из ключевых семейств культурных растений в сельском хозяйстве России (Власюк, 1969).

$X_9$  — количество азота на снежном покрове, измеряемое в килограммах на квадратный километр. Исследователи-химики выявляют, что избыточное содержание азота приводит к увеличению периода вегетации, снижению урожая и его качества, так же, как и недоста-

ток данного элемента (Архипова, Смирнов, 2020). Как и в случае содержания серы ( $X_8$ ), недостаток элемента можно отнести к более низкому уровню контроля за почвой, а переизбыток, помимо того, к химическому почвенному загрязнению.

$X_{10}$  — уровень кислотности (рН) снежного покрытия. Вероятно, большинство культур хорошо растут в нейтральных или немного щелочных почвах, к умеренно-щелочным относят, к примеру, чернозем (Jordanova, 2016). Предполагается, что чем ниже кислотность, тем лучше для большинства культур, которые произрастают в выбранных муниципалитетах.

$X_{11}$  — вывоз твердых коммунальных отходов в год (тыс. м<sup>3</sup> на душу населения). Данный фактор отражает эффективность сбора и вывоза бытового мусора, исходя из предпосылки, что в среднем россияне в сельской местности имеют одинаковый уровень выброса мусора. Считается, что твердые бытовые отходы мало применимы в сельском хозяйстве и, мало того, почвы и растительность вблизи мусорных полигонов испытывают всевозможные негативные воздействия (Архипова, Смирнов, 2020).

$X_{12}$  — качество воды водоемов суммарно по классам загрязнения 4 и 5, согласно методике подсчета Гидрохимического института. К 4-му и 5-му классам загрязнения поверхностных вод относятся разряды: грязная, грязная, очень грязная, очень грязная, экстремально грязная вода (Архипова, Смирнов, 2020; Перечень..., 1999). Классификация производится по величинам комбинаторного индекса загрязненности воды с учетом ряда дополнительных факторов. В работе использованы готовые индексы, предоставленные Гидрохимическим институтом.

$X_{13}$  — диффузное горизонтальное облучение, инсоляция (среднегодовое количество солнечной радиации на квадратный метр территории), кВт / м<sup>2</sup>. Данный показатель демонстрирует уровень солнечной радиации, необходимой для роста растений и фотосинтеза, а также расположение регионов РФ по северной широте. Предполагается получить линейное положительное влияние на урожайность, так как в России не так много районов повышенных температур и стабильной засухи.

$X_{14}$  — наличие или отсутствие загрязнения города в муниципалитете, фиктивная переменная. Рассчитывается получить значимую положительную взаимосвязь с урожайностью в рублях из-за возможностей сбыта продукции без затрат на транспортировку и хранения

продукции, а также из-за большей оснащенности агрокомплексов необходимыми промышленными товарами и услугами рынка труда.

В качестве результирующего признака  $Y$  использовался показатель «урожайность зерновых культур поля муниципалитета» (руб. на га засеянной площади). При выборе зависимой переменной мы опирались на работу Сальвати и соавторов (Salvati et al., 2010), в которой авторы использовали данный показатель для изучения урожайности полей, аргументируя это тем, что он позволяет отразить продуктивность поля как в физическом объеме, так и в денежном выражении. Результирующий признак был прологарифмирован, так как традиционные регрессионные модели требуют нормального распределения зависимой переменной.

Для реализации основной цели работы (прогнозирование урожайности полей муниципалитетов сельскохозяйственных регионов России), а также статистической проверки гипотезы  $H_2$  использованы регрессионные модели (традиционная регрессионная модель, основанная на методе наименьших квадратов, регрессионная модель по урезанной выборке) и нейронные сети (многослойный перцептрон (*MLP*), свёрточная нейронная сеть (*CNN*) и общая модель с двумя ветвями по смешанным данным). Отметим, что в статье представлены нейросетевые модели, показавшие лучший вариант. Подробную информацию об использовании традиционной регрессионной модели, основанной на методе наименьших квадратов и регрессионной модели по урезанной выборке для прогнозирования урожайности зерновых культур, можно найти в работе М.Ю. Архиповой и А.Ю. Смирнова (Архипова, Смирнов, 2020).

Нейронная сеть — это попытка с помощью математических моделей воспроизвести работу человеческого мозга. Нейронные сети, согласно С. Осовскому (Осовский, 2004), — это раздел искусственного интеллекта, в котором для обработки сигналов используются явления, аналогичные происходящим в нейронах живых существ. Важнейшая особенность сети, свидетельствующая о ее широких возможностях и огромном потенциале, состоит в параллельной обработке информации всеми звеньями, что позволяет значительно ускорить данный процесс. Кроме того, при большом количестве межнейронных соединений сеть приобретает устойчивость к ошибкам и не претерпевает существенных возмущений. Другое не менее важное свойство — способность к обучению и обобщению накопленных знаний.

Для прогнозирования урожайности в работе использованы три типа нейросетевых моделей, среди которых многослойный перцептрон (*MLP*), свёрточная нейронная сеть (*CNN*) и модель на смешанных данных.

Выбор первой модели нейронных сетей (многослойный перцептрон) опирался на статью (Haghverdi, Washington-Allen, Leib, 2018), в которой, согласно полученным авторами результатам, она показала вполне приемлемые результаты, даже без учета информации закодированных спутниковых изображений. Модель принимает значения урожайности и табличные значения независимых переменных.

