

Установление породы древесины по спектрам диффузного отражения в ближней инфракрасной области с применением линейного дискриминантного анализа

А.Н. Хох¹, В.Б. Звягинцев²

¹ Государственное учреждение «Научно-практический центр Государственного комитета судебных экспертиз Республики Беларусь», Минск 220114, Республика Беларусь

² Учреждение образования «Белорусский государственный технологический университет», Минск 220006, Республика Беларусь

Аннотация. Представлен алгоритм исследования древесных материалов хвойных пород (*Pinus sylvestris* L., *Pinus sibirica* Du Tour, *Larix decidua* Mill., *Abies alba* L. и *Picea abies* (L.) Karst.) методом спектроскопии в ближней инфракрасной области. Древесина представляет собой сложный органический материал, состоящий из целлюлозы, гемицеллюлозы, лигнина и экстрактивных веществ. Данные соединения генерируют полосы поглощения в ближней инфракрасной области, в основном представляющие собой обертоны и полосы комбинации функциональных групп О-Н, N-H и С-Н. Определены оптимальные способы коррекции рассеяния с учетом специфики исследуемых образцов. На основе применения линейного дискриминантного анализа предложена методика по автоматическому установлению видовой принадлежности древесины по спектрам диффузного отражения в ближней инфракрасной области, а также определены ключевые параметры, достаточные для проведения процедуры классификации. Полученные LDA-модели показали высокую предсказательную способность, общая средняя точность классификации достигала 97,6%. Следует констатировать, что метод спектроскопии в ближней инфракрасной области подходит для решения задач, связанных с установлением видовой принадлежности древесины, является эффективным инструментом для разделения древесных пород.

Ключевые слова: древесина, БИК-спектроскