


doi 10.18699/vjgb-25-34

Идентификация грибных болезней земляники садовой на основе анализа гиперспектральных изображений методами машинного обучения

А.Ф. Чешкова 

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

 cheshanna@yandex.ru

Аннотация. Белая, бурая и угловатая пятнистости являются наиболее распространенными грибными болезнями земляники садовой в Западной Сибири, значительно влияющими на ее урожайность и качество. Точная, быстрая и неинвазивная диагностика этих заболеваний имеет важное значение в промышленном производстве земляники. В настоящей статье исследуются возможности применения методов машинного обучения и гиперспектральной визуализации для обнаружения и дифференциации на листьях земляники симптомов, вызванных патогенными грибами *Ramularia tulasnei* Sacc., *Marssonina potentillae* Desm. и *Dendrophoma obscurans* Anders. Спектр отражения листьев регистрировали гиперспектральной камерой Photonfocus MV1-D2048x1088-HS05-96-G2-10 в лабораторных условиях методом линейного сканирования. Для дифференциации здоровых и пораженных областей листьев изучено пять методов машинного обучения: метод опорных векторов (SVM), метод К-ближайших соседей (KNN), линейный дискриминантный анализ (LDA), дискриминантный анализ частичных наименьших квадратов (PLS-DA) и случайный лес (RF). С целью уменьшения высокой размерности извлеченных спектральных данных и увеличения скорости их обработки было отобрано несколько подмножеств оптимальных длин волн, несущих наиболее важную спектральную информацию. Рассмотрены следующие методы сокращения размерности: метод анализа ROC-кривых, метод анализа производных, метод PLS-DA, метод ReliefF. Кроме того, 16 вегетационных индексов задействовано в качестве информативных признаков. Наибольшую точность классификации, 89.9 %, показал метод опорных векторов на полном спектре значений. При использовании вегетационных индексов и наборов оптимальных длин волн общая точность классификации всех методов снизилась незначительно по сравнению с классификацией на полном спектре значений. Результаты исследования подтверждают перспективность применения методов гиперспектральной визуализации в сочетании с методами машинного обучения для дифференциации грибных болезней земляники садовой.

Ключевые слова: гиперспектральные изображения; грибные болезни земляники; методы машинного обучения; сокращение размерности

Для цитирования: Чешкова А.Ф. Идентификация грибных болезней земляники садовой на основе анализа гиперспектральных изображений методами машинного обучения. *Вавиловский журнал генетики и селекции*. 2025;29(2):310-319. doi 10.18699/vjgb-25-34

Identification of fungal diseases in strawberry by analysis of hyperspectral images using machine learning methods

A.F. Cheshkova 

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

 cheshanna@yandex.ru

Abstract. Leaf spot, leaf scorch and phomopsis leaf blight are the most common fungal diseases of strawberry in Western Siberia, which significantly reduce its yield and quality. Accurate, fast and non-invasive diagnosis of these diseases is important for strawberry production. This article explores the ability of hyperspectral imaging to detect and differentiate symptoms caused to strawberry leaves by pathogenic fungi *Ramularia tulasnei* Sacc., *Marssonina potentillae* Desm. and *Dendrophoma obscurans* Anders. The reflection spectrum of leaves was acquired with a Photonfocus MV1-D2048x1088-HS05-96-G2-10 hyperspectral camera under laboratory conditions using the line scanning method. Five machine learning methods were considered to differentiate between healthy and diseased leaf areas: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Partial Least Squares Discriminant Analysis (PLS-DA), and Random Forest (RF). In order to reduce the high dimensionality of the extracted spectral data and to increase the speed of their processing, several subsets of optimal wavelengths were selected. The following dimensionality reduction methods were explored: ROC curve analysis method, derivative analysis method, PLS-DA method, and ReliefF method. In addition, 16 vegetation indices were used as features. The support vector machine method demonstrated the highest classification accuracy of 89.9 % on the full range spectral data. When using vegeta-

tion indices and optimal wavelengths, the overall classification accuracy of all methods decreased slightly compared to the classification on the full range spectral data. The results of the study confirm the potential of using hyperspectral imaging methods in combination with machine learning for differentiating fungal diseases of strawberries.

Key words: hyperspectral imaging; fungal diseases of strawberries; machine learning methods; dimensionality reduction

For citation: Cheshkova A.F. Identification of fungal diseases in strawberry by analysis of hyperspectral images using machine learning methods. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2025;29(2):310-319. doi 10.18699/vjgb-25-34

Введение

Земляника садовая – одна из самых популярных ягодных культур среди потребителей во всем мире благодаря своим высоким вкусовым качествам и питательной ценности (Zheng et al., 2021). Кроме того, земляника отличается способностью к быстрому вегетативному размножению, высокой урожайностью и рентабельностью. К сдерживающим факторам увеличения производства ягод земляники садовой относится значительное поражение возделываемых сортов грибными болезнями, что приводит к снижению урожайности и экономическим потерям. Наиболее распространенные грибные болезни земляники садовой в Западной Сибири – рамуляриоз (белая пятнистость), бурая и угловатая пятнистости (Говорова, Говоров, 2015). Раннее выявление данных заболеваний имеет решающее значение для целенаправленного применения соответствующих мер защиты растений.

Традиционные методы диагностики, такие как визуальная экспертиза и микробиологический лабораторный анализ, высокочувствительны по времени и человеческим ресурсам, что ограничивает возможность их применения в крупных сельскохозяйственных предприятиях. В последние годы большой потенциал продемонстрировал метод анализа гиперспектральных изображений (Mishra et al., 2017; Mahlein et al., 2018; Чешкова, 2022) в качестве эффективного и неинвазивного инструмента мониторинга биотического и абиотического стресса растений. При воздействии патогенов происходят изменения биохимических и биофизических свойств растений, создается спектр отражения, отличный от спектра здоровых листьев. Современные оптические сенсоры регистрируют до нескольких сотен полос электромагнитного спектра в широком диапазоне длин волн и формируют спектральный профиль для каждого отдельного пикселя, тем самым комбинируя спектральную и пространственную информацию (Mishra et al., 2017). Метод гиперспектральной визуализации соединяет в себе преимущества методов машинного зрения и оптической спектроскопии, позволяя одновременно оценивать как физиологические, так и морфологические параметры. В настоящее время в научных публикациях приводятся примеры успешного применения метода гиперспектральной визуализации для распознавания различных заболеваний земляники, таких как мучнистая роса (Mahmud et al., 2020), антракноз (Lu et al., 2017; Jiang et al., 2021), вертициллезное увядание (Cockerton et al., 2019), серая гниль (Wu et al., 2023), пятнистости (Cheshkova, 2023).

Для анализа большого объема полученных многомерных данных наилучшим инструментом являются методы машинного обучения (Nagaraju et al., 2020; Venos et al., 2021). Множество исследований показывает, что использо-

вание вегетационных индексов в качестве информативных признаков для построения моделей машинного обучения позволяет добиться хороших результатов в обнаружении и распознавании болезней сельскохозяйственных культур (Mahlein et al., 2013; Lu et al., 2017).

Гиперспектральные данные характеризуются высокой коллинеарностью. Большое количество длин волн усложняет модели и снижает производительность. Один из важных этапов предварительной обработки данных – уменьшение размерности, целью которого является устранение спектральной избыточности при сохранении важной информации (Liu et al., 2014; Sun, Du, 2019). Уменьшение размерности может быть достигнуто за счет выбора определенных длин волн либо выделения информативных признаков.

Цель настоящей работы – исследование возможности применения метода гиперспектральной визуализации в сочетании с методами машинного обучения для дифференцирования здоровых и пораженных белой (*Ramularia tulasnei* Sacc.), бурой (*Marssonina potentillae* Desm.) и угловатой (*Dendrophoma obscurans* Anders.) пятнистостями листьев земляники садовой; сравнение и выбор наиболее точного метода классификации гиперспектральных изображений, позволяющего идентифицировать грибные болезни земляники садовой; исследование возможности использования методов сокращения размерности и вегетационных индексов для оптимизации моделей машинного обучения.

Материалы и методы

Растительный материал. Исследованы три вида грибных болезней земляники садовой, наиболее распространенных в Западной Сибири: белая, бурая и угловатая пятнистости.

Белая пятнистость (рамуляриоз). Возбудитель: *Mycosphaerella fragariae* (Tul.) Lindau; конидиальная стадия – *Ramularia tulasnei* Sacc. На пораженных листьях появляются мелкие пурпуровые или красновато-бурые округлые пятна, которые разрастаются в диаметре от 1 до 8 мм, в центре становятся белыми, с темно-бурым или пурпуровым ободком.

Бурая пятнистость земляники. Возбудитель: *Diplocarpon earliana* (Ell. et Ev.) Wolf; конидиальная стадия – *Marssonina potentillae* (Desm.) P. Magn., *M. fragariae* (Lib.) Ohl. Пораженные листья покрываются многочисленными, от очень мелких (0.1 мм), до крупных (0.5 см), темнопурпуровыми пятнами, часто расплывчатыми, иногда сливающимися.

Угловатая пятнистость (фомопсис, побурение). Возбудитель: *Dendrophoma obscurans* (Ell. et Ev.) H.W. Ander-

son (синонимы *Phomopsis obscurans* (Ell. et Ev.) Sutton). Болезнь характеризуется появлением на листьях (чаще старых) округлых или неопределенной формы светло-, темно- или красно-коричневых пятен, часто со светлым центром. Вокруг пятна обычно темно-коричневая или темно-красная кайма. Пятна расположены по краю листовой пластинки или вдоль средней жилки листа.

Исследования проводили на образцах земляники садовой, выращенных на биополигоне Сибирского физико-технического института аграрных проблем (СибФТИ) Сибирского федерального научного центра агробиотехнологий РАН (Новосибирская область, р.п. Краснообск) в 2021–2023 гг. В течение вегетационного периода было отобрано 120 растений, включающих 30 здоровых растений, а также по 30 растений, имеющих видимые симптомы поражения белой, бурой или угловатой пятнистостями. На каждом растении было выбрано по одному листу. Листья были отделены от растений для проведения дальнейших исследований в лаборатории. Идентификация болезни и степени поражения осуществлялась посредством визуальной экспертизы по видимым симптомам заболевания (Garrido et al., 2011; Говорова, Говоров, 2015).

Получение и калибровка изображения. Съемку проводили гиперспектральной камерой Photonfocus MV1-D2048x1088-HS05-96-G2-10, с сенсором IMEC CMV2K-LS150-VNIR (Photonfocus AG, Швейцария), диапазон длин волн 470–900 нм, спектральное разрешение 3 нм, пространственное разрешение 2048×1088 пикселей методом линейного сканирования с использованием подвижной платформы, управляемой контроллером. Программно-аппаратный комплекс для выполнения исследований разработан совместно Институтом автоматизации и электротехники СО РАН и СибФТИ СФНЦА РАН (Максимов и др., 2023). Листья земляники для проведения съемки помещали на площадку белого цвета. Источник освещения – две галогенные лампы. Шаг сканирования, диафрагма и высота закрепления камеры были определены экспериментально. Дополнительно в конце каждого сканирования создавали два калибровочных файла с «черным» и «белым» эталонными спектрами. «Черный» спектр получали, закрыв объектив камеры непрозрачной крышкой. «Белый» спектр регистрировали с использованием стандартной белой поверхности для получения максимально возможной интенсивности пикселей на каждой длине волны. Затем исходное гиперспектральное изображение калибровали по формуле:

$$R = \frac{I_S - I_D}{I_W - I_D},$$

где I_S – исходное изображение; I_D – черное эталонное изображение; I_W – белое эталонное изображение; R – результирующее изображение.

По результатам сканирования были сформированы трехмерные массивы данных (гиперкубы), содержащие два измерения пространственной информации и дополнительно одно измерение спектральной информации (2048×1088×131).

Извлечение и обработка данных. Полученные файлы изображений были разделены на две группы: 96 изображений листьев для формирования обучающей выборки

(по 24 в каждом из 4 классов) и 24 изображения листьев для формирования валидационной выборки (по 6 в каждом из 4 классов). Извлечение спектральных данных выполняли в программе ENVI 5.2 (NV5 Geospatial Solutions, Inc., США). На изображении листьев земляники вручную были выделены области интереса (ROI), соответствующие здоровой ткани листа и цветковым пятнам белой, бурой и угловатой пятнистостей (рис. 1). Из каждой выделенной области случайным образом выбирали по 250 пикселей. В результате были сформированы обучающая выборка из 24 000 значений спектра (по 6000 пикселей для класса) и валидационная выборка из 6000 значений (по 1500 пикселей).