Вторая модель (свёрточная нейронная сеть) принимает значения зависимой переменной и спутниковые изображения полей муниципальных образований в виде векторов, каждое число которых обозначает интенсивность красного, зеленого, синего и черного цветов в зависимости от вегетации полей. При этом преимуществом модели является возможность учета объединенного изображения из двух фотографий NDVI: на начало осени и начала лета. Необходимость учета двух фотоснимков обусловлена тем, что по одному снимку вегетации делать выводы о произрастании зеленой массы на поле ошибочно, так как существуют озимые культуры, которые зреют в начале лета, и яровые, урожай которых приходится на осень.

Третья модель (модель на смешанных данных) принимает итоговые значения слоев обеих моделей и прогнозирует урожайность, как с помощью табличных значений, так и с помощью фотографий.

Интерпретация модели искусственной нейронной сети не всегда является «черным ящиком», результаты модели по смешанным данным могут и должны объяснять влияние факторов на результирующую переменную. Для данных целей создано большое количество интерпретаторов, но одними из наиболее часто используемых являются *LIME* (*Local Interpretable Model-agnostic Explanations*) и *SHAP* (*SHapley Additive exPlanations*). *LIME* — это методика, которая объясняет предсказание любой модели интерпретируемым и точным способом, изучая модель локально вокруг предсказанного значения. Цель данного интерпретатора — минимизация функции:

$$\xi(x) = \arg \min_{g \in G} L(f, g, \pi_x) + \Omega(g), \quad (1)$$

где  $f$  — зависимая переменная;  $x$  — регрессоры;  $g$  — модель;  $\pi$  — мера близости для определения окрестности вокруг  $x$ .

На первом шаге алгоритма нужно выбрать интересующее наблюдение по муниципальному образованию из тестовой выборки, для которого необходимо получить объяснение прогноза «черного ящика». *LIME* последовательно меняет эффекты переменных в исследуемой модели для того чтобы понять, насколько прогноз чувствителен к тому или иному фактору (Molnar, 2018). Берутся новые наблюдения в соответствии с их близостью к исследуемому наблюдению, таким образом, объяснение создается путем локального приближения базовой модели к интерпретируемой. Интерпретируемые модели, как правило, — это линейные модели с сильной регуляризацией. Они обучаются на небольших возмущениях исходного экземпляра и обеспечивают только локальное приближение, а не глобальное. В связи с этим фактом существует критика интерпретатора, который, кроме того, может упустить значимые нелинейные связи из-за локального приближения изучаемой модели к линейной.

Второй интерпретатор — *SHAP* относится к классу моделей, называемых аддитивными методами атрибуции признаков, где объяснение выражается в виде линейной функции признаков. Значение *SHAP* оценивает разницу с выходным значением функции путем включения регрессора для всех комбинаций функций, кроме того, величина Шепли — это средневзвешенная от всех предельных вкладов всех возможных сочетаний признаков. Данный интерпретатор берет свое начало из теории игр и сообщает, как оптимально распределить эффекты среди переменных. В алгоритме использована формула значения *SHAP*:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)], \quad (2)$$

где  $f$  — модель;  $S$  — количество переменных без той переменной, которой рассматривается *SHAP* value;  $M$  — количество всех наблюдений.

Также в работе были использованы активационная функция *ReLU*

$$(R(x) = \max(0, x) \rightarrow \frac{\partial f}{\partial x} = 1, \text{ если } x \geq 0, \text{ иначе } 0)$$

и линейная функция активации  $g(x) = x$ . Помимо простоты вычислений, объективным важным преимуществом данных функций выпрямителя является то, что они способны выдавать истинное нулевое значение, что может быть очень важным для упрощения модели во избежание переобучения (Goodfellow, Bengio, Courville, 2016). Функции активации

принимают сумму весов от предыдущего слоя и выдают значения для следующего в соответствии с формулой. Также тестировались Softmax, гиперболический тангенс, и сигмоидная функции активации, однако модели с их применением были менее точными.

Отметим, что напрямую сравнивать результаты эконометрического моделирования с использованием регрессионных моделей и нейронных сетей затруднительно. Использование традиционных подходов для выбора наиболее точных моделей, таких как классические информационные критерии, для нейронных сетей не представляется возможным (Anders, Korn, 1999), поскольку в алгоритме нейронных сетей отсутствует собственно функция правдоподобия. Однако некоторые исследователи сравнивают предсказательную способность эконометрических моделей и нейронных сетей с помощью средней квадратичной ошибки (MSE) (Moshiri, Cameron, 2000; Vajracharya, 2011), которая является одной из самых распространенных функций потерь для решения задач такого рода (см, например, (Ширяев, 2013)), но при этом имеет несколько недостатков. В случае нечастых, но больших по модулю ошибок моделей слишком увеличивается вес ошибки, в обратном случае, когда ошибки малы по модулю, но достаточно часты, MSE недооценивает ошибки. В этой связи в работе дополнительно использовались и другие варианты сравнения моделей, среди которых средняя абсолютная ошибка (MAE) и средневзвешенная абсолютная ошибка в процентах (WMAPE).

При моделировании урожайности исходный набор данных, состоящий из 100 наблюдений, был разбит на две части, 75 наблюдений использовались для обучения моделей, а оставшиеся 25 — для проверки их предсказательной способности.

### Полученные результаты

Использование традиционного эконометрического инструментария для прогнозирования урожайности зерновых культур полей сельскохозяйственных регионов представлено в работе (Архипова, Смирнов, 2020). Согласно полученным результатам, преимуществом обладает регрессионная модель по урезанной выборке, которая продемонстрировала более высокие показатели точности и адекватности по сравнению с МНК регрессией.

