

МОДЕЛИ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ЦИФРОВОЙ ДИАГНОСТИКИ БОЛЕЗНЕЙ ЗЕРНОВЫХ КУЛЬТУР (НА ПРИМЕРЕ *PYRENOPHORA TERES* ОЗИМОГО ЯЧМЕНЯ)

И. В. Ариничев¹, кандидат экономических наук, **И. В. Ариничева**², доктор биологических наук,
Г. В. Волкова³, член-корреспондент РАН, доктор биологических наук,
Я. В. Яхник³, аспирант

¹Кубанский государственный университет,
350040, Краснодар, ул. Ставропольская, 149

²Кубанский государственный аграрный университет им. И. Т. Трубилина,
350044, Краснодар, ул. Калинина, 13

³Федеральный научный центр биологической защиты растений,
350039, Краснодар, п/о 39
E-mail: galvol.bpp@yandex.ru

Исследования проводили с целью обоснования возможности использования цифровых интеллектуальных технологий при прогнозировании развития сетчатой пятнистости озимого ячменя. Предложенный подход представляет собой бинарное дерево решений, способное прогнозировать сценарии развития сетчатой пятнистости: депрессивное, умеренное и эпифитотийное. Для настройки параметров алгоритма в 2021–2023 гг. были проведены полевые и лабораторные опыты по подготовке данных. В качестве входных факторов, фиксируемых в выборке, использовали наблюдаемую степень пораженности листьев, тип устойчивости сорта, фазу вегетации, в которую произошло первичное заражение, среднюю относительную влажность воздуха за фазу вегетации, в которой произошло заражение. Общий объем выборки составил 144 наблюдения. Обученная модель показала высокое качество классификации как на обучающей, так и на тестовой выборке с долей правильных ответов более 96%. Согласно результатам статистической оценки важности факторов, влияющих на развитие сетчатой пятнистости ячменя, наибольшим весом обладает текущая пораженность листьев (74,3%), второй по важности фактор – средняя относительная влажность воздуха (11,9%), далее следуют устойчивость сорта к болезни (10,4%) и фаза вегетации, в которую произошло заражение (3,4%). Предложенное решение имеет важное практическое значение, так как предоставляет новые возможности для процесса диагностики сетчатой пятнистости озимого ячменя, среди которых: высокая скорость диагностики, точность составления прогнозов, применимость в полевых условиях.

AI SOLUTIONS FOR DIGITAL DIAGNOSTICS OF DISEASES OF GRAIN CROPS (USING THE EXAMPLE OF *PYRENOPHORA TERES* OF WINTER BARLEY)

I. V. Arinichev¹, **I. V. Arinicheva**², **G. V. Volkova**³, **Y. V. Yakhnik**³

¹Kuban State University,
350040, Krasnodar, ul. Stavropol'skaya, 149

²Trubilin Kuban State Agrarian University,
350040, Krasnodar, ul. Kalinina, 13

³Federal Research Center of Biological Plant Protection,
350039, Krasnodar, p/o 39
E-mail: galvol.bpp@yandex.ru

The aim of the research is to justify the feasibility of using digital intelligent technologies in forecasting the development of spot blotch in winter barley. The AI solution developed is a binary decision tree capable of predicting scenarios of net blotch development: depressive, moderate, and epiphytotic. To configure the algorithm's parameters from 2021 to 2023, field and laboratory experiments were conducted at the Federal Scientific Center for Biological Plant Protection. The data preparation involved several stages, including setting up field plots to create an artificial infection background, preparing an inoculum, sowing highly susceptible and resistant winter barley varieties, and artificial inoculation. The selected input factors included the observed degree of leaf damage, type of variety resistance, vegetation phase at the time of primary infection, average relative air humidity during the vegetation phase of infection. The total sample size comprised 144 observations. The trained model demonstrated high classification accuracy on both the training and test datasets, with an accuracy rate exceeding 96%. Based on a statistical assessment of the factors influencing the development of spot blotch in barley, it is shown that the most influential factor is the current degree of leaf infection (74,3%), followed by the average relative air humidity (11,9%), the variety's resistance to the disease (10,4%), and the stage development during which infection occurred (3,4%). The proposed solution holds significant practical importance as it provides new opportunities for the diagnostic process of spot blotch in winter barley. Among these are high diagnostic speed, accuracy in forecast predictions, and applicability in field conditions.