Далее для сглаживания спектра и коррекции рассеивания к спектральным данным были применены фильтр Савицкого–Голя (Savitzky, Golay, 1964) и нормализация стандартным преобразованием (standard normal variate) (Vidal, Amigo, 2012).

Методика определения оптимальных длин волн. В нашем исследовании рассмотрены следующие методы сокращения размерности:

- метод анализа ROC-кривых (Luo et al., 2012); в качестве метрики, определяющей значимость признака, используется значение AUC (area under curve); из всего спектра данных оставляют те длины волн, для которых AUC превышает определенное пороговое значение;
- метод анализа производных (Savitzky, Golay, 1964); из всего спектра данных оставляют те длины волн, для которых абсолютное значение 2-й производной имеет локальный экстремум;
- метод PLS-DA (Mehmood et al., 2012); после построения модели для каждого класса выбирают длины волн, для которых абсолютная величина коэффициента регрессии превышает определенное пороговое значение;
- метод ReliefF (Kononenko, 1994; Urbanowicz et al., 2018); для ранжирования признаков используют итеративный алгоритм, который присваивает признакам разные веса в зависимости от категории и корреляции каждого объекта, признаки с весами ниже определенного порогового значения удаляются.

Выбор вегетационных индексов. Вегетационные индексы являются алгебраическими комбинациями, рассчитываемыми на основе значений спектра отражения для двух или более выбранных длин волн.

Для нашего исследования было отобрано 16 вегетационных индексов (табл. 1), характеризующих содержание хлорофилла, пигментов, воды, азота и углерода в тканях растений (Wu et al., 2023). Эти вегетационных индексы связаны с фотохимической активностью (PRI), изменением пигментов (PSSRa, PSSRb, CRI, ARI), содержанием азота (NRI), углерода (PSRI), хлорофилла (MCARI, TVI, VOG1, VOG2, VOG3) и общим физиологическим состоянием листьев растений (NDVI, RENDVI, RVSI, PhRI).

Методы моделирования и анализа данных. Для дифференциации здоровых и пораженных областей листьев земляники в нашем исследовании было рассмотрено пять методов машинного обучения (SVM, KNN, LDA, PLS-DA, RF), наиболее часто используемых в классификации гиперспектральных данных (Singh et al., 2016; Benos et al., 2021).

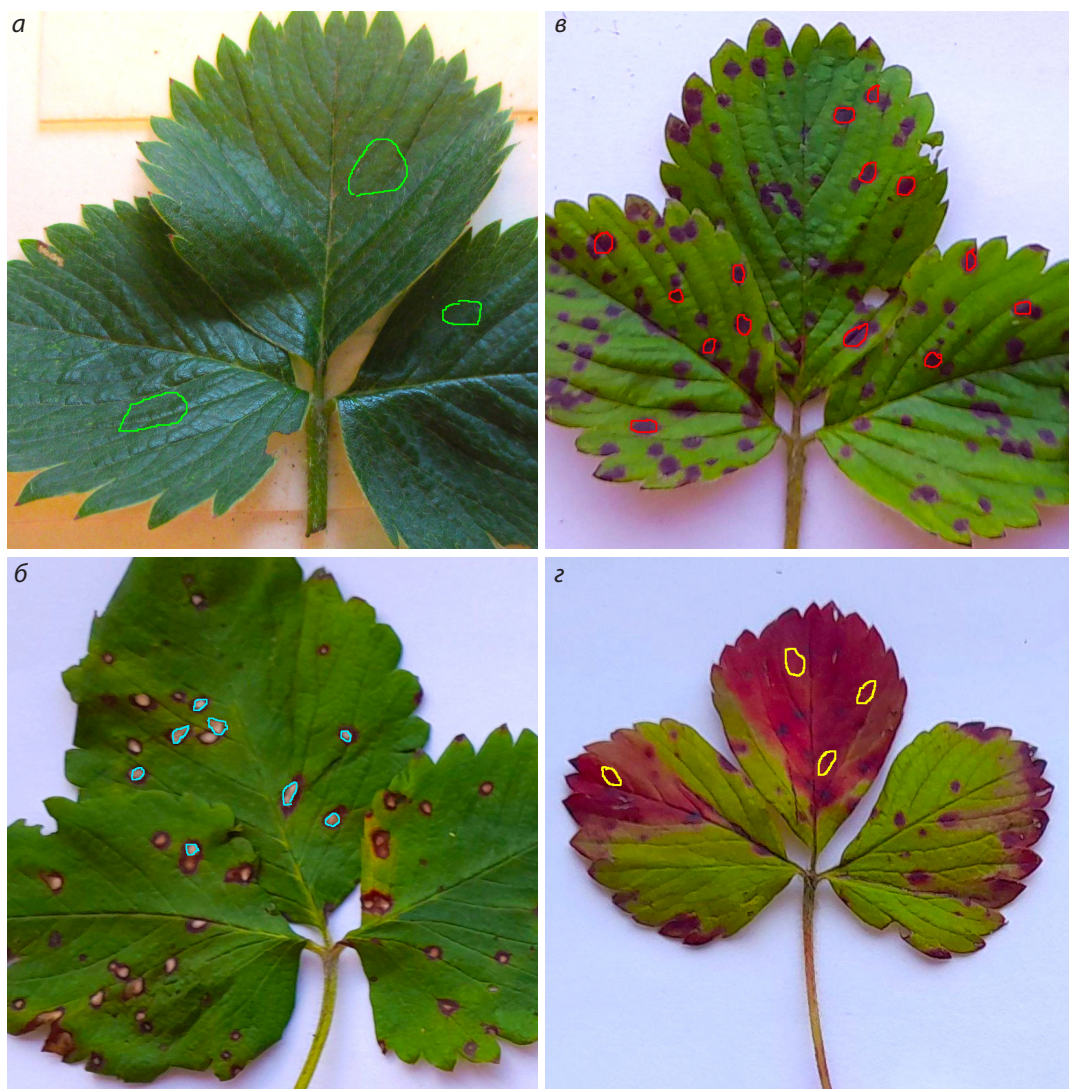


Рис. 1. Выделение ROI: а – здоровые области, б – области белой пятнистости, в – области бурой пятнистости, з – области угловой пятнистости.