Рассмотрим результаты, полученные с использованием нейросетевых моделей. Первой была построена модель многослойного перцептрона MLP по всем переменным, так как удале-

ние регрессоров только ухудшало предсказательную способность модели на тестовой выборке. Стандартизация данных проходила отдельно для обучающей и тестовой выборок. Были произведены попытки подбора различных параметров модели для лучшей точности на тестовой выборке. В результате была выбрана структура сети, состоящая из 3 скрытых и одного внешнего слоя по 20, 7, 1 и 1 нейрону соответственно для каждого наблюдения. Методика последовательного включения большего числа нейронов от одного скрытого слоя и регулирования описана в (Neural Network Design, 2014). Для решения поставленной задачи был выбран оптимизатор Adam (Adaptive Moment Optimization), который является комбинацией RMSprop и стохастического градиентного спуска (SGD) с импульсом. Для оптимизатора Adam эмпирически были выбраны гиперпараметры  $\beta_0$  и  $\beta_1$  ( $\beta_0 = 0,9$ ;  $\beta_1 = 0,999$ ). Для каждого веса в модели Adam применяются следующие вычисления:

$$v_t = \beta_1 v_{t-1} - (1 - \beta_1) g_t, \quad s_t = \beta_2 s_{t-1} - (1 - \beta_2) g_t^2, \\ \Delta w_t = -\eta \frac{v_t}{\sqrt{s_t - \epsilon}} g_t, \quad w_{t+1} = w_t + \Delta w_t, \quad (3)$$

где  $\eta$  — начальный коэффициент обучения;  $g_t$  — градиент для  $w_j$ ;  $v_t$  — экспоненциальное скользящее среднее градиентов для  $w_j$ ;  $s_t$  — экспоненциальное скользящее среднее квадратов градиентов для  $w_j$ .

В работе также были рассмотрены следующие варианты оптимизаторов: стохастический градиентный спуск, RMSprop, Adagrad, Adadelata, Adamax, Nadam. Однако модели, построенные с их использованием, уступали по характеристикам модели с оптимизатором Adam, а некоторые из них даже ухудшали предсказательную способность сети. В качестве функции потерь использовалась средняя квадратичная ошибка (MSE) и активационная функция *ReLU*

$$(R(x) = \max(0, x) \rightarrow \frac{\partial f}{\partial x} = 1, \text{ если } x \geq 0, \text{ иначе } 0)$$

и линейная функция активации  $g(x) = x$ .

Помимо простоты вычислений, объективным важным преимуществом данных функций выпрямителя является их способность выдавать истинное нулевое значение, что может быть очень важным для упрощения модели во избежание переобучения (Goodfellow, Bengio, Courville, 2016). Функции активации принимают сумму весов от предыдущего слоя и выдают значения для следующего в соответствии с формулой. Также тестировались

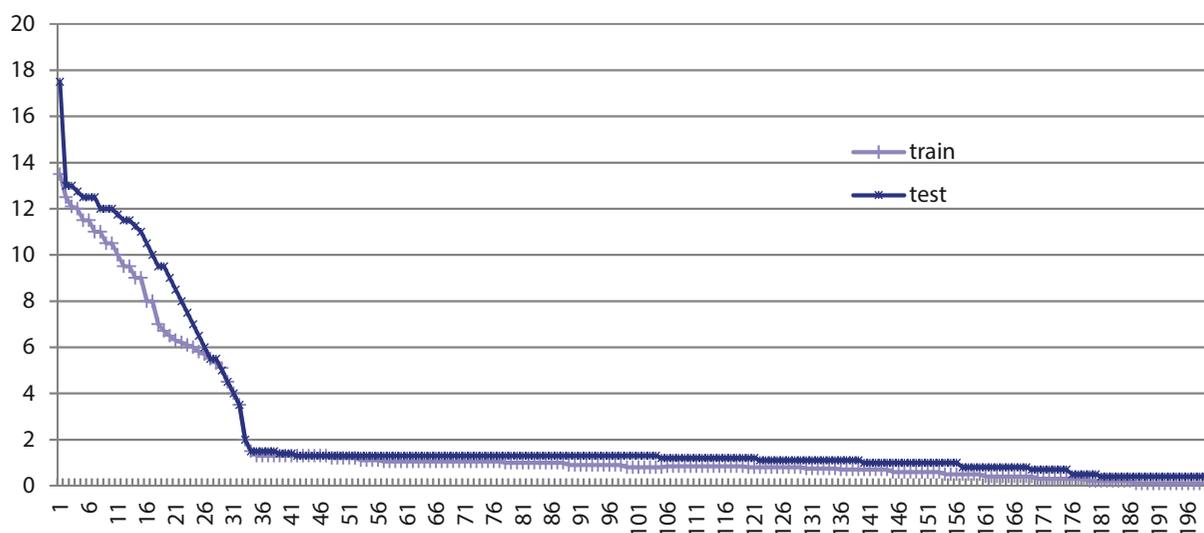


Рис. 2. MSE для обучающей и тестовой выборок модели на смешанных данных (источник: расчеты авторов)

Fig. 2. MSE for training and test sets of the mixed data model

Softmax, гиперболический тангенс, и сигмоидная функции активации, однако модели с их применением были менее точными.

Во второй модели нейронных сетей (CNN) были оставлены те же параметры, что и для модели MLP, кроме количества эпох, которое было выбрано равным сорока, фотографии вегетации полей за июнь и сентябрь были сжаты до одинакового размера 64 на 64 пикселя и соединены в одно изображение. Была выбрана структура сети из четырех скрытых и одного внешнего слоя по 16, 16, 4, 4 и 1 нейрону соответственно для каждого наблюдения.

Для первой составной части модели на табличных значениях была выбрана структура сети, аналогичная модели MLP. Для второй части — модели с обработкой фотографий полей — использовалась структура, выбранная для одинарной модели CNN.

