

МЕТОДЫ КЛАССИФИКАЦИИ ЭЛЕМЕНТОВ ЗЕРНОВЫХ СМЕСЕЙ НА ОСНОВЕ АНАЛИЗА СПЕКТРАЛЬНЫХ ХАРАКТЕРИСТИК В ВИДИМОМ И ИНФРАКРАСНОМ ДИАПАЗОНАХ ДЛИН ВОЛН

А. О. Донских*, Д. А. Минаков**, А. А. Сирота*, В. А. Шульгин*

**Воронежский государственный университет,*

***ВУНЦ ВВС «Военно-воздушная академия им. проф. Н. Е. Жуковского и Ю. А. Гагарина» (г. Воронеж)*

Поступила в редакцию 20.03.2016 г.

Аннотация. Рассматривается задача классификации элементов зерновых смесей на основе методов спектрального анализа, выполняемого в различных участках диапазона длин волн. Для проведения классификации используются различные варианты получения первичных измерений с использованием спектров отражения, спектров пропускания и их комбинаций. Для последующей обработки получаемых первичных спектральных измерений используются нейросетевые алгоритмы классификации элементов зерновых смесей. Приводятся результаты оценки вероятности ошибочной классификации в зависимости от количества используемых спектральных участков и комбинаций методов, применяемых для получения первичных измерений.

Ключевые слова: спектральный анализ, компоненты зерновых смесей, распознавание образов, нейронные сети.

Annotation. Methods of classification of grain mixture components based on spectral analysis in different wavelength ranges are considered. Various measurement approaches – reflection, transmission and combined spectrum methods – are used. To further process the results of the spectral measurement, neural network based classification algorithms are used. The probabilities of incorrect recognition for various numbers of spectral parts and combinations of spectral methods are estimated.

Keywords: spectral analysis, components of grain mixtures, pattern recognition, neural networks.

ВВЕДЕНИЕ

Эффективное решение задачи классификации элементов зерновых смесей (ЭЗС) в интересах анализа качества используемого семенного материала возможно путем интеграции новейших достижений в области компьютерных технологий, электроники, средств высокоскоростной регистрации и цифровой обработки сигналов и изображений. Одним из перспективных подходов для решения указанной задачи является использование методов глубокого спектрального анализа в сочетании с современными методами машин-

ного обучения для последующей обработки результатов первичных измерений.

Вариации качества семенного материала, обусловленные технологией выращивания, почвенно-климатическими условиями, материнскими, мутагенными факторами и т. п., часто приводят к снижению урожайности. Поэтому всегда важно иметь семенной материал с высокими показателями сортовой чистоты. Например, для пшеницы [1] в категориях семян оригинальные семена (ОС) и элитные семена (ЭС) сортовая чистота должна составлять не менее 99,7 %, в категории репродукционные семена (РС) – 98 %, в категории репродукционные семена товарной продукции (РСТ) – 95 %. В биологическом растениеводстве также большое значение

© Донских А. О., Минаков Д. А., Сирота А. А., Шульгин В. А., 2016

имеет анализ показателей засоренности ЭЗС сорняками, зараженности их болезнями, наличия в них вредителей, поврежденных и невыполненных семян.

Как правило, качество используемого семенного материала определяется химическими методами. Весьма эффективным является метод молекулярного маркирования, заключающийся в исследовании полиморфизма белков (прежде всего запасных белков) и нуклеиновых кислот в ЭЗС [2–4]. Для оценки биологической ценности и всхожести ЭЗС применяют стандартные химические методы [5, 6]. Относительный состав биологически значимых показателей семян, таких как белок, крахмал, жир и др. является отличительным генетическим признаком сорта [7]. Все эти методы показали свою эффективность в области селекции и семеноводства, однако их применение для анализа больших партий ЭЗС имеет весьма существенные ограничения, поскольку их реализация требует относительно большого времени анализа, дорогостоящего химического оборудования, реактивов, специально подготовленного персонала.

Альтернативный подход основан на применении методов оптической спектроскопии, которые в настоящее время находят все большее применение для контроля качества сельскохозяйственной и пищевой продукции. Кроме того, как показано, в частности в [8–12], эти методы могут быть использованы при создании датчиков потокового типа, применяемых в высокопроизводительных оптических сепараторах, ориентированных на сортировку интенсивных потоков ЭЗС в реальном времени.

Спектральный анализ состава сложных многокомпонентных образцов, как показали многочисленные исследования, целесообразно проводить в ближнем инфракрасном диапазоне длин волн (ИК), исследуя, например, колебания O-H, C-H, C-O и N-H связей и тем самым определять биологическую ценность ЭЗС. Метод не является разрушающим, при этом его реализация является относительно недорогой. В литературе можно встретить большое число работ, посвященных применению ближней ИК спектроскопии при

оценке качества ЭЗС различных сельскохозяйственных культур, определения пораженных различными заболеваниями и внешними факторами зерновок [13–23]. Для компьютерной обработки данных спектральных измерений применялись различные алгоритмы машинного обучения: *partial least squares discriminant analysis* (дискриминантный анализ с помощью регрессии на латентные структуры), *soft independent modeling of class analogy* (формальное независимое моделирование аналогий классов), *K-nearest neighbors* (классификация по *K* ближайшим соседям), *and least squares support vector machines* (метод опорных векторов).

Успешность решения задач оценки биологической ценности различных ЭЗС подтверждает наличие на рынке разнообразных экспресс-анализаторов в ближней ИК области, так называемых БИК-анализаторах [24–28]. Как правило, в существующих экспресс-анализаторах используется какой-либо один способ спектрального анализа, например, анализ спектров отражения или пропускания, поглощения или фотолюминесценции. Возможности повышения информативности анализа биологической ценности сырья, продуктов или ЭЗС, а также расширения области применения экспресс-анализаторов, по нашему мнению, заключаются в комбинированном использовании сразу нескольких способов спектрального анализа. Кроме того, в существующих БИК-анализаторах проводится спектральный анализ либо прошедшего, либо отраженного излучения, полученного сразу от множества ЭЗС. Это приемлемо для анализа ЭЗС в пищевой или кормовой промышленности, но в области семеноводства возникает необходимость контроля каждого объекта в отдельности.