Метод опорных векторов (support vector machine). Основная идея метода SVM – перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с наибольшим зазором в этом пространстве. В качестве ядра классификатора была взята радиальная базисная функция Гаусса.

Метод К-ближайших соседей (K-nearest neighbors). При классификации объект присваивается к классу, наиболее распространенному среди К-ближайших соседей этого элемента, классы которых уже известны. Близость объектов определяет значение функции расстояния. В нашем исследовании в качестве функции расстояния использована евклидова метрика.

Линейный дискриминантный анализ (linear discriminant analysis). Проецирует многомерные данные в пространство меньшей размерности, чтобы способствовать разделению классов. Оптимальная проекция в классическом LDA получается путем максимизации расстояния между разными классами и минимизации расстояния внутри одного класса.

Дискриминантный анализ частичных наименьших квадратов (partial least squares discriminant analysis). Является вариантом объединения регрессии PLSR в пространстве оптимальных латентных структур и дискриминантного анализа (DA). В отличие от классического дискриминантного анализа, в котором выполняется поиск гиперплоскостей максимальной дисперсии независимых предикторов, PLS-DA строит модель линейной регрессии, проецируя прогнозируемые и наблюдаемые переменные в новое редуцированное пространство.

Случайный лес (Random Forest). Нелинейная модель машинного обучения, основанная на использовании ансамбля решающих деревьев.

В качестве метрики оценки качества моделей используется общая точность классификации (overall accuracy), вычисляемая как процент отношения правильно классифицированных объектов к общему числу объектов.

Все расчеты и анализ данных были проведены в программной среде R с помощью пакетов caret, kernlab, randomForest, klaR, pls, CORElearn, class, MASS, terra.

Таблица 1. Вегетационные индексы, использованные в качестве информативных признаков

№	Индекс	Формула
1	NDVI (normalized difference vegetation index)	$(R_{800} - R_{670}) / (R_{800} + R_{670})$
2	RENDVI (red edge normalized difference vegetation index)	$(R_{750} - R_{705}) / (R_{750} + R_{705})$
3	PhRI (physiological reflectance index)	$(R_{550} - R_{531}) / (R_{550} + R_{531})$
4	RVSI (red-edge vegetation stress index)	$[(R_{712} + R_{752}) / 2] - R_{732}$
5	MCARI (modified chlorophyll absorption ratio index)	$[(R_{700} - R_{670}) - 0.2 \cdot (R_{700} - R_{550})] \cdot (R_{700} / R_{670})$
6	TVI (triangular vegetation index)	$0.5 \cdot [120 \cdot (R_{750} - R_{550}) - 200 \cdot (R_{670} - R_{550})]$
7	VOG1 (Vogelman index 1)	R_{740} / R_{720}
8	VOG2 (Vogelman index 2)	$(R_{734} - R_{747}) / (R_{715} + R_{726})$
9	VOG3 (Vogelman index 3)	R_{715} / R_{705}
10	PSRI (plant senescence reflectance index)	$(R_{680} - R_{500}) / R_{750}$
11	NRI (nitrogen reflectance index)	$(R_{570} - R_{670}) / (R_{570} + R_{670})$
12	PSSRa	R_{800} / R_{680}
13	PSSRb (pigments specific simple ratio)	R_{800} / R_{635}
14	CRI (carotenoid reflectance index)	$1 / R_{510} - 1 / R_{550}$
15	ARI (anthocyanin reflectance index)	$1 / R_{550} - 1 / R_{700}$
16	PRI (photochemical/physiological reflectance index)	$(R_{531} - R_{570}) / (R_{531} + R_{570})$

Примечание. NDVI – нормализованный разностный вегетационный индекс; RENDVI – нормализованный разностный вегетационный индекс в крайнем красном спектре; PhRI – индекс физиологического отражения; RVSI – вегетационный индекс состояния в крайнем красном спектре; MCARI – модифицированный относительный индекс поглощения в хлорофилле; TVI – треугольный вегетационный индекс; VOG1, VOG2, VOG3 – индексы Вогельмана; PSRI – индекс отражения огрубевшего углерода в растительных тканях; NRI – индекс отражения азота; PSSRa, PSSRb – индексы удельного соотношения пигментов; CRI – индекс отражения каротиноидов; ARI – индекс отражения антоцианов; PRI – фотохимический индекс отражения.

Результаты

Спектральные характеристики образцов

На рис. 2 представлены усредненные значения спектра отражения здоровых и пораженных грибными болезнями листьев земляники. Спектральные кривые имеют типичные для растительной ткани характеристики (Mishra et al., 2017). Общей чертой всех спектральных кривых является более низкий коэффициент отражения в видимом диапазоне длин волн по сравнению с ближним инфракрасным диапазоном. На длине волн около 670 нм наблюдается снижение коэффициента отражения, что связано с сильным поглощением света хлорофиллом в листьях. В диапазоне длин волн 670–760 нм отражательная способность листьев резко возрастает по причине рассеивания света в межклеточном пространстве. В диапазоне длин волн 760–900 нм коэффициент отражения остается высоким.

Наблюдаются определенные различия между спектрами. Так, у здоровых зеленых тканей листа отмечается характерный пик на длине волны 550 нм (зона поглощения азота), в то время как пораженные пятнистостями ткани имеют спад в этой области. В диапазоне 720–810 нм здоровая и пораженная бурой пятнистостью ткани имеют более высокий коэффициент отражения, по сравнению с бурой и угловатой пятнистостями. А в диапазоне 810–900 нм, наоборот, более низкий. Отражательная характеристика листьев, пораженных белой пятнистостью, равномерно растет на всем диапазоне длин волн.

Выбор оптимальных длин волн

Предварительно проведенный дисперсионный анализ (ANOVA) выявил значимость различий средних значений

коэффициентов отражения в исследуемых областях для каждой из длин волн согласно F-критерию с уровнем значимости $p < 0.001$. Кроме того, для каждой из рассматриваемых моделей был применен пошаговый метод сокращения размерности, в результате которого также было обнаружено, что все длины волн значимы для классификации.

Для сокращения размерности исходных данных были рассмотрены четыре различные методики, определившие разные наборы информативных длин волн (рис. S1–S4 Приложения)¹.