Предсказательная способность модели CNN только по фотографиям вегетации невысока по сравнению с моделью MLP по всему пулу переменных, поэтому было принято решение о соединении моделей и использовании модели на смешанных данных, что позволило получить новую модель с лучшими характеристиками качества по сравнению с моделями MLP и CNN, а также моделью только с константой.

В третьей модели (модель на смешанных данных) было принято решение использовать тот же набор параметров модели, что и для моделей MLP и CNN. Качество модели оценивалось по виду функций потерь.

Функция потерь показывает усредненный квадрат ошибки между наблюдаемыми и модельными значениями зависимой переменной и является одной из самых распространенных

функций потерь в машинном обучении. Она позволяет оценить, насколько хорошо настроилась нейронная сеть в отношении обучающей выборки (*train*) и ожидаемых ответов (*test*).

Сопоставление трех моделей нейронных сетей показало, что предсказательная способность модели по смешанным данным согласно статистике MSE выше, чем с использованием моделей нейросетей MLP или CNN. Средняя квадратичная ошибка MSE для тестовой выборки в модели на смешанных данных составила 0,52, в то время как для моделей MLP и свёрточной сети метрика составила 0,68 и 1,04, соответственно. Данный результат подтверждает выводы, полученные в исследовании (Kung et al., 2016), где ансамбль нейронных сетей сравнивался с отдельными сетями.

На рисунке 2 показано, что если в начале периода для модели на смешанных данных наблюдается высокая ошибка прогноза, то затем MSE резко уменьшается и стремится к нулю, что свидетельствует о хорошем качестве модели.

Данный результат позволил сделать выбор в пользу модели на смешанных данных, которая была использована в качестве итоговой модели для прогнозирования урожайности полей муниципальных образований и проведения дальнейшего анализа.

Отметим, что результат, полученный по модели на смешанных данных, также превосходит по точности лучшую эконометрическую модель по урезанной выборке, для которой значение статистики составило 1,59. Аналогичные результаты получаются и при использовании критерия MAE (*Mean Absolute Error*). Для эконометрической модели значение критерия соста-

Таблица  
Сравнение итоговых эконометрической  
и нейросетевой моделей

Table  
Comparison of final econometric and neural network  
models

Модель	Ошибки модели		
	MSE	MAE	WMAPE
Нейросетевая модель по смешанным данным	0,52	0,54	15,8
Модель по урезанной выборке (CNN)	1,59	0,96	28,1

вило 0,96, а для нейронной сети на смешанных данных — 0,54, что говорит о более высокой точности сети. С помощью статистики *WMAPE* (средневзвешенной абсолютной процентной ошибки прогнозирования) можно сделать такой же вывод: 28,1 % у эконометрической модели против 15,8 % у нейронной сети (табл. 1).

Преимущество нейросетевых моделей по сравнению с традиционным эконометрическим инструментарием можно объяснить недостаточной гибкостью эконометрических моделей, не позволяющих быстро и оперативно проводить корректировку моделей при изменяющихся условиях. Большие временные затраты на настройку параметров новых моделей являются факторами, снижающими эффективность процесса управления.

Отметим, что даже незначительное увеличение точности прогнозных оценок позволяет получить существенный экономический эффект. Это вполне объяснимо, так как более точные прогнозные показатели позволяют сбалансировать управленческие решения, опирающиеся на расчет стоимости хранения зерновых культур, подготовку зернохранилищ, выстраивание логистических цепочек, заключение договоров с подрядными организациями, уменьшить штрафы за сорванные поставки.

Таким образом, не отвергается выдвинутая в начале работы гипотеза  $H_2$  (нейронные сети позволяют получить более точные результаты по сравнению с классическим эконометрическим инструментарием и могут быть использованы для прогнозирования урожайности полей муниципальных образований).

Рассмотрим выводы, полученные на основе используемых интерпретаторов для объяснения влияния факторов для каждого конкретного предсказанного значения. В качестве примера рассмотрим работу предложенного алгоритма для Красноярского района Самарской области. Интерпретатор *LIME* для нейронной сети *MLP* по табличной информации показал, что самым важным фактором при моделировании уро-

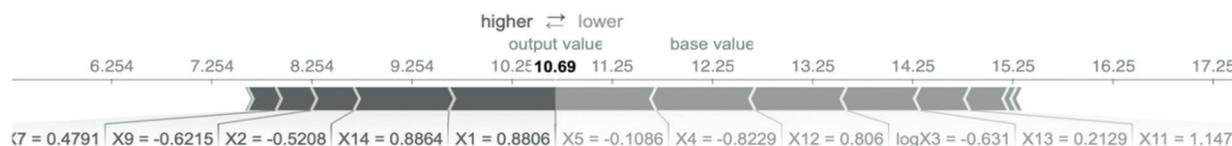
жайности МО сельскохозяйственных территорий является показатель  $X_2$  (доля посевной площади от площади МО), значение данного фактора превышает среднее значение по выборке на 30 %. Второй по важности признак, который также положительно влияет на урожайность — доля растениеводства в сельском хозяйстве МО, которая на 28 % больше среднего. Третий признак — наличие поселков городского типа: Волжский, Мирный и Новосемейкино, чье население относится к городскому. Наличие городского населения, как и ожидалось, увеличивает вероятность высокой урожайности сельскохозяйственных культур.

Интерпретатор *SHAP* для табличных данных для Красноярского района Самарской области подтвердил лидерование влияния показателя  $X_2$  на урожайность. Однако на второе место попал фактор внесения минеральных удобрений, значение которого которое на 22 % ниже среднего уровня. Данный вывод может говорить о возможном недостатке питательных веществ почвы. На третьем месте с положительным влиянием на урожайность — фактор наличия поселков городского типа. На четвертом месте — обеспеченность тракторами, что совпадает с выводами интерпретатора *LIME*.