В связи с этим представляется актуальной задача исследования различных методов спектрального определения качества семенного материала, чистоты сорта и исследования возможности повышения эффективности решения данной задачи при одновременном использовании нескольких спектральных методов, т. е. на основе мультиспектральной обработки (МСО), а также построенных на ос-

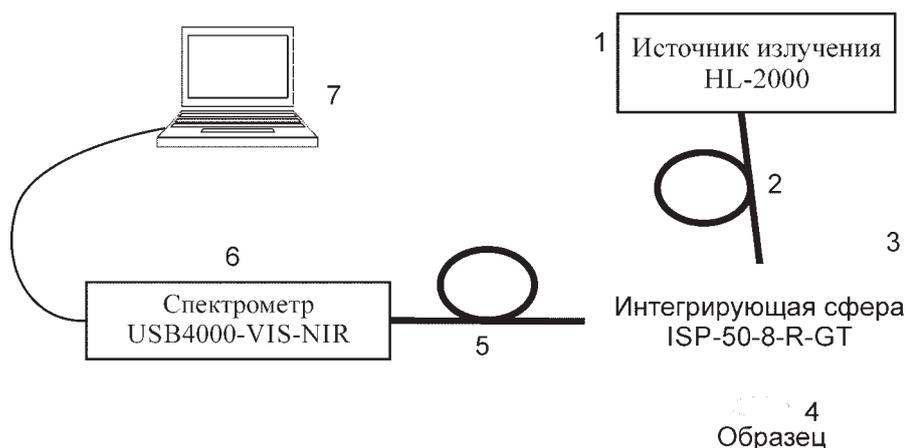


Рис. 1. Схема экспериментальной установки для измерения спектров отражения ЭЗС

нове данных МСО алгоритмов компьютерной обработки в интересах классификации ЭЗС для анализаторов комбинированного типа.

МЕТОДЫ ПОЛУЧЕНИЯ ПЕРВИЧНЫХ ИЗМЕРЕНИЙ

В качестве объектов исследования были выбраны три группы семян: два сорта озимых «Доктрина 110» и «Горка», а также сорт яровой тритикале – «Укро». Семена были выведены и любезно предоставлены ГНУ Воронежской ИСХ Россельхозакадемии имени В. В. Докучаева. Для исследования возможности спектрального определения яровых и озимых семян тритикале были использованы методы спектроскопии отражения и рассеяния в видимом и ближнем ИК диапазоне. Как известно зерна тритикале – это гибрид мягкой пшеницы и озимой ржи [29, 30]. Для формирования эталонных групп семян каждого сорта применялся метод молекулярного маркирования, путем записи электрофоретических спектров запасных белков семян тритикале. Отметим, что внешне семена этих сортов схожи по цвету и по форме.

Спектры отражения и рассеяния в видимом и ближнем ИК диапазоне измеряли с помощью экспериментальных установок, созданных на базе волоконных спектрометров USB4000-VIS-NIR и NirQuest 512 (Ocean Optics).

Экспериментальная установка для измерения спектров отражения показана на рис. 1.

Здесь свет от источника излучения HL-2000 (Ocean Optics) 1 поступает через световод 2 в интегрирующую сферу ISP-50-8-R-GT (Ocean Optics) 3 под углом 8° . После чего, отраженный от образца 4 свет под углом 90° выходит из интегрирующей сферы и по световоду 5 поступает к спектрометру 6, соединённому с компьютером 7. Для получения опорного сигнала использовали эталон отражения WS-1. Для регистрации спектров отражения в ближнем ИК диапазоне (до 2100 нм) спектрометр USB4000-VIS-NIR заменяли на спектрометр NirQuest 512, а также применяли интегрирующую сферу ISP-REF со встроенным источником излучения.

Экспериментальная установка для измерения спектров коллимированного пропускания (рассеяния вперед) семян представлена на рис. 2. Здесь свет от источника излучения HL-2000 (Ocean Optics) 1 поступает через световод 2 и направляется к коллиматору 3. После чего коллимированный пучок света направляется на образец 4. Прошедший через образец свет собирается коллиматором 5, сопряженным с выходным световодом 6 и передается в спектрометр 7, соединенным с компьютером 8. Оптические оси обоих коллиматоров и образец располагаются на одной линии. Для этого коллиматоры и образец устанавливаются на специальные трехкоординатные подвижки (Thorlabs). Для увеличения динамического диапазона работы установки может применяться дополнительная ирисовая диафрагма, которая устанавливается после коллиматора 3.

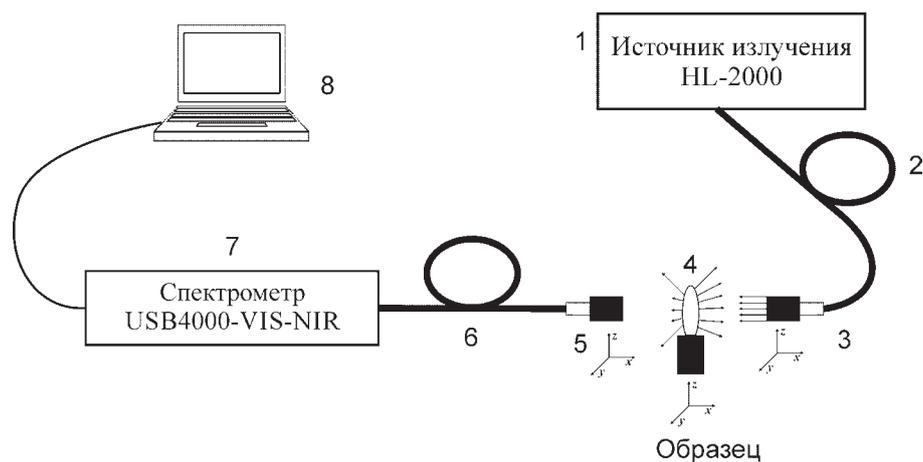


Рис. 2. Схема экспериментальной установки для измерения спектров пропускания ЭЗС

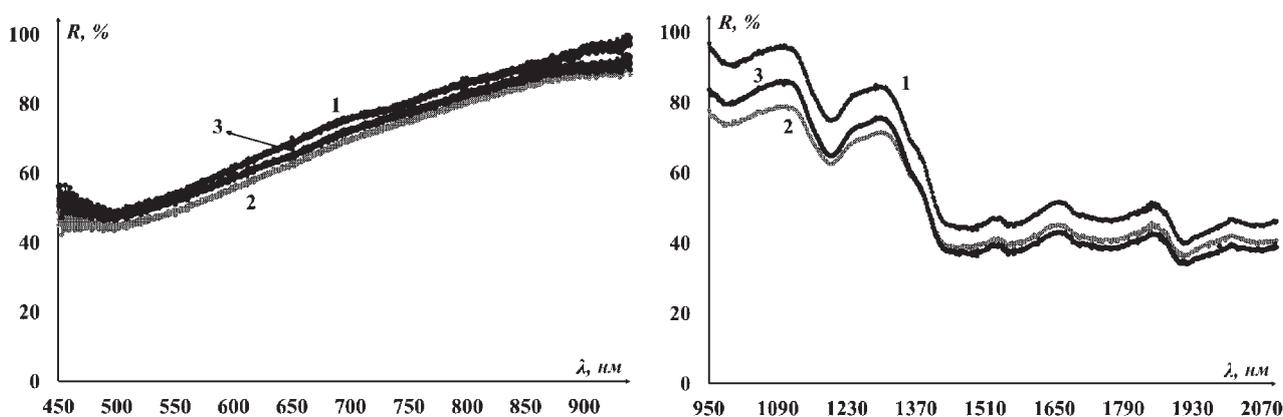


Рис. 3. Усредненные спектры отражения семян тритикале. Кривая 1 соответствует семенам «Укро»; 2 – «Доктрина 110»; 3 – «Горка»

Поскольку на практике семена даже одного класса могут быть сравнительно неоднородны, то для повышения точности спектральных исследований в каждом классе произвольным образом выбиралось 25 семян, причем для каждого проводилась регистрация не менее 10 спектров отражения и 10 спектров пропускания. Отметим также, что для корректного исследования спектров пропускания семян всех классов их толщина выдерживалась примерно одинаковой. Незначительные вариации толщины были существенно ниже, чем общая толщина объектов.

РЕЗУЛЬТАТЫ СПЕКТРАЛЬНОГО АНАЛИЗА

На рис. 3а, б представлены усредненные спектры отражения в видимом и ближнем ИК диапазоне семян тритикале. Усреднение проводили по всем 250 спектрам каждого класса образцов.

Из рис. 3а видно, что в видимом диапазоне спектры отражения исследованных семян трех сортов имеют схожую структуру и очень близки друг к другу. В ближнем ИК диапазоне уже наблюдается небольшое различие в поведении спектральных кривых (см. рис. 3б). Здесь спектры имеют довольно сложную структуру, которая связана с многокомпонентным химическим составом семян. Формы полос всех сортов семян тритикале очень схожи во всем спектральном диапазоне, но отличаются по величине коэффициента отражения. Спектральная кривая семян «Доктрины 110» лежит ниже, чем кривая для семян «Горки». Наибольшим отражением обладает яровой сорт тритикале.

На рис. 4а, б представлены усредненные спектры коллимированного пропускания семян тритикале в видимом и ближнем ИК диапазоне.

Как видно из рис. 4а, б спектральные кривые озимых сортов тритикале практиче-

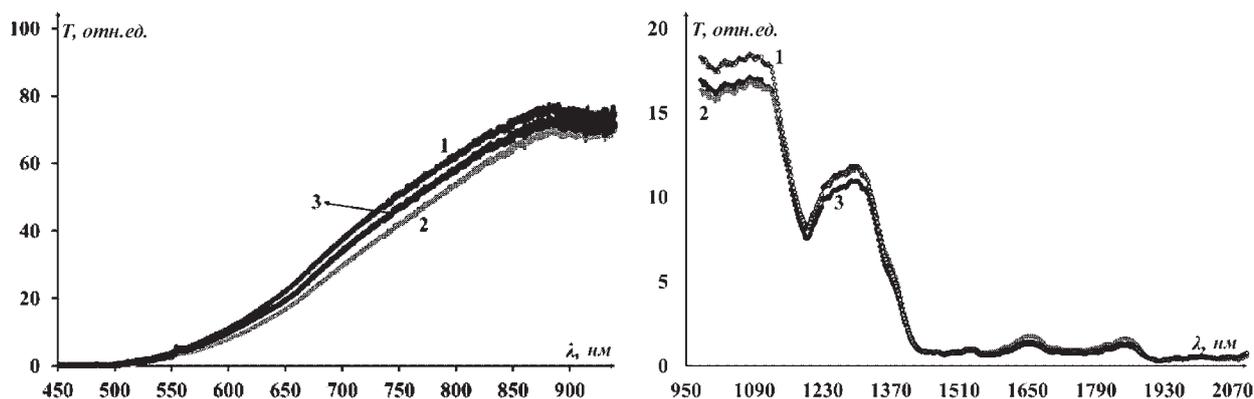


Рис. 4. Усредненные спектры коллимированного пропускания семян тритикале. Кривая 1 соответствует семенам «Укро»; 2 – «Доктрина 110»; 3 – «Горка»