Методом анализа ROC-кривых были определены 23 длины волны (нм): [541.39, 545.04, 548.92, 550.41, 553.99, 557.94, 561.3, 565.18, 568.58, 745.48, 748.98, 751.75, 756.45, 759.36, 763.0, 765.97, 769.44, 772.39, 775.92, 778.56, 781.11, 784.53, 787.2], для которых AUC превысило пороговое значение 0.99.

Методом анализа производных при пороговом значении 1.0 выбраны следующие 15 длин волн (нм): [677.11, 680.47, 682.99, 685.28, 688.76, 691.62, 695.25, 697.97, 709.54, 712.19, 729.07, 732.25, 736.15, 739.20, 742.67].

Методом PLS-DA для порогового значения коэффициентов регрессии 0.4 выбраны следующие 16 длин волн (нм): [498.68, 502.7, 505.97, 510.11, 513.5, 517.33, 522.39, 526.49, 529.98, 533.99, 541.39, 680.47, 682.99, 688.76, 691.62, 722.02].

Методом ReliefF для порогового значения коэффициентов значимости 0.5 выбраны следующие 24 длины волны (нм): [537.27, 541.39, 545.04, 548.92, 550.41, 557.94,

¹ Рис. S1–S4 Приложения см. по адресу: <https://vavilovj-icg.ru/download/pict-2025-29/appx10.pdf>

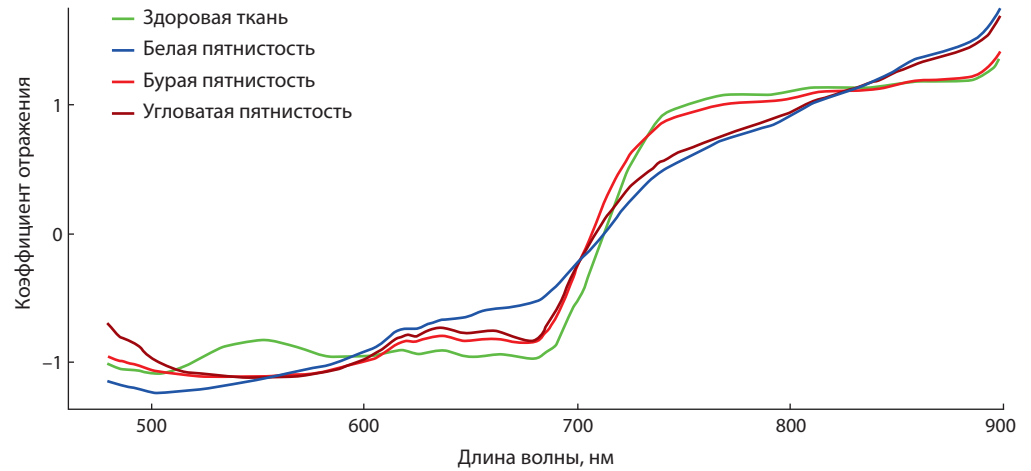


Рис. 2. Средний спектр отражения здоровых и инфицированных областей листьев земляники садовой.

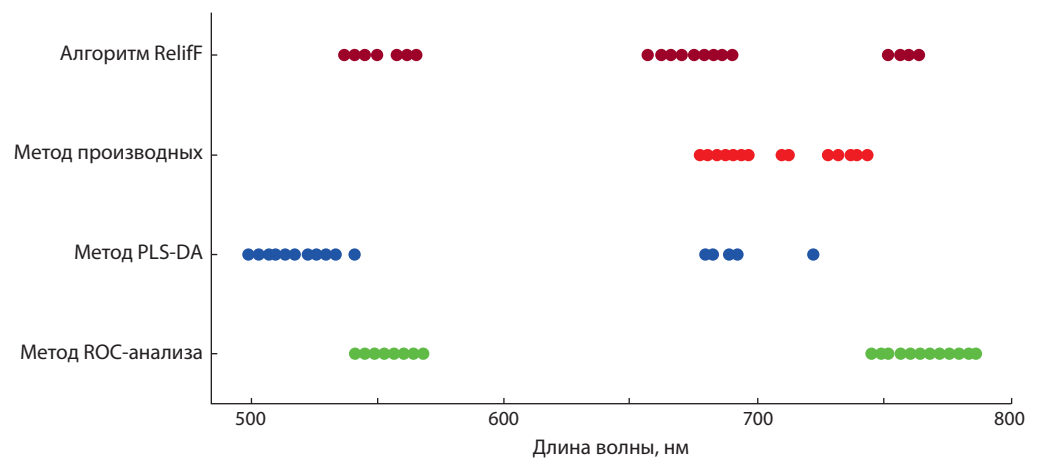


Рис. 3. Наборы оптимальных длин волн, определенных различными методами.

561.3, 565.18, 657.08, 662.06, 664.83, 668.35, 670.93, 674.46, 677.11, 680.47, 682.99, 685.28, 688.76, 691.62, 751.75, 756.45, 759.36, 763.00].

Сравнение между собой выбранных различными методами наборов оптимальных длин волн (рис. 3) позволяет сделать вывод, что наиболее информативны для классификации диапазоны длин волн [542–565 нм] и [680–691 нм].

Расчет вегетационных индексов

Шестнадцать вегетационных индексов были рассчитаны по соответствующим формулам (см. табл. 1) для каждого объекта в обучающей выборке. Для каждого индекса был выполнен дисперсионный анализ (ANOVA) с целью выявления значимости различий средних значений индексов по видам болезней. Все 16 индексов имели уровень значимости $p < 0.001$.

Результаты классификации на полном спектре значений

В нашем исследовании пять различных моделей, SVM, KNN, LDA, PLS-DA, RF, были применены для классификации здоровых и пораженных грибными болезнями листьев земляники. Сначала модели были построены для полного спектра длин волн (131 длина волны в диапазо-

не 470–900 нм). Методом кросс-валидации были подобраны следующие оптимальные гиперпараметры: SVM ($\sigma = 0.03$, $C = 6$), KNN ($K = 9$), RF ($mtry = 11$), PLS-DA ($ncomp = 38$). Результаты классификации приведены в табл. 2. Анализ данных показывает, что основные ошибки в классификации происходят при дифференциации бурой и угловатой пятнистостей, так как эти области имеют сходный спектр отражения.

Метод опорных векторов на полном спектре значений показал наибольшую точность классификации (90 %), а метод К-ближайших соседей – наименьшую (85 %).