Ключевые слова: зерновое производство, озимый ячмень (*Hordeum vulgare* L.), сетчатая пятнистость (*Pyrenophora teres*), цифровая диагностика, искусственный интеллект.

Key words: grain production, winter barley (*Hordeum vulgare* L.), spot blotch (*Pyrenophora teres*), digital diagnostics, artificial intelligence.

Зерновое производство – наиболее крупная подотрасль сельского хозяйства, устойчивое развитие которой играет критическую роль в достижении продовольственной безопасности страны и имеет страте-

гическое значение в решении ряда задач, среди которых обеспечение населения страны безопасной, качественной и доступной сельскохозяйственной продукцией, создание рабочих мест в сельских районах, развитие

аграрной инфраструктуры, укрепление экономического положения страны на мировых рынках и др. Зерновые культуры, такие как пшеница, ячмень, рис и кукуруза, служат фундаментом мирового продовольственного рынка, обеспечивая стабильный поток продуктов как для внутреннего потребления, так и для экспорта.

Несмотря на устойчивый прирост валового сбора зерна за последние 5 лет (со 113,3 млн т в 2018 г. до 157,7 млн т в 2022 г.) [1] и уверенное превышение порогового значения доли зерна отечественного производства, определенного Доктриной продовольственной безопасности Российской Федерации на уровне 95 % (в 2022 г. уровень самообеспеченности превысил 150%) [2], перед отраслью возникают новые все более серьезные вызовы, самый значимый из которых – угроза фитосанитарному состоянию посевов.

Среди основных факторов массового распространения болезней и вредителей на посевах зерновых решающее значение имеют следующие:

ухудшение погодно-климатических условий, включая повышение среднегодовых температур на большей части зернопроизводящих регионов, непредвиденные и продолжительные осадки в более поздние сроки вегетации, экстремальные погодные явления (засухи, наводнения и аномальные температуры), которые способствуют распространению болезней и вредителей;

частое и несвоевременное использование химических средств защиты растений, которое приводит к развитию резистентности болезнетворных организмов к применяемым препаратам, делает обычные методы борьбы менее эффективными и требует более интегрированных подходов;

монокультура, способствующая накоплению патогенов и вредителей в почве, что увеличивает риск распространения болезней и усугубляет проблемы с фитосанитарным состоянием;

недостаточное внимание к системе мониторинговых мероприятий, осуществляемых в рамках производственной деятельности и имеющих решающее значение для обнаружения, прогнозирования и контроля фитосанитарного состояния посевов.

Не вызывает сомнений, что решение одной из главных проблем зернового производства – массового распространения болезней и вредителей – невозможно без перехода к инновационным методам ведения производства. Сегодня на передний план всех инновационных решений выходит искусственный интеллект, определяющий ядро таких инноваций и обладающий огромным потенциалом для цифровой трансформации.

Освоение интеллектуальных решений открывает новые возможности для повышения устойчивости, снижения зависимости и минимизации ручного труда и, следовательно, увеличения эффективности всей отрасли, что способствует решению сложных вызовов и обеспечению продовольственной безопасности хозяйства, региона, страны [3].

Ранее была проведена интеллектуальная диагностика сетчатой пятнистости ячменя, вызываемой грибом *Pyrenophora teres* Drechsler, одним из доминантных патогенов в ценозе культуры, как на юге России, так и во всем мире. Ежегодные потери урожая от его воздействия составляют от 15 до 50%, а частота эпифитотий – 5 раз в 10 лет. При благоприятном сочетании факторов (погода, восприимчивость сорта, фаза вегетации) распространенность болезни может достигать до 100%, развитие – 50...90% [4, 5].