Новый вывод, который был менее заметен на сводке предыдущего интерпретатора, состоит в том, что фактор низкого качества водных ресурсов негативно влияет на урожайность. Процент 4-го и 5-го классов загрязненной воды на 63 % выше среднего по выборке, что говорит о возможных проблемах водных ресурсов, используемых в сельском хозяйстве. Еще один негативный эффект выявлен у фактора заработной платы сельскохозяйственных работников, который меньше среднего на 14 %, что может говорить о недостаточной экономической мотивации трудоустроенных в агрокомплексе района.

Вывод интерпретатора *SHAP* для всех факторов для данного района представлен на рисунке 3. На темных участках изображены факторы, повышающие прогнозное значение урожайности, на светлых — понижающие.

Согласно рисунку, наибольшее положительное влияние на прогнозное значение урожайности имеют показатели  $X_1$  и  $X_{14}$ , связанные с долей растениеводства в сельском хозяйстве муниципального образования и экологической ситуацией в городе (им соответствуют большие прямоугольники на рисунке 3, закрашенные в темный цвет). Негативное влияние — у показателей  $X_4$ ,  $X_5$  и  $X_{12}$ , что может быть объяснено устаревшей материально-технической базой



**Рис. 3.** Вывод LIME для всех факторов для табличной информации о Красноярском районе Самарской области (источник: расчеты авторов)

**Fig. 3.** LIME results for all table factors concerning Krasnoyarsky district of Samara oblast

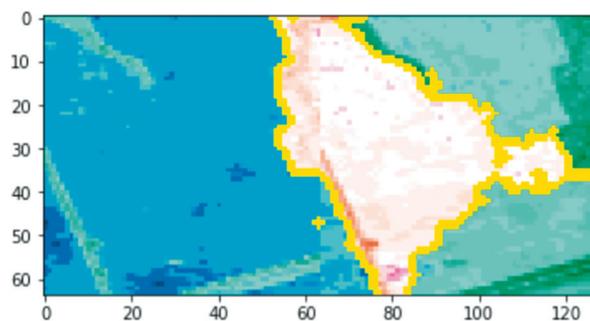
и неэффективностью ведения сельского хозяйства на территориях с высоким уровнем загрязнения вод (им соответствуют большие серые прямоугольники).

В рамках интерпретатора LIME для сверточной сети CNN фотография поля была разбита на 5 суперпикселей, представляющих крупные характерные участки поля, по виду которых модель предсказывает урожайность. Для случайно выбранного Дивеевского района Нижегородской области фактором высокого прогнозного значения урожая, которое выше среднего на 10 %, стала фотография начала осени, на которой видно, что поле убрано после урожая и достаточно однородно. Снижает прогноз фактор фотографии весны, на котором видна относительно невысокая урожайность. Описанные эффекты видны на рисунке 4. Темные тона показывают те участки поля, которые положительно влияют на урожайность, а белым и светло-серым цветами представлен сектор поля, который характеризуется низкой урожайностью для весеннего периода (низкой всхожимостью культуры).

Отличие метода SHAP от метода LIME заключается в том, что данный метод оценивает каждый пиксель изображения, а не суперпиксели, как это делает LIME. Светло-серые точки на правой части иллюстрации сообщают о тех участках поля, которые уменьшают прогнозное значение урожайности, темно-серые — наоборот (рис. 5). Как можно заметить, фактор небольшой реки в северо-западной части поля положительно влияет на урожайность, в то время как небольшой выходящий на поле в юго-западной части лес негативно сказывается на урожайности.

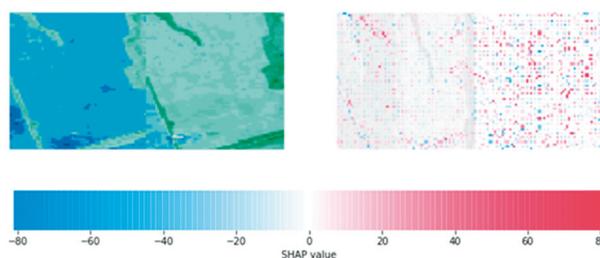
Однако следует отметить, что в целом выводы интерпретаторов не противоречат друг другу: наличие небольшой реки улучшает урожайность поля, присутствие лесного массива площадью 1900 м<sup>2</sup> в месте, препятствующему движению комбайнов, тракторов и прочей техники, негативно сказываются на урожайности.

Важным результатом моделирования является статистическая значимость фактора спутниковых фотографий полей для повышения точности прогнозных значений урожайности (гипотеза H<sub>1</sub>).



**Рис. 4.** LIME для фотографии поля Дивеевского района Нижегородской области (источник: расчеты авторов в Python)

**Fig. 4.** LIME for a photograph of a field in Diveyevsky district of Nizhny Novgorod oblast



**Рис. 5.** SHAP для фотографии поля Дивеевского района Нижегородской области (источник: расчеты авторов в Python)

**Fig. 5.** SHAP for a photograph of a field in Diveyevsky district of Nizhny Novgorod oblast

### Заключение

В результате исследования были получены ответы на основные исследовательские вопросы, которые были поставлены перед проведением исследования. Так, было статистически доказано, что спутниковые фотоснимки являются статистически значимым фактором для прогнозирования урожайности полей муниципальных образований (гипотеза H<sub>1</sub>) и повышают точность прогнозных моделей при их использовании (гипотеза H<sub>2</sub>).