Результаты классификации на наборах оптимальных длин волн и на вегетационных индексах

Каждая из пяти моделей классификации, SVM, KNN, LDA, PLS-DA, RF, была протестирована на наборах оптимальных длин волн, полученных в результате применения четырех различных методов сокращения размерности (анализа ROC-кривых, анализа производных, PLS-DA, ReliefF), а также на наборе значений 16 вегетационных индексов (табл. 3).

Как видно из табл. 3, общая точность классификации всех методов снизилась, по сравнению с классификацией

Таблица 2. Матрицы ошибок для классификации гиперспектральных изображений различными методами на полном спектре значений

Результат классификации	Реальные значения класса, %				Общая точность классификации, %
	Здоровые области	Белая пятнистость	Бурая пятнистость	Угловатая пятнистость	
SVM					89.90
Здоровые области	98.07	0.73	0.07	0.00	
Белая пятнистость	0.73	95.13	1.07	1.27	
Бурая пятнистость	0.33	3.20	92.67	25.00	
Угловатая пятнистость	0.87	0.93	6.20	73.73	
KNN					85.17
Здоровые области	98.73	1.13	0.33	0.00	
Белая пятнистость	0.27	90.80	1.27	1.73	
Бурая пятнистость	0.27	5.33	92.13	39.27	
Угловатая пятнистость	0.73	2.73	6.27	59.00	
RF					86.93
Здоровые области	98.20	0.87	0.20	0.00	
Белая пятнистость	0.40	93.27	0.67	0.47	
Бурая пятнистость	0.27	3.53	90.40	33.67	
Угловатая пятнистость	1.13	2.33	8.73	65.87	
LDA					89.15
Здоровые области	99.00	1.00	0.07	0.0	
Белая пятнистость	0.07	93.07	0.93	0.4	
Бурая пятнистость	0.93	3.53	89.53	24.6	
Угловатая пятнистость	0.00	2.40	9.47	75.0	
PLS-DA					87.63
Здоровые области	99.93	1.87	1.07	0.00	
Белая пятнистость	0.00	94.60	1.87	0.33	
Бурая пятнистость	0.00	1.47	81.13	24.80	
Угловатая пятнистость	0.07	2.07	15.93	74.87	

Примечание. SVM – метод опорных векторов; KNN – метод К-ближайших соседей; LDA – линейный дискриминантный анализ; PLS-DA – дискриминантный анализ частичных наименьших квадратов; RF – случайный лес.

Таблица 3. Общая точность классификации (%) гиперспектральных изображений различными методами на выбранных длинах волн и на вегетационных индексах

Модель	Метод сокращения размерности				На 16 вегетационных индексах
	анализа ROC-кривых	анализа производных	PLS-DA	ReliefF	
SVM	77.37	83.03	85.13	78.32	89.75
KNN	76.40	81.60	83.27	77.48	82.65
RF	77.55	82.67	83.50	78.57	84.25
LDA	74.32	80.55	85.88	76.93	84.62
PLS-DA	69.00	75.45	85.60	77.15	84.70

Примечание. SVM – метод опорных векторов; KNN – метод К-ближайших соседей; RF – случайный лес; LDA – линейный дискриминантный анализ; PLS-DA – дискриминантный анализ частичных наименьших квадратов.