Актуальность цифровой диагностики сетчатой пятнистости обусловлена малой эффективностью

классических методов борьбы с этим заболеванием, включающих агротехнологические приемы, протравливание семян, возделывание устойчивых сортов и др. Гарантированное сохранение урожая можно обеспечить только путем применения эффективных фунгицидов. При этом контрольной точкой в процессе производства служит принятие решения о целесообразности их использования в конкретный заданный промежуток времени.

Для поддержки принятия таких решений ранее были разработаны модели на основе искусственного интеллекта (ИИ-модели), с использованием которых были решены задачи обнаружения и классификации *Pyrenophora teres*, по отношению к другим видам болезней со схожими симптомами [6], а также локализации пораженных участков и определения степени пораженности листьев [7], что служит одним из основных сигналов для применения химических средств защиты.

Однако, наряду с перечисленными задачами, где ИИ-решения уже показали высокие результаты, как вследствие сокращения времени диагностики, так и благодаря росту доли числа точных прогнозов, высокой практической ценности и актуальностью обладают вопросы, связанные с прогнозированием развития болезни. Важно не просто диагностировать заболевание и констатировать степень пораженности растений на текущий момент, но и составить прогноз течения болезни.

Цель исследования – обосновать возможность использования цифровых интеллектуальных технологий в процессе прогнозирования развития сетчатой пятнистости озимого ячменя.

Методика. Для достижения поставленной цели в 2021–2023 гг. на площадках ФГБНУ ФНЦБЗР были выполнены полевые и лабораторные исследования, предусматривающие заражение растений популяцией *Pyrenophora teres*; мониторинг первичного проявления и динамики развития. В исследованиях использовали 3 сорта озимого ячменя (Виват, Рубеж, Романс), высеваемых на юге России и отличающихся по устойчивости к возбудителю сетчатой пятнистости (устойчивые и восприимчивые). При проведении экспериментов использовали классические фитопатологические методы и подходы. Идентификацию *P. teres* осуществляли по определителю Билай В. И. [8]. Нарработку инокулюма гриба и инокуляцию в фазе полного кушения в полевых условиях осуществляли по стандартным методикам [9]. Учеты проводили начиная с первичного проявления заболевания и до фазы молочно-восковой спелости зерна с интервалом 10...12 суток. Степень пораженности листьев и других органов сетчатой пятнистостью определяли по шкале Э. Э. Гешеле.

В качестве входных факторов модели, фиксируемых в процессе реализации опыта и влияющих на степень развития сетчатой пятнистости, использовали наблюдаемую степень пораженности листьев (%); тип устойчивости сорта (У = устойчивый, В = восприимчивый); фаза вегетации, в которую произошло первичное заражение (кушение, выход в трубку, флаг-лист); средняя относительная влажность воздуха за фазу вегетации, в которой произошло заражение (%).

Выходная целевая переменная – степень развития болезни в фазе ранней молочной спелости со следующими возможными значениями: Д – депрессия, У – умеренное развитие, Э – эпифитотия. Метки классов результирующей переменной были выбраны на основе работы [10], авторы которой прогнозировали фитосанитарное состояние посевов пшеницы.

Объем выборки в общей сложности составил 144 наблюдения, соответствующих различным комбинациям искомых признаков. Из них случайным образом 115 объектов использовали в качестве тренировочного набора и 29 – для итоговой оценки качества модели. Разбиение выборки на обучающую и тестовую – важный этап методологии машинного обучения и прикладной статистики, цель которого заключается в оценке производительности модели и ее обобщающей способности. Основная идея, стоящая за таким разбиением, «спрятать» от алгоритма часть данных, которые используют при настройке параметров модели, чтобы на нужном этапе выполнить на них проверку качества. Если модель может делать точные прогнозы на тестовой части данных, которые ранее не подавались на вход, то с ее использованием возможно обобщение и применение знаний, полученных в процессе обучения, к новым данным.

В процессе анализа пригодности различных моделей для прогнозирования развития сетчатой пятнистости озимого ячменя в качестве базового было выбрано семейство алгоритмов деревьев решений. Обучение итоговой модели, детерминирующей объекты выборки по разным классам развития сетчатой пятнистости, проводили в соответствии с общепринятыми методами машинного обучения (стохастического градиентного спуска, обратного распространения ошибки и др.).

Результаты и обсуждение. Выбор деревьев решений для прогнозирования развития сетчатой пятнистости озимого ячменя обоснован несколькими ключевыми причинами:

древовидные алгоритмы предоставляют четкие и интерпретируемые результаты. Это особенно важно для сельского хозяйства, где товаропроизводители должны составлять прогнозы без использования специальных программных средств. При использовании выбранного алгоритма достаточно пройти из корневой вершины к одной из конечных, проверяя соблюдение условий в узлах модели, и получить прогноз развития болезни;

применение дерева решений дает возможность относительно быстро обрабатывать как небольшие, так и огромные объемы данных, что позволяет в конечном итоге оперативно выполнять предиктивную функцию;

деревья решений могут масштабироваться под новые условия и требования задачи, что делает их пригодными для разнообразных ситуаций, в частности, для других культур и их патогенов;

модели на основе деревьев решений могут работать с категориальными признаками, не имеющими выражения в числовой шкале, что особенно ценно для аграрного производства, где качественные факторы часто выступают предикторами. В нашем случае к типу категориальных относились два из четырех факторов.

При построении дерева решения использовали рекурсивную процедуру машинного обучения *Induction of Decision 3* (ID3). Наиболее важным ее шагом выступает поиск предиката (бинарной функции), который ставится в соответствие каждой внутренней вершине дерева и разделяет в ней выборку на две части. Он сводится к оптимизации информационного критерия ветвления – специальной метрики, помогающей алгоритму определить, какой признак лучше использовать для разделения данных на каждом уровне дерева.

В работе использовали наиболее распространенный и часто применяемый критерий Джини, показывающий сколько пар объектов, лежащих в одном и том же классе, одновременно попадут либо в левую, либо в правую дочернюю вершину дерева (значения предиката на них совпадают):

$$I(\beta, X^l) = \#\{(x_i, x_j) : y_i = y_j, \beta(x_i) = \beta(x_j)\},$$

где X^l – обучающая выборка, x_i, x_j – признаки (факторы) объектов обучающей выборки, β – предикат, $\#$ – оператор, подсчитывающий количество пар, удовлетворяющих условию.

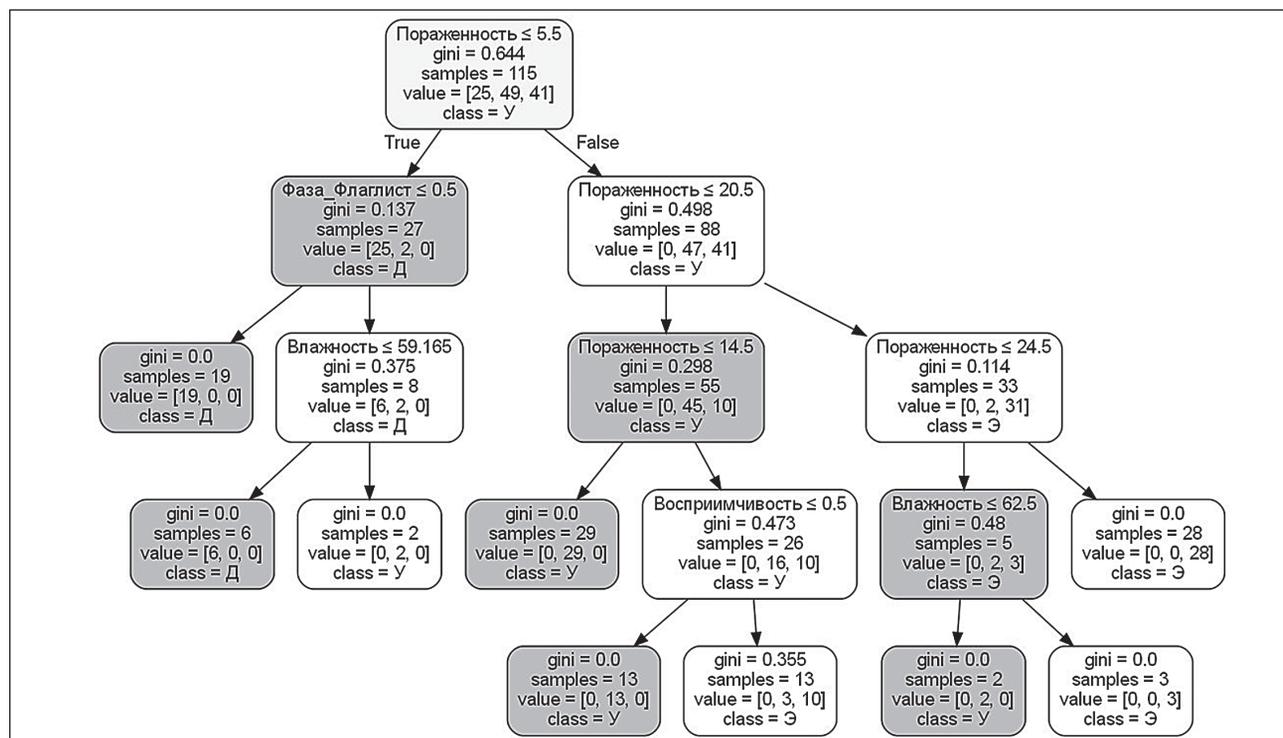


Рис. 1. Обученное дерево решений, прогнозирующее развитие сетчатой пятнистости озимого ячменя.

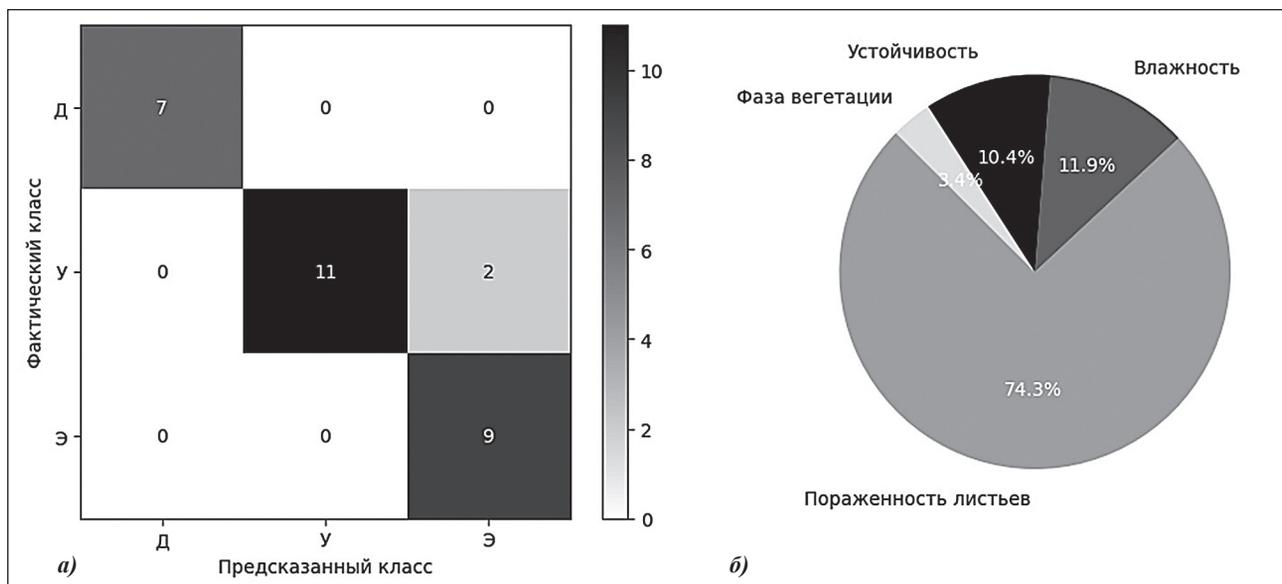


Рис. 2. Результаты моделирования:
а) матрица ошибок алгоритма; б) важность факторов при моделировании.