Сопоставление эконометрических моделей и моделей нейронных сетей позволило сделать выбор в пользу последних, которые показали лучшие результаты при прогнозировании урожайности сельскохозяйственных культур для муниципальных образований сельскохозяйственных регионов. Было статистически показано, что нейронные сети могут эффективно решать задачу регрессии, в том числе

и на малой выборке наблюдений (гипотеза  $H_2$ ). Ошибки  $MSE$  на тестовой выборке для нейронной сети по смешанным данным оказалась меньше, чем для эконометрических моделей.

Что касается экологических факторов (гипотеза  $H_3$ ), то было показано, что они в разной степени влияют на урожайность. Так, если факторы загрязнения воздуха и почвы оказались статистически незначимыми, то фактор, отвечающий за эффективность сбора и вывоза бытового мусора, показал положительное влияние на урожайность.

Согласно интерпретатору *SHAP* (и менее выражено в *LIME*), качество водных ресурсов также является важным фактором, оказывающим влияние на урожайность полей наравне со сбором и вывозом бытового мусора.

Кроме того, получен и ряд дополнительных выводов. Так, например, статистически доказано, что высокая урожайность полей в растениеводстве во многом связана со специализацией района и большими возможностями диверсификации труда и всех процессов земледелия, хранения, транспортировки и продажи урожая. Высокая оснащенность МО тракторами и иными сельскохозяйственными машинами, и комбайнами (на гектар засеянной площади) также оказывает позитивное воздействие на результирующий признак. Связано это с инновационным развитием региона и инвестициями бизнеса в агрокультурные предприятия муниципального образования. Негативное воз-

действие на урожайность оказал фактор чрезмерно высокой доли засеянных территории МО. Объяснение данного факта состоит в том, что для повышения эффективности сельскохозяйственного производства недостаточно увеличивать площадь полей (экстенсивный фактор), необходимо развивать всю инфраструктуру сельских и городских территорий.

Также в процессе исследования было показано преимущество модели по смешанным данным для прогнозирования урожайности полей по сравнению с моделями *MLP* и *CNN* в отдельности.

К перспективам исследования относятся обучение и проверка моделей на большей выборке как табличных данных, так и спутниковых фотоснимков. Также интерес представляет расширение предсказательной способности нейронных сетей в виде настройки параметров за счет включения новых переменных, таких как тип возделываемой культуры, информация о минерализации и состоянии отдельных полей. Использование данных переменных позволит непосредственно проводить планирование посадок и севооборота. Вероятно, автоматизация сбора, хранения и обработки данных о полях и МО в едином реестре с последующим анализом урожайности полей и районов в целом потенциально интересна для частных фермеров, агрохолдингов, инвесторов в сельское хозяйство России и регулирующих государственных органов.

### Список источников

- Анализ данных: учебник для академического бакалавриата. / В. С. Мхитарян [и др.]. Москва : Юрайт, 2017. 490 с.
- Архипова М. Ю., Смирнов А. И. Современные направления прогнозирования урожайности сельскохозяйственных культур на основе использования эконометрических моделей // Вопросы статистики. 2020. № 27 (4). С. 88–101
- Власюк П. А. Биологические элементы в жизнедеятельности растений. Киев : Наукова думка, 1969. 516 с.
- Осовский С. Нейронные сети для обработки информации. Москва : Финансы и статистика, 2004. 343 с.
- Перечень предельно допустимых концентраций и ориентировочно безопасных уровней воздействия вредных веществ для воды рыбохозяйственных водоемов / С. Н. Анисова, С. А. Соколова, А. Т. Лебедев. Москва : Мединор, 1995. 220 с.
- Перечень рыбохозяйственных нормативов предельно-допустимых концентраций (ПДК) и ориентировочно безопасных уровней воздействия (ОБУВ) вредных веществ для воды водных объектов, имеющих рыбохозяйственное значение / Н. А. Шиленко и др. Москва : ВНИРО, 1999. С. 28–249.
- Ширяев В. И. Финансовые рынки. Нейронные сети, хаос и нелинейная динамика: Учебное пособие. Москва : Либроком, 2013. 232 с.
- Anders U., Korn O. Model selection in neural networks // Neural networks. 1999. Vol. 12, No. 2. P. 309–323. DOI: doi.org/10.1016/S0893-6080(98)00117-8.
- Bajracharya D. Econometric Modeling vs Artificial Neural Networks: A Sales Forecasting Comparison. 2011. URL: <http://docplayer.net/14695913-Econometric-modeling-vs-artificial-neural.html> (date accessed: 19.05.2020).
- Evaluating environmental sensitivity at the basin scale through the use of geographic information systems and remotely sensed data: an example covering the Agri basin (Southern Italy) / F. Basso, et al. // Catena. 2000. Vol. 40, No. 1. P. 19–35. DOI: doi.org/10.1016/S0341-8162(99)00062-4.
- Soybean crop coverage estimation from NDVI images with different spatial resolution to evaluate yield variability in a plot / A. De la Casa, et al. // ISPRS journal of photogrammetry and remote sensing. 2018. Vol. 146. P. 531–547. DOI: doi.org/10.1016/j.isprsjprs.2018.10.018.

Dharmadhikari N. L. Economic Modeling of Agricultural Production in North Dakota Using Transportation Analysis and Forecasting. URL: [https://library.ndsu.edu/ir/bitstream/handle/10365/29017/Dharmadhikari\\_ndsu\\_0157D\\_12238.pdf?sequence=4&isAllowed=y](https://library.ndsu.edu/ir/bitstream/handle/10365/29017/Dharmadhikari_ndsu_0157D_12238.pdf?sequence=4&isAllowed=y).