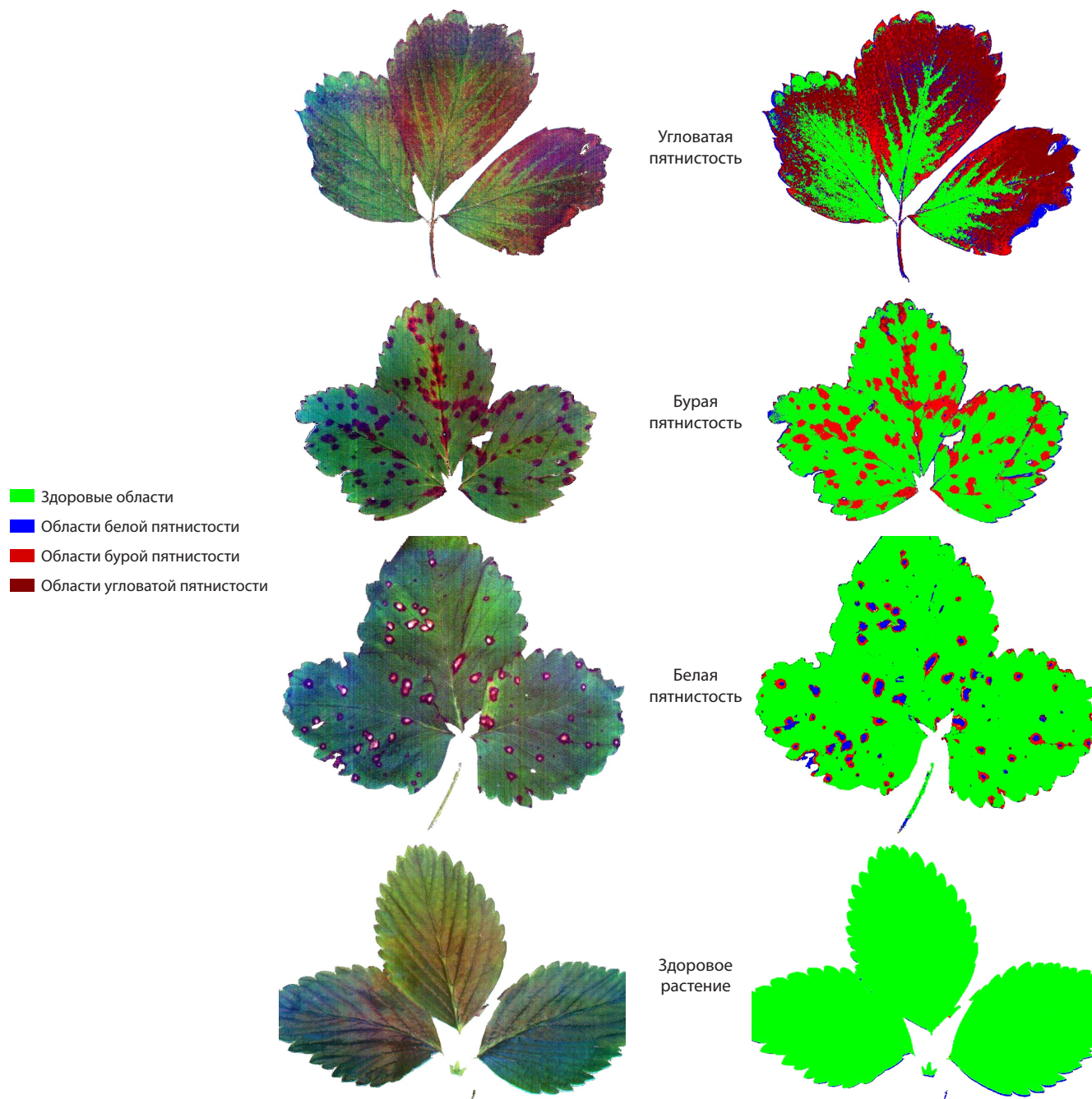


Рис. 4. Визуализация классификации видов болезней земляники методом опорных векторов на полном спектре значений.

на полном спектре значений. Наибольшая точность классификации для всех моделей получена для вегетационных индексов в качестве информативных признаков и для набора длин волн, выбранных методом PLS-DA.

Диагностика грибных болезней земляники с использованием оптимизированных моделей

Построенные оптимизированные модели могут быть использованы для обнаружения и дифференциации грибных болезней земляники. На рис. 4 представлен пример применения модели SVM для диагностики различных видов грибных болезней.

Обсуждение

Анализ гиперспектральных изображений методами машинного обучения уже успешно применялся в научных исследованиях для выявления болезней земляники. Так, G. Wu с коллегами (2023) изучали возможности использования гиперспектральной визуализации в сочетании со спектральными характеристиками, вегетационными индексами и текстурными характеристиками для обнаружения серой гнили на листьях земляники в лабораторных условиях. Были построены и оптимизированы три модели машинного обучения: ELM, KNN, SVM. Общая точность классификации моделей достигала 96 %.

В работе (Jiang et al., 2021) на основе выбранных спектральных признаков применено шесть методов машинного обучения, SVM, ELM, KNN, PLS-DA, RF, NB, для раннего обнаружения антракноза и серой гнили земляники в лабораторных условиях. Большинство классификационных моделей показали высокую точность (до 100 %), распознавая классы инфекций до появления видимых симптомов заболевания. В нашем исследовании полученная точность классификации заболеваний не превысила 90 %. Данный результат можно объяснить несколькими причинами. Во-первых, рассматривались сразу три вида заболеваний, а не одно или два, как в других исследованиях. Во-вторых, для успешной дифференциации заболеваний необходимо различие спектральных характеристик пораженных патогенами листьев растений. Мы выявили, что основные ошибки в классификации происходят при дифференциации бурой и угловатой пятнистостей, так как эти области имеют сходный спектр отражения. Возможный путь решения данной проблемы – использование сверточных нейронных сетей, учитывающих не только спектральные, но и текстурные характеристики пораженных листьев, такие как форма и расположение пятен.

Выбор метода классификации зависит от культуры и изучаемых болезней. Среди рассмотренных нами пяти популярных моделей машинного обучения, SVM, KNN, LDA, PLS-DA, RF, наилучшую точность классификации показал метод опорных векторов (SVM), что согласуется с результатами других работ (Benos et al., 2021).

Для снижения размерности и выбора оптимальных длин волн при построении моделей ученые применяют различные методы. Так, в работах (Luo et al., 2012; Jiang et al., 2021; Wu et al., 2023) задействованы алгоритмы CARS, CARS-RF, ReliefF и ROC. Сокращение размерности во многих исследованиях не снижает точности моделей, однако в нашем случае все исходные длины волн оказались информативными, и точность классификации незначительно снизилась по сравнению с полным спектром.

Полученные результаты лабораторных исследований говорят о перспективности применения методов гиперспектральной визуализации для диагностики грибных болезней земляники садовой в сельскохозяйственном производстве. В литературе уже описаны примеры успешного использования гиперспектральных сенсоров, установленных на БПЛА, для диагностики биотических и абиотических стрессов растений (Yang et al., 2017).

В дальнейшем мы планируем опробовать применение методов гиперспектральной визуализации в полевых условиях с целью автоматизации диагностики грибных болезней земляники.

Заключение

Анализ гиперспектральных изображений с использованием методов машинного обучения позволяет обнаружить и дифференцировать грибные болезни земляники (белую, бурую и угловатую пятнистости) при наличии видимых симптомов заболеваний. Для эффективного определения вида заболевания были построены и оценены пять различных моделей классификации: SVM, KNN, LDA, PLS-DA и RF. Наибольшую точность классификации, 89.9 %, на полном спектре значений (131 длина волны в диапазоне

470–900 нм) показал метод опорных векторов (SVM). С целью упрощения моделей и повышения скорости обработки данных было рассмотрено четыре различных метода сокращения размерности (анализа ROC-кривых, анализа производных, PLS-DA, ReliefF). Кроме того, модели были протестированы на наборе из 16 вегетационных индексов в качестве исходных признаков. Общая точность классификации всех методов снизилась незначительно по сравнению с классификацией на полном спектре значений. Набор из 16 оптимальных длин волн, полученный методом PLS-DA, и набор из 16 вегетационных индексов имели более высокую точность классификации, чем остальные наборы длин волн.