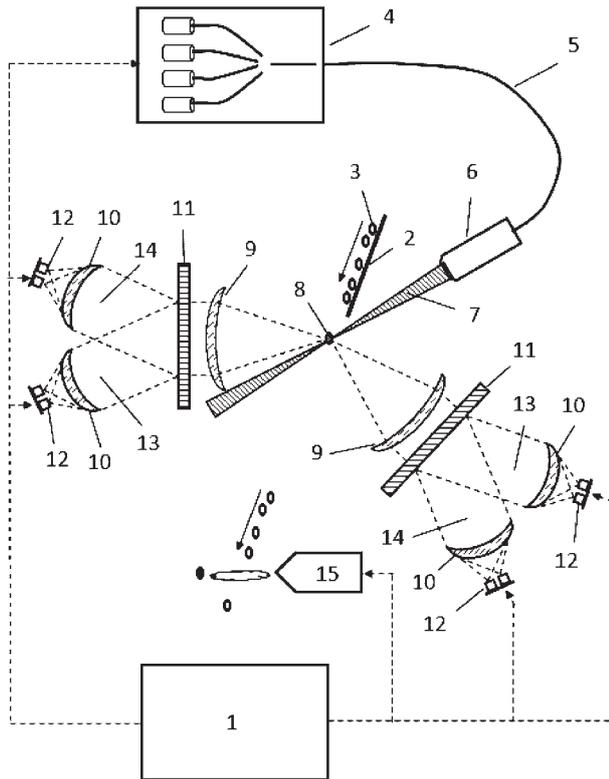
ски идентичны друг другу. Несколько большее пропускание в ближнем ИК диапазоне (от 950 нм до 1400 нм) характерно для семян «Укро». В диапазоне от 1400 нм наблюдается уже существенное уменьшение интенсивности прошедшего излучения для всех сортов тритикале.

Обнаруженная сложная структура спектральных кривых в спектрах отражения и пропускания обусловлена откликами молекул воды, белка, жиров, углеводов и других компонент. Визуальный анализ полученных результатов свидетельствует о том, что эффективность спектрального определения семян тритикале с одной стороны зависит от количества анализируемых спектральных интервалов, а с другой стороны от комбинации спектральных методов. Действительно, как в видимом диапазоне, так и в ближнем ИК применение сразу двух спектральных методов: спектроскопии отражения и рассеяния, потенциально может привести к существенному снижению ошибки классификации. При этом, как показывает визуальный анализ, даже использование небольшого числа спектральных участков дает высокую степень делимости классов ЭЗС. В связи с этим чрезвычайно перспективна разработка анализаторов комбинированного типа, способных использовать одновременно отраженное и прошедшее через объекты излучение.

ВИДЕОСПЕКТРАЛЬНЫЙ ЛАЗЕРНЫЙ АНАЛИЗАТОР

Как один из возможных вариантов решения конечной задачи классификации и разделения семенного материала на основе МСО предлагается рассмотреть предлагаемое нами устройство – видеоспектральный лазерный анализатор (ВЛА), в котором реализован принцип одновременного анализа движущихся объектов по спектральным изображениям отраженного и прошедшего лазерного излучения в четырех спектральных интервалах. В зависимости от типа линейного сенсора возможен выбор спектрального интервала как видимого, так и ближнего ИК диапазона. Выбор сенсора обусловлен конкретной задачей. В устройстве реализована также возможность распознавания и сортировки объектов по цвету, размерам, форме и иным характеристикам, доступным анализу видеоспектральных изображений средствами машинного зрения. На рис. 5 изображена схема ВЛА.

Освещение анализируемых объектов, движущихся по скатному лотку в форме монослоя осуществляется лазерным излучением различного спектрального состава с линейной разверткой, сфокусированной в предметной плоскости регистрации изображения. Для суммирования двух и более источников лазерного освещения используется оригинальное оптоволоконное устройство [12]. Отличительной особенностью ВЛА является получение цифрового изображения движущихся с высокой скоростью объектов



1 – устройство считывания и обработки изображения; 2 – скатный лоток; 3 – анализируемые объекты; 4 – оптоволоконное устройство суммирования источников лазерного освещения; 5 – волоконный световод; 6 – система цилиндрических линз, формирующих пространственное распределение пучка 7 в виде линии 8 в зоне регистрации изображения объектов; 9 – коллимирующие объективы; 10 – камерные объективы; 11 – дифракционные решетки; 12 – монохромные линейные сенсоры, сгруппированные в +1 и -1 порядках дифракции 13 и 14.

Рис. 5. Оптическая блок-схема видеоспектрального лазерного фотосепаратора

видеокамерой с линейными монохромными сенсорами без применения светофильтров, а также получение пиксельного изображения, в котором каждой точке объекта соответствует полный набор спектральных диапазонов лазерного освещения.

Применение освещения в виде линии, спектр которой представляет сумму нескольких источников лазерного излучения, позволил сформировать устройство, аналогичное спектрометру. Линейное освещение в данном случае выполняет функцию щели спектрометра. Отраженное и прошедшее через объекты излучение после прохождения коллиматорного объектива и дифракции на фазовой диф-

ракционной решетке, разлагается на нулевой, +1-й и -1-й порядки дифракции. В первых порядках присутствуют все спектральные составляющие лазерных источников, углы между которыми определяются длинами волн лазеров и постоянной решетки. Для двух длин волн достаточно просто выполнить независимую настройку совмещения изображений с линейными видеодатчиками в фокальной плоскости камерного объектива в одном из порядков дифракции.