Построение модели прогноза сетчатой пятнистости осуществляли на языке программирования *Python*. Для имплементации алгоритма дерева решений, определения параметров модели и ее отрисовки (рис. 1) задействовали библиотеку *scikit-learn 1.0.2*, которая предоставляет собой обширный функционал для машинного обучения. Каждый блок построенного дерева имеет однотипную структуру:

наименование фактора – признак, по которому разбивается выборка на две подгруппы. Узел дерева сравнивает значение этого признака с некоторым порогом, и, в зависимости от результата, данные направляются в одну из ветвей дерева;

индекс Джини («gini») – измеряет степень «перепутанности» классов в узле. Чем меньше его значение, тем «чище» узел;

объем выборки (samples) – количество наблюдений (экземпляров) в узле дерева;

значение (value) – содержит информацию о распределении классов в узле. Например, если узел разбивает выборку на две подгруппы, значение будет представлять собой количество экземпляров каждого класса в каждой подгруппе;

класс (class) – если узел листовой (конечный), то есть не имеет дочерних узлов, этот параметр отражает прогнозируемый класс для такого листа.

В каждой внутренней вершине дерева отражен признак, по которому проводится разбиение выборки на два множества. При этом для всех качественных признаков предварительно провели «быстрое кодирование» (*one-hot encoding*), представляющее категориальные признаки в виде бинарных векторов, что делает их пригодными для использования в моделях машинного обучения, включая деревья решений. Так, для фактора «устойчивость сорта» было выполнено обычное бинарное кодирование: 0 – устойчивый сорт, 1 – восприимчивый. Для фактора «фаза вегетации» кодирование проводили уже с использованием двух бинарных переменных:

$$\text{Фаза_кущения} = \begin{cases} 0, & \text{наблюдение не относится к фазе кущения,} \\ 1, & \text{наблюдение относится к фазе кущения.} \end{cases}$$

$$\text{Фаза_флаглист} = \begin{cases} 0, & \text{наблюдение не относится к фазе флаглиста,} \\ 1, & \text{наблюдение относится к фазе флаглиста.} \end{cases}$$

После этого три возможные фазы вегетации, которые фиксировали в опыте, представляли в виде парных комбинаций значений фиктивных переменных: (1, 0) – фаза кущения; (0, 1) – фаза флаглиста; (0, 0) – фаза выхода в трубку, а затем подавали на вход модели.

Результатом классификации служили метки, содержащиеся в терминальных (листовых) вершинах. Для оценки качества работы алгоритма на обучающей выборке достаточно сравнить распределение истинных значений целевой переменной для всех объектов, попавших в листовую вершину (вектор *value*) со значением прогнозируемого класса. Например, если в листовой вершине вектор *value* = [0, 13, 0] и модельный прогноз выдает «Умеренное развитие», то можно констатировать, что все 13 наблюдений были проклассифицированы корректно. В случае, когда *value* = [0, 3, 10] и моделью прогнозируется «Эпифитотийное развитие», то алгоритм ошибся трижды, отнеся три наблюдения с умеренным развитием сетчатой пятнистости к наиболее пессимистичному сценарию. В наших исследованиях (см. рис. 1) такой случай ошибочной классификации выявлен в одной единственной конечной вершине. Доля правильных ответов алгоритма (*accuracy*) на обучающей выборке составила 98,2%.

Естественно, что оценка качества моделирования только по обучающей выборке может быть недостаточно информативной. Это связано с явлением, называемым переобучением (*overfitting*), когда модель слишком подстраивается под обучающие данные, но плохо работает на новых, тестовых наборах.

При детальном рассмотрении матрицы ошибок алгоритма на тестовой выборке (рис. 2а) можно сделать заключение, что обученная модель корректно классифицировала большинство объектов на основе новых данных, совершив ошибку всего на одном наблюдении. Это свидетельствует о её способности обобщать данные, отсутствии переобучения и пригодности для практического применения.

Информация о важности (весах) факторов (рис. 2б), используемых моделью при принятии решений, показывает, насколько каждый признак влияет на прогноз. Самое значимое влияние на развитие сетчатой

пятнистости ячменя оказывает пораженность листьев (74,3%). Доли устойчивости сорта и относительной влажности находились на уровне 10%, а наименее важным признаком оказалась «фаза вегетации» ячменя.