Список литературы / References

- Говорова Г.Ф., Говоров Д.Н. Грибные болезни земляники садовой, селекция на иммунитет и другие методы защиты. М.: РГАУ-МСХА, 2015
- [Govorova G.F., Govorov D.N. Fungal Diseases of Garden Strawberries: Breeding for Immunity and Other Methods of Protection. Moscow, 2015 (in Russian)]
- Максимов Л.В., Гурова Т.А., Елкин О.В. Разработка программно-аппаратного комплекса для фенотипирования стресса зерновых культур по гиперспектральным изображениям. В: Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов (SDM-2023). Новосибирск: ФИЦ ИВТ, 2023;289-293. doi 10.25743/sdm.2023.42.73.049
- [Maximov L.V., Gurova T.A., Elkin O.V. Development of software and hardware complex for phenotyping of grain crop stress by hyperspectral images. In: Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes (SDM-2023). Novosibirsk, 2023;289-293. doi 10.25743/sdm.2023.42.73.049 (in Russian)]
- Чешкова А.Ф. Обзор современных методов обнаружения и идентификации болезней растений на основе анализа гиперспектральных изображений. *Вавиловский журнал генетики и селекции*. 2022;26(2):202-213. doi 10.18699/VJGB-22-25
- [Cheshkova A.F. A review of hyperspectral image analysis techniques for plant disease detection and identification. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2022;26(2):202-213. doi 10.18699/VJGB-22-25]
- Benos L., Tagarakis A., Dolias G., Berruto R., Kateris D., Bochtis D. Machine learning in agriculture: a comprehensive updated review. *Sensors*. 2021;21:3758. doi 10.3390/s21113758
- Cheshkova A.F. Application of machine learning methods for the differentiation of fungal diseases in strawberry based on hyperspectral image analysis. *E3S Web Conf*. 2023;390:03018. doi 10.1051/e3s-conf/202339003018
- Cockerton H.M., Li B., Vickerstaff R.J., Eyre C.A., Sargent D.J., Armitage A.D., Marina-Montes C., Garcia-Cruz A., Passey A.J., Simpson D.W., Harrison R.J. Identifying *Verticillium dahlia* resistance in strawberry through disease screening of multiple populations and image based phenotyping. *Front Plant Sci*. 2019;10:924. doi 10.3389/fpls.2019.00924
- Garrido C., Carbu M., Fernandez-Acero J.F., Gonzalez-Rodriguez V.E., Cantoral J.M., New insights in the study of strawberry fungal pathogens. In: Husaini A.M., Mercado J.A. (Eds) Genomics, Transgenics, Molecular Breeding and Biotechnology of Strawberry. Global Science Books, UK, 2011;24-39
- Jiang Q., Wu G., Tian C., Li N., Yang H., Bai Y., Zhang B. Hyperspectral imaging for early identification of strawberry leaves diseases with machine learning and spectral fingerprint features. *Infrared Phys Technol*. 2021;118:103898. doi 10.1016/j.infrared.2021.103898
- Kononenko I. Estimating attributes: analysis and extensions of RELIEF. In: Bergadano F., De Raedt L. (Eds) Machine Learning: ECML-94. Lecture Notes in Computer Science. Vol. 784. Berlin: Springer, 1994;171-182. doi 10.1007/3-540-57868-4_57

- Liu D., Sun D.W., Zeng X.A. Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry. *Food Bioprocess Technol.* 2014;7:307-323. doi 10.1007/s11947-013-1193-6
- Lu J., Ehsani R., Shi Y., Abdulridha J., de Castro A.I., Xu Y. Field detection of anthracnose crown rot in strawberry using spectroscopy technology. *Comput Electron Agric.* 2017;135:289-299. doi 10.1016/j.compag.2017.01.017
- Luo X., Takahashi T., Kyo K., Zhang S. Wavelength selection in vis/NIR spectra for detection of bruises on apples by ROC analysis. *J Food Eng.* 2012;109(3):457-466. doi 10.1016/j.jfoodeng.2011.10.035
- Mahlein A.-K., Rumpf T., Welke P., Dehne H.-W., Plümer L., Steiner U., Oerke E.-C. Development of spectral indices for detecting and identifying plant diseases. *Remote Sens Environ.* 2013;128:21-30. doi 10.1016/j.rse.2012.09.019
- Mahlein A.-K., Kuska M.T., Behmann J., Polder G., Walter A. Hyperspectral sensors and imaging technologies in phytopathology: state of the art. *Annu Rev Phytopathol.* 2018;56:535-558. doi 10.1146/annurev-phyto-080417-050100
- Mahmud M.S., Zaman Q.U., Esau T.J., Chang Y.K., Price G.W., Prithiviraj B. Real-time detection of strawberry powdery mildew disease using a mobile machine vision system. *Agronomy.* 2020;10:1027. doi 10.3390/agronomy10071027
- Mehmood T., Liland K.H., Snipen L., Sæbø S. A review of variable selection methods in Partial Least Squares Regression. *Chemom Intell Lab Syst.* 2012;118:62-69. doi 10.1016/j.chemolab.2012.07.010
- Mishra P., Asaari M., Herrero-Langreo A., Lohumi S., Diezma B., Scheunders P. Close range hyperspectral imaging of plants: a review. *Biosyst Eng.* 2017;164:49-67. doi 10.1016/j.biosystemseng.2017.09.009
- Nagaraju M., Chawla P. Systematic review of deep learning techniques in plant disease detection. *Int J Syst Assur Eng Manag.* 2020;11(3):547-560. doi 10.1007/s13198-020-00972-1
- Savitzky A., Golay M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem.* 1964;36(8):1627-1639
- Singh A., Ganapathysubramanian B., Singh A.K., Sarkar S. Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 2016;21(2):110-124. doi 10.1016/j.tplants.2015.10.015
- Sun W., Du Q. Hyperspectral band selection: a review. *IEEE Geosci Remote Sens Mag.* 2019;7(2):118-139. doi 10.1109/MGRS.2019.2911100
- Urbanowicz R.J., Meeker M., Cava W., Olson R.S., Moore J.H. Relief-based feature selection: introduction and review. *J Biomed Inf.* 2018;85:189-203. doi 10.1016/j.jbi.2018.07.014
- Vidal M., Amigo J.M. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemom Intell Lab.* 2012;117:138-148. doi 10.1016/j.chemolab.2012.05.009
- Wu G., Fang Y., Jiang Q., Cui M., Li N., Ou Y., Diao Z., Zhang B. Early identification of strawberry leaves disease utilizing hyperspectral imaging combing with spectral features, multiple vegetation indices and textural features. *Comput Electron Agric.* 2023;204:107553. doi 10.1016/j.compag.2022.107553
- Yang G., Liu J., Zhao C., Li Z., Huang Y., Yu H., Xu B., Yang X., Zhu D., Zhang X., Zhang R., Feng H., Zhao X., Li Z., Li H., Yang H. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front Plant Sci.* 2017;8:1111. doi 10.3389/fpls.2017.01111
- Zheng C., Abd-Elrahman A., Whitaker V. Remote sensing and machine learning in crop phenotyping and management, with an emphasis on applications in strawberry farming. *Remote Sens.* 2021;13:531. doi 10.3390/rs13030531

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию 27.04.2024. После доработки 07.11.2024. Принята к публикации 07.11.2024.