КЛАССИФИКАЦИЯ ЭЛЕМЕНТОВ ЗЕРНОВЫХ СМЕСЕЙ ПО ДАННЫМ МУЛЬТИСПЕКТРАЛЬНОЙ ОБРАБОТКИ

Для обоснования эффективности подобных схем измерений и оценки выигрыша, получаемого за счет мультиспектральной обработки, потребовалось провести синтез и анализ алгоритмов машинного обучения, обеспечивающих классификацию элементов зерновых смесей как при использовании различных методов спектрального анализа в отдельности, так и при их комбинированном использовании. Кроме того, в задачах такого типа возникает проблема ограниченности объемов обучающих данных в связи с трудоемкостью проведения первичных измерений. Этот вопрос также потребовал отдельной проработки.

Ранее в подобных исследованиях [8–10] авторы рассматривали различные методы машинного обучения и, прежде всего, искусственные нейронные сети и метод машин опорных векторов. Учитывая, что применение нейронных сетей (НС) в этих исследованиях показало лучшие результаты, а также известные свойства их сходимости к оптимальным классификаторам данных, в данной работе авторы ограничились рассмотрением только методов машинного обучения, связанных с применением НС.

Для решения задачи классификации элементов зерновых смесей использовалась НС класса MLP (многослойный перцептрон). Сеть содержит один скрытый слой с сигмоидальной функцией активации и один выход-

ной слой с линейной функцией активации. Количество входных контактов сети n соответствует количеству используемых признаков распознавания, формируемых на основе компонентов спектров, полученных в ходе экспериментальных исследований, которые подаются на вход, а количество нейронов в выходном слое m_2 равно числу сортов семян (в данном случае 3), при этом значение «1» на выходе нейрона означает, что рассматриваемый образец принадлежит к соответствующему сорту, а «0» – не принадлежит. Количество нейронов в скрытом слое m_1 выбиралось из диапазона значений $n \leq m_1 \leq 2n + 1$. Сеть создавалась и тестировалась в среде MATLAB, для обучения использовался алгоритм Левенберга – Марквардта.

Для обеспечения инвариантности обработки исходных данных – измеренных спектров отражения и пропускания, первоначально проводилась их нормировка по следующей формуле:

$$s'_{i,j} = \frac{(s'_{i,j} - s'_{\min})}{(s'_{\max} - s'_{\min})}, \quad i = \overline{1, N}, \quad j = \overline{1, L},$$

$$s'_{\min} = \min(s'_{i,j}), \quad s'_{\max} = \max(s'_{i,j}),$$

где s' – матрица изначально зарегистрированных значений для спектра отражения или пропускания, N – количество спектральных компонентов, получаемых на установке для рассматриваемого спектра, L – количество проведенных на установке измерений.

Первичные измерения по каждому ЭЗС содержали по 512 спектральных составляющих для спектров пропускания и отражения в ближнем ИК диапазоне для каждого измерения. Однако в рассматриваемой задаче использование такого количества спектральных составляющих является нецелесообразным. Для получения n спектральных участков из N спектральных составляющих сначала проводилось усреднение по Δ соседним участкам для каждой составляющей:

$$\tilde{s}'_{i,j} = \sum_{t=i}^{i+\Delta-1} s_{t,j} / \Delta, \quad i = \overline{1, N}, \quad j = \overline{1, L},$$

$$\Delta \geq \text{floor}\left(\frac{N}{n}\right),$$

где $n < N$ – итоговое количество спектральных участков; $\text{floor}()$ – операция округления

в меньшую сторону. Затем выполнялось «прореживание» строк полученной матрицы \tilde{s}' таким образом, чтобы в результате получилась новая матрица усредненных спектральных компонент \tilde{s} размером $n \times L$:

$$\tilde{s}_{k,j} = \tilde{s}'_{ind,j}, \quad k = \overline{1, n}, \quad j = \overline{1, L},$$

$$ind = (k-1) * \text{floor}\left(\frac{N}{n}\right) + 1.$$

Кроме того, был исследован альтернативный способ снижения размерности исходных данных, основанный на применении метода главных компонент (РСА).

ОБУЧЕНИЕ И ТЕСТИРОВАНИЕ АЛГОРИТМА КЛАССИФИКАЦИИ

Для оценки эффективности классификации на экспериментальной установке было проведено по 10 измерений для 25 образцов. Таким образом, всего в качестве исходных данных для обучения НС по спектрам пропускания и отражения было получено по $L = 250$ измерений. При малом объеме данных традиционное разделение на обучающую и тестовую выборки в соотношении 2:1 или 1:1 нецелесообразно, так как объем как обучающей, так и тестовой выборки получается недостаточным для адекватной оценки эффективности. Поэтому при тестировании использовался метод скользящего контроля. Его суть состоит в последовательном изъятии из обучающей выборки одного образа, проведении обучения алгоритма на оставшихся данных и его последующем тестировании с использованием изъятых образцов. Полученные результаты усредняются по всем изъятым образцам. Полученные таким образом оценки могут быть использованы для подбора настраиваемых параметров алгоритма. Конкретно оценки вероятностей ошибок $\tilde{\alpha}(X^L)$, $\tilde{\beta}(X^L)$ для случая двух классов формируются на основе следующих выражений:

$$\tilde{\alpha}(X^L) = \frac{1}{L_1} \sum_{k=1}^{L_1} \tilde{\alpha}_k(X_k^{L_1-1}, X^{L_2}),$$

$$\tilde{\beta}(X^L) = \frac{1}{L_2} \sum_{k=1}^{L_2} \tilde{\beta}_k(X^{L_1}, X_k^{L_2-1}),$$

где $\tilde{\alpha}_k(X_k^{L_1-1}, X^{L_2})$ – индикатор ошибки пер-

вого рода, принимающий значение 0 или 1 при предъявлении для тестирования образа $x^{(1,k)}$ для алгоритма, построенного на основе обучающих выборок $X_k^{L_1-1}, X_k^{L_2}$, т. е. на основе общей выборки за вычетом образа $x^{(1,k)}$; $\tilde{\beta}_k(X_k^{L_1}, X_k^{L_2-1})$ – индикатор ошибки второго рода, принимающий значение 0 или 1 при предъявлении для тестирования образа $x^{(2,k)}$ для алгоритма, построенного на основе обучающих выборок $X_k^{L_1}, X_k^{L_2-1}$ т. е. на основе общей выборки за вычетом образа $x^{(2,k)}$.