Представленные результаты имеют ряд важных практических и научных приложений:

учитывая способность деревьев решений к масштабированию, при наличии собранных данных по другим культурам и болезням, представленные ИИ-решения могут стать инструментом для принятия управленческих решений, так как предоставляют оперативную и объективную оценку степени их развития;

использование модели способствует более эффективному распределению временных и материальных ресурсов, что в конечном итоге влияет на эффективность производственной деятельности (благодаря сокращению производственных затрат).

Выводы. На основе современных методов машинного обучения разработана ИИ-модель прогнозирования развития сетчатой пятнистости озимого ячменя. В качестве базового алгоритма было выбрано дерево решений. Настроенная по обучающей выборке модель продемонстрировала высокую прогнозную способность на тестовых данных, обеспечив долю правильных ответов 98% на обучающей выборке и 97% на отложенной.

Основными факторами, влияющими на развитие сетчатой пятнистости ячменя, выступают текущая пораженность листьев (вклад 74,3%), средняя относительная влажность воздуха (11,9%), устойчивость сорта к болезни (10,4%) и фаза вегетации, в которую произошло заражение (3,4%).

ФИНАНСИРОВАНИЕ РАБОТЫ.

Исследование выполнено при финансовой поддержке Кубанского научного фонда в рамках научного проекта № МФИ-20.1/121. Никаких дополнительных грантов на проведение или руководство данным конкретным исследованием получено не было.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ.

В этой работе отсутствуют исследования человека или животных.

КОНФЛИКТ ИНТЕРЕСОВ.

Авторы этой работы заявляют, что у них нет конфликта интересов.

Литература.

1. Федеральная служба государственной статистика. *Сельское хозяйство, охота и лесное хозяйство*. URL: https://rosstat.gov.ru/enterprise_economy (дата обращения: 01.10.2023).
2. Указ Президента Российской Федерации от 21.01.2020 г. № 20 «Об утверждении Доктрины продовольственной безопасности Российской Федерации». URL: <http://www.kremlin.ru/acts/bank/45106> (дата обращения: 01.10.2023).
3. Применение технологий искусственного интеллекта в сельском хозяйстве / Е. А. Скворцов, В. И. Набоков, К. В. Некрасов и др. // *Аграрный вестник Урала*. 2019. № 8 (187). С. 91–98. doi: 10.32417/article_5d908ed78f7fc7.89378141.
4. Волкова Г. В., Астанчук И. Л. Распространение *Ryzenophora teres* на посевах ячменя в Северо-Кавказском регионе // *Зерновое хозяйство России*. 2019. № 2 (62). С. 63–68. doi: 10.31367/2079-8725-2019-62-2-63-68.
5. Орловская Е. Н., Астанчук И. Л., Волкова Г. В. Влияние системы обработки почвы на поражаемость озимого ячменя сетчатой пятнистостью // *Вестник Ульяновской государственной сельскохозяйственной академии*. 2018. № 4 (44). С. 110–117. doi: 10.18286/1816-4501-2018-4-110-117.
6. Ариничева И. В., Ариничев И. В., Репко Н. В. Автоматизированная диагностика грибных болезней ячменя // *Труды Кубанского государственного аграрного университета*. 2022. № 95. С. 69–72. doi: 10.21515/1999-1703-95-69-72.
7. Ариничев И. В., Полянских С. В., Ариничева И. В. Семантическая сегментация ржавчин и пятнистостей пшеницы // *Компьютерная оптика*. 2023. Т. 47. № 1. С. 118–125. doi: 10.18287/2412-6179-СО-1130.
8. Билай В. И. Микроорганизмы – возбудители болезней растений (справочник). Киев: Наукова думка, 1988. 549 с.
9. Хасанов Б. А. *Определитель грибов-возбудителей «гельминтоспориозов» растений из родов Bipolaris, Drechslera и Exserohilum*. Ташкент: Фан, 1992. 180 с.
10. Ибрагимов Т. З., Санин С. С. Цифровая защита растений и интеллектуальный анализ фитосанитарной информации // *Защита и карантин растений*. 2019. № 4. С. 15–18.

Поступила в редакцию 10.11.2023

После доработки 19.12.2023

Принята к публикации 09.01.2024