Goodfellow I., Bengio Y., Courville A. Deep learning // MIT press. 2016. URL: <http://www.deeplearningbook.org/> (date accessed: 19.05.2020).

Haghverdi A., Washington-Allen R. A., Leib B. G. Prediction of cotton lint yield from phenology of crop indices using artificial neural networks // Computers and Electronics in Agriculture. 2018. Vol. 152. P. 186–197. DOI: doi.org/10.1016/j.compag.2018.07.021.

Jordanova N. Soil magnetism: Applications in pedology, environmental science and agriculture. Academic Press, 2016. 466 p.

On large-batch training for deep learning: Generalization gap and sharp minima / N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, P. T. P. Tang // arXiv preprint arXiv:1609.04836. 2016. URL: <https://arxiv.org/pdf/1609.04836.pdf>

Accuracy analysis mechanism for agriculture data using the ensemble neural network method / H. Y. Kung, et al. // Sustainability. 2016. Vol. 8, No. 8. P. 735. DOI: doi.org/10.3390/su8080735.

Masters D., Luschi C. Revisiting small batch training for deep neural networks // arXiv preprint arXiv:1804.07612. 2018. P. 29–43

Molnar C. Interpretable Machine Learning-A Guide for Making Black Box Models Explainable. URL: <https://christophm.github.io/interpretable-ml-book/interpretable-ml.pdf>.

Moshiri S., Cameron N. Neural network versus econometric models in forecasting inflation // Journal of forecasting. 2000. Vol. 19. No.3. P. 201–217. DOI: doi.org/10.1002/(SICI)1099-131X(200004)19:33.3.CO;2-W.

Neural Network Design / H. B. Demuth, M. H. Beale, D. J. Orlando, M. Hagan. URL: <https://hagan.okstate.edu/NNDesign.pdf>.

Wheat yield prediction using machine learning and advanced sensing techniques / X. E. Pantazi, et al. // Computers and Electronics in Agriculture. 2016. Vol. 121. P. 57–65. DOI: doi.org/10.1016/j.compag.2015.11.018.

Pöldaru R., Roots J., Viira A. H. Estimating econometric model of average total milk cost: A support vector machine regression approach // Economics and rural development. 2005. Vol. 1, No. 1. P. 23–31.

Irrigated pinto bean crop stress and yield assessment using ground based low altitude remote sensing technology / R. Ranjan, et al. // Information Processing in Agriculture. 2019. Vol. 6, No. 4. P. 502–514. DOI: doi.org/10.1016/j.inpa.2019.01.005.

Exploring the relationship between agricultural productivity and land degradation in a dry region of Southern Europe / L. Salvati, et al. // New Medit. 2010. Vol. 9, No. 1. P. 35–40.

Machine-learned prediction of annual crop planting in the US Corn Belt based on historical crop planting maps / C. Zhang, et al. // Computers and Electronics in Agriculture. 2019. Vol. 166. P. 104989. DOI: doi.org/10.1016/j.compag.2019.104989.

Zhang L., Lei L., Yan D. Comparison of two regression models for predicting crop yield // 2010 IEEE International Geoscience and Remote Sensing Symposium. Ieee, 2010. P. 1521–1524.

## References

Anders, U. & Korn, O. (1999). Model selection in neural networks. *Neural networks*, 12(2), 309–323. DOI: 10.1016/S0893-6080(98)00117-8.

Anisova, S. N., Sokolova, S. A. & Lebedev, A. T. (1995). *Perechen predelno dopustimyykh kontsentratsiy i orientirovochno bezopasnykh urovney vozdeystviya vrednykh veshchestv dlya vody rybokhozyaystvennykh vodoemov [List of maximum permissible concentrations and approximately safe impact levels of hazardous substances for water of fishery water bodies]*. Moscow: TOO “Medinor”, 220. (In Russ.)

Arkhipova, M. Yu. & Smirnov, A. I. (2020). Current Trends in Crop Yield Forecasting Based on the Use of Econometric Models. *Voprosy statistiki*, 27(4), 88–101. (In Russ.)

Bajracharya, D. (2011). *Econometric Modeling vs Artificial Neural Networks: A Sales Forecasting Comparison*. Retrieved from: <http://docplayer.net/14695913-Econometric-modeling-vs-artificial-neural.html> (Date of access: 19.05.2020).

Basso, F., Bove, E., Dumontet, S., Ferrara, A., Pisante, M., Quaranta, G. & Taberner, M. (2000). Evaluating environmental sensitivity at the basin scale through the use of geographic information systems and remotely sensed data: an example covering the Agri basin (Southern Italy). *Catena*, 40(1), 19–35. DOI: 10.1016/S0341-8162(99)00062-4.

de la Casa, A., Ovando, G., Bressanini, L., Martinez, J., Diaz, G. & Miranda, C. (2018). Soybean crop coverage estimation from NDVI images with different spatial resolution to evaluate yield variability in a plot. *ISPRS journal of photogrammetry and remote sensing*, 146, 531–547. DOI: 10.1016/j.isprsjprs.2018.10.018.

Demuth, H. B., Beale, M. H., De Jess, O. & Hagan, M. T. (2014). *Neural network design*. Martin Hagan. Retrieved from: <https://hagan.okstate.edu/NNDesign.pdf>.

Dharmadhikari, N. L. (2018). *Economic Modeling of Agricultural Production in North Dakota Using Transportation Analysis and Forecasting. Doctoral dissertation*. North Dakota State University. Retrieved from: [https://library.ndsu.edu/ir/bitstream/handle/10365/29017/Dharmadhikari\\_ndsu\\_0157D\\_12238.pdf?sequence=4&isAllowed=y](https://library.ndsu.edu/ir/bitstream/handle/10365/29017/Dharmadhikari_ndsu_0157D_12238.pdf?sequence=4&isAllowed=y).