На рис. 6 приведены полученные зависимости вероятности ошибки распознавания сорта семян тритикале от количества n анализируемых спектральных участков (рис. 6а) или главных компонент (рис. 6б). Видно, что при увеличении и числа спектральных участков, и числа главных компонент, участвующих в обучении НС, ошибка распознавания быстро уменьшается. При превышении порогового значения $n \geq 10$ вероятность ошибочного распознавания достаточно мала и существенно не снижается при дальнейшем увеличении n . Отсюда можно сделать вывод, что в составе измерительной установки допустимо использование более простых датчиков, регистрирующих небольшое число участков спектра, что весьма существенно для построения высокопроизводительных анализаторов, действующих в потоковом режиме.

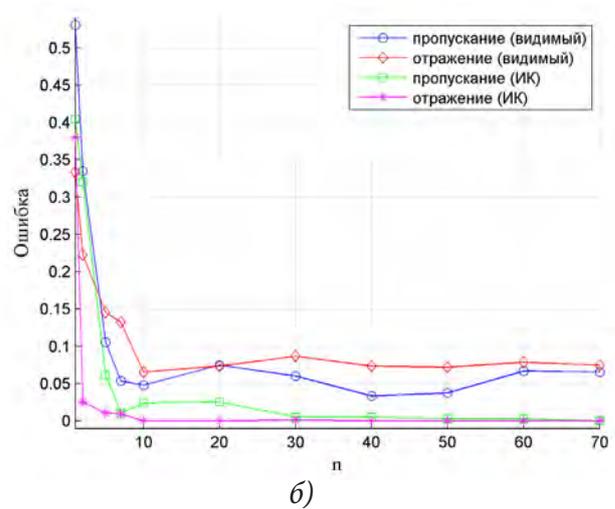
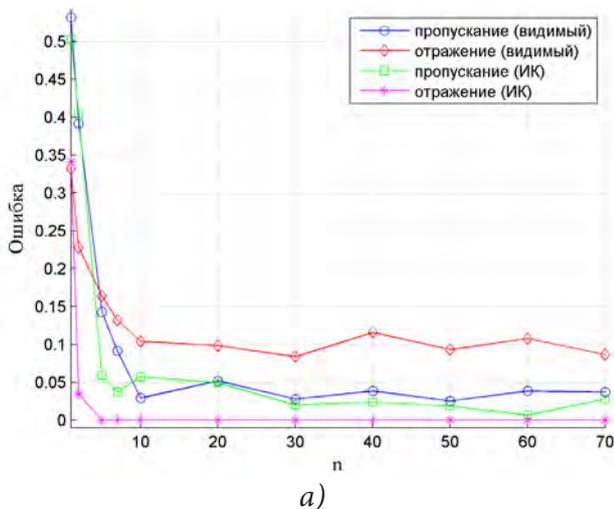


Рис. 6. Зависимость вероятности ошибочного распознавания сорта семян тритикале от а) – количества анализируемых спектральных участков; б) – количества анализируемых главных компонент

ПРОБЛЕМА ОГРАНИЧЕННОГО ОБЪЕМА ОБУЧАЮЩЕЙ ВЫБОРКИ

Для повышения устойчивости классификатора в условиях ограниченного объема обучающих данных было предложено использовать искусственное размножение данных в обучающей выборке за счет случайно сгенерированных данных. В данной работе такие данные формировались на основе оценок векторов математического ожидания и матриц ковариации «прореженных» до n спектральных участков первичных измерений, полученных для каждого класса ЭЗС. Сначала проводилась генерация L' случайных векторов, обладающих заданной матрицей ковариации нулевым вектором математического ожидания с использованием разложения Холецкого. Затем каждый из случайных векторов суммировался с вектором математического ожидания первичных измерений. В результате получалось L' случайно сгенерированных измерений для n спектральных участков, обладающих тем же вектором математического ожидания и матрицей ковариации, что и полученные экспериментально данные. Искусственно сгенерированные данные добавлялись только к обучающей выборке, для тестирования использовались исклю-

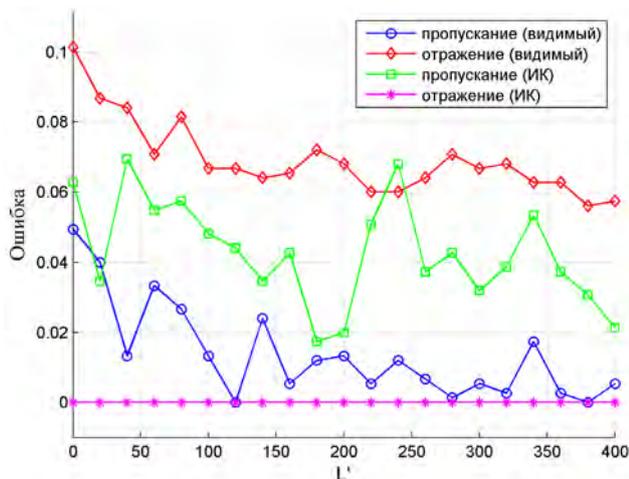


Рис. 7. Зависимость вероятности ошибочного распознавания сорта семян тритикале от объема искусственно сгенерированных данных L' в обучающей выборке для $n = 10$ спектральных участков

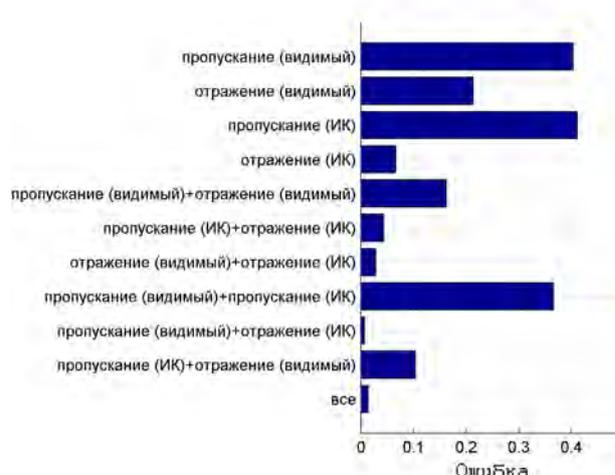
чительно данные, полученные путем проведения измерений. Полученные результаты в виде зависимостей вероятности ошибки от объема искусственно сгенерированных данных представлены на рис. 7.

На рис. 7 видно, что искусственное размножение данных, реализованное с помощью описанного метода, позволило уменьшить процент ошибок распознавания 10 % до 6 % для отражения в видимом диапазоне и с 4,5 % до 0–0,2 % для пропускания в видимом диапазоне.

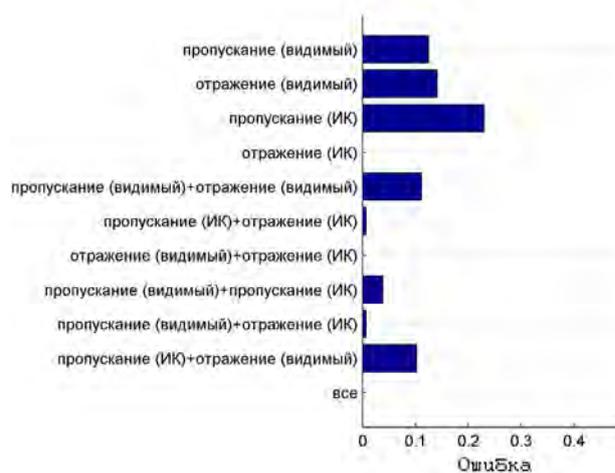
КОМБИНИРОВАННОЕ ИСПОЛЬЗОВАНИЕ СПЕКТРАЛЬНЫХ МЕТОДОВ

В качестве одного из способов повышения достоверности распознавания было предложено комбинирование нескольких спектральных методов. В этом случае на вход нейронной сети подавалось по n результатов измерения спектральных составляющих, полученных одновременно несколькими методами спектрального анализа.

Как видно из рис. 8, комбинирование некоторых спектральных методов уже при $n = 2$ позволяет получить заметно более высокую достоверность распознавания по сравнению с применением этих же методов по отдельности. Так, при отдельном использовании спек-



а)



б)

Рис. 8. Вероятность ошибочного распознавания сорта семян тритикале при использовании различных комбинаций спектральных методов а) для двух участков спектра; б) для четырех участков спектра

тров пропускания и отражения в ИК диапазоне процент ошибок составляет 41 % и 6,7 % соответственно, а при использовании комбинации – 4,4 %. При $n = 4$ в целом наблюдается похожая ситуация за исключением спектра отражения в ИК диапазоне, который обеспечивает максимальную эффективность даже при одиночном использовании.

ЗАКЛЮЧЕНИЕ

В целом анализ полученных результатов позволяет сделать вывод о том, что применение предлагаемых технологий проведения

спектральных измерений в сочетании с нейросетевыми алгоритмами классификации потенциально способно обеспечить достаточно высокую достоверность классификации сортов семян. Использование комбинированных спектральных методов позволило достичь вероятности ошибки, близкой к нулю. При этом достоверность классификации является достаточно высокой даже при малом числе анализируемых спектральных участков и, начиная с определенного порогового значения, не существенно зависит от их числа.