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from: <http://www.deeplearningbook.org/> (Date of access: 19.05.2020).

- Haghverdi, A., Washington-Allen, R. A. & Leib, B. G. (2018). Prediction of cotton lint yield from phenology of crop indices using artificial neural networks. *Computers and Electronics in Agriculture*, 152, 186–197. DOI: 10.1016/j.compag.2018.07.021.
- Jordanova, N. (2016). *Soil magnetism: Applications in pedology, environmental science and agriculture*. Academic Press, 466.
- Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M. & Tang, P. T. P. (2016). *On large-batch training for deep learning: Generalization gap and sharp minima*. arXiv preprint arXiv:1609.04836. Retrieved from: <https://arxiv.org/pdf/1609.04836.pdf>
- Kung, H. Y., Kuo, T. H., Chen, C. H. & Tsai, P. Y. (2016). Accuracy analysis mechanism for agriculture data using the ensemble neural network method. *Sustainability*, 8(8), 735. DOI: 10.3390/su8080735.
- Masters, D. & Luschi, C. (2018). *Revisiting small batch training for deep neural networks*. Pp. 29–43. arXiv preprint arXiv:1804.07612.
- Mkhitarian, V. S. (2017). *Analiz dannykh: uchebnik dlya akademicheskogo bakalavriata [Data Analysis: An Academic Undergraduate Textbook]*. Moscow: Urait, 490. (In Russ.)
- Molnar C. (2018). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Retrieved from: <https://christophm.github.io/interpretable-ml-book/interpretable-ml.pdf>.
- Moshiri, S. & Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of forecasting*, 19(3), 201–217. DOI: 10.1002/(SICI)1099-131X(200004)19:3<201::AID-FORE2019>3.0.CO;2-W.
- Osovskiy, S. (2004). *Neyronnye seti dlya obrabotki informatsii [Neural networks for data processing]*. Moscow: Finansy i statistika, 343. (In Russ.)
- Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L. & Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*, 121, 57–65. DOI: 10.1016/j.compag.2015.11.018.
- Pöldaru, R., Roots, J. & Viira, A. H. (2005). Estimating econometric model of average total milk cost: A support vector machine regression approach. *Economics and rural development*, 1(1), 23–31.
- Ranjan, R., Chandel, A. K., Khot, L. R., Bahlol, H. Y., Zhou, J., Boydston, R. A. & Miklas, P. N. (2019). Irrigated pinto bean crop stress and yield assessment using ground based low altitude remote sensing technology. *Information Processing in Agriculture*, 6(4), 502–514. DOI: 10.1016/j.inpa.2019.01.005.
- Salvati, L. (2010). Exploring the relationship between agricultural productivity and land degradation in a dry region of Southern Europe. *New Medit*, 9(1), 35–40.
- Shilenko, N. A., Sokolova, S. A., Anisova, S. N., Lesnikov, L. A., Lebedev, A. T. & Semionova, I. V. (1999). *Perechen rybokhozyaystvennykh normativov predelno-dopustimyykh kontsentratsiy (PDK) i orientirovochno bezopasnykh urovney vozdeystviya (OBUV) vrednykh veshchestv dlya vody vodnykh obektov, imeyushchikh rybokhozyaystvennoe znachenie [The List of Fishery Standards: Maximum Permissible Concentrations (MPC) and Tentatively Safe Levels of Exposure (SLE) of Harmful Substances to Water of Water Bodies of Fishery Importance]*. M.: VNIRO, 304. (In Russ.)
- Shiryaev, V. I. (2013). *Finansovye rynki. Neyronnye seti, khaos i nelineynaya dinamika: Uchebnoe posobie [Financial Markets: Neural Networks, Chaos, and Nonlinear Dynamics]*. Moscow: Librocom, 232. (In Russ.)
- Vlasyuk, P. A. (1969). *Biologicheskie elementy v zhiznedeyatel'nosti rasteniy [Biological Elements in the Life Activity of Plants]*. Kiev: Naukova dumka, 516. (In Russ.)
- Zhang, C., Di, L., Lin, L. & Guo, L. (2019). Machine-learned prediction of annual crop planting in the US Corn Belt based on historical crop planting maps. *Computers and Electronics in Agriculture*, 166, 104989. DOI: 10.1016/j.compag.2019.104989.
- Zhang, L., Ji, L., Lei, L. & Yan, D. (2010). Comparison of two regression models for predicting crop yield. In: *2010 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1521–1524). IEEE.

### Информация об авторе

**Архипова Марина Юрьевна** — доктор экономических наук, профессор, Национальный исследовательский университет «Высшая школа экономики»; Scopus Author ID: 57191839300; <https://orcid.org/0000-0002-9022-7385> (Российская Федерация, 101000, г. Москва, ул. Мясницкая, д. 20; 121352, г. Москва, ул. Мясницкая, д. 20; e-mail: marhipova@hse.ru).

### About the author

**Marina Yu. Arkhipova** — Dr. Sci. (Econ.), Professor, National Research University “Higher School of Economics”; Scopus Author ID: 57191839300; <http://orcid.org/0000-0002-9022-7385> (20, Myasnitskaya St., Moscow, 101000, Russian Federation; e-mail: marhipova@hse.ru).

Дата поступления рукописи: 05.07.2020.

Прошла рецензирование: 17.09.2020.

Принято решение о публикации: 07.04.2022.

Received: 05 Jul 2020.

Reviewed: 17 Sep 2020.

Accepted: 07 Apr 2022.