ЛИТЕРАТУРА

1. ГОСТ Р 52325-2005. Семена сельскохозяйственных растений. Сортные и посевные качества. Общие технические условия. – 2005. – 24 с.
2. Конарев А. В., Конарев В. Г., Губарева Н. К., Пенева Т. И. Белки семян как маркеры в решении проблем генетических ресурсов растений, селекции и семеноводства Цитология и генетика. – 2000. – Т. 34. – № 2. – С. 91–104.
3. Романова Ю. А., Губарева Н. К., Конарев А. В., Митрофанова О. П., Ляпунова О. А., Анфилова Н. А., Стрельченко П. П. Исследование коллекции вида пшеницы *Triticum spelta* L. по полиморфизму глатинов Генетика. – 2001. – Т. 37, № 9. – С. 1258–1265.
4. Алматыева Н. В., Губарева Н. К. Характеристика староместных сортов озимой мягкой пшеницы по электрофоретическим спектрам высокомолекулярного глютеина // Аграрная Россия, 2002. – № 3. – С. 24–28.
5. Коренман Я. И., Лисицкая Р. П. Практикум по аналитической химии. Анализ пищевых продуктов. – Воронеж : Воронеж. гос. технол. акад., 2002. – 408 с.
6. Методы биохимического исследования растений / Ред. Ермаков А. И. – Л. : Колос, 1972. – 455 с.
7. Bayea T. M., Pearson T. C., Settles A. M. Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy *Journal of Cereal Science*. – 43 (2006). – P. 236–243.
8. Algazinov E. K., Dryuchenko M. A., Minakov D. A., Sirota A. A., Shul'gin V. A. Methods of measuring the spectral characteristics and identifying the components of grain mixtures in real-time separation systems // *Measurement Techniques*. – 2014. – Vol. 57, No. 1. – P. 54–61.
9. Algazinov E. K., Dryuchenko M. A., Minakov D. A., Sirota A. A., Shul'gin V. A. Hardware–software complex for the analysis of a nonuniform flow of objects in real-time optical sorting systems // *Measurement Techniques*. – 2014. – Vol. 57, No. 5. – P. 509–518.
10. Algazinov E. K., Dryuchenko M. A., Minakov D. A., Sirota A. A., Shul'gin V. A. Methods of identification elements grain mixtures by measuring spectral characteristics in real-time separation systems // *Bulletin of the Voronezh State University. Series: systems analysis and information technology*. – 2013. – № 2. – P. 9–19. (in Russian)
11. Laser sorter // PatentRF № 2489215 C1 // Invention, 2013.
12. Fiber Laser Sorter // PatentRF № 2 521 215 C1 // Invention, 2014.
13. Burns D. and Ciurczak E. (2007). *Handbook of Near Infrared Analysis 3rd Edition*. CRC Press, New York, U.S.A.
14. Agelet L. E., Ellis D. D., Duvick D., Goggi A. S., Hurburgh C. R., Gardner C. A. Feasibility of near infrared spectroscopy for analyzing corn kernel damage and viability of soybean and corn kernels *Journal of Cereal Science*. – 55 (2012). P. 160–165.
15. Janni J., Weinstock B. A., Hagen L., Wright S. 2008. Novel near-infrared sampling apparatus for single kernel analysis of oil content in maize. *Applied Spectroscopy*. – 62 (4). – P. 423–426.
16. Spielbauer G., Armstrong P., Baier J. W., Allen W. B., Richradson K., Shen B., Settles M. 2009. High-throughput nearinfrared reflectance spectroscopy for predicting quantitative and qualitative composition phenotypes of individual maize kernels. *Cereal Chemistry*. – 86 (5). – P. 556–564.
17. Weinstock B. A., Janni J., Hagen L., Wright S. 2006. Prediction of oil and oleic acid concentrations in individual corn (*zea mays* L.) kernels using near-infrared reflectance

hyperspectral imaging and multivariate analysis. *Applied Spectroscopy*. – 60 (1), P. 9–16.

18. *Campbell M. R., Sykes J., Glover D. V.* 2000. Classification of single- and doublemutant corn endosperm genotypes by near-infrared transmittance spectroscopy. *Cereal Chemistry*. – 77 (6). P. 774–778.

19. *Wang D., Dowell F. E., Dempster R.* 2002a. Determining vitreous subclasses of hard red spring wheat using visible/near-infrared spectroscopy. *Cereal Chemistry*. – 79 (3). P. 418–422.

20. *Wang D., Dowell F. E., Dempster R.* 2002a. Determining vitreous subclasses of hard red spring wheat using visible/near-infrared spectroscopy. *Cereal Chemistry*. – 79 (3). – P. 418–422.

21. *Shahin M. A., Symons S. J., Hatcher D. W.* Quantification of Mildew Damage in Soft Red Winter Wheat Based on Spectral Characteristics of Bulk Samples: A Comparison of Visible-Near-Infrared Imaging and Near-Infrared Spectroscopy *Food Bioprocess Technol.* – (2014) 7:224–234.

22. *McGoverin C. M., Engelbrecht P., Geladi P., Manley M.* Characterisation of non-viable whole

barley, wheat and sorghum grains using near-infrared hyperspectral data and chemometrics *Anal Bioanal Chem.* – (2011) 401:2283–2289.

23. *Sundaram J., Kandala C. V., Butts C. L.* Application of near infrared spectroscopy to peanut grading and quality analysis: overview *Sens. & Instrumen. Food Qual.* – (2009) 3:156–164.

24. Сайт компании Perten Instruments. – (<http://www.perten.com>)

25. Сайт компании FOSS. – (<http://www.foss.dk>)

26. Сайт компании DICKEY-john Corporation. – (<http://www.dickey-john.com>)

27. Сайт компании Kett. – (<http://www.kett.com>)

28. Сайт группы компаний «Люмэкс». – (<http://www.lumex.ru>)

29. Тритикале – первая зерновая культура, созданная человеком. / под ред. Ю.Л. Гужова. – М. : Колос, 1978. – 285 с.

30. *Якубке Х.-Д., Ешкайт Х.* Аминокислоты, пептиды, белки. / под ред. Ю. В. Митина. – М. : Мир, 1985. – 456 с.

Донских А. О. – магистрант кафедры технологий обработки и защиты информации, факультет компьютерных наук, Воронежский государственный университет.
E-mail: a.o.donskih@yandex.ru

Минаков Д. А. – к. ф.-м.н., доцент, доцент кафедры физики и химии, Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н. Е. Жуковского и Ю. А. Гагарина» (г. Воронеж).
E-mail: minakov_d_a@mail.ru

Сирота А. А. – д. т. н., профессор, заведующий кафедрой технологий обработки и защиты информации, факультет компьютерных наук, Воронежский государственный университет.
E-mail: sir@cs.vsu.ru

Шульгин В. А. – к. т. н., старший научный сотрудник кафедры информационных систем, Воронежский государственный университет.

Donskikh A. O. – master's degree student, Department of Processing Technology and Information Security, Computer Sciences Faculty, Voronezh State University.
E-mail: a.o.donskih@yandex.ru

Minakov D. A. – Candidate of Physical and Mathematical Sciences, Assistant Professor, Assistant Professor of Department of Physics and Chemistry, Military Educational and Scientific Center of the Air Force «N. E. Zhukovsky and Y. A. Gagarin Air Force Academy» (Voronezh).
E-mail: minakov_d_a@mail.ru

Sirota A. A. – Doctor of Technical Sciences, Professor, Head of Department of Processing Technology and Information Security, Computer Sciences Faculty, Voronezh State University.
E-mail: sir@cs.vsu.ru

Shulgin V. A. – Senior Researcher, Chair of Information Systems, Voronezh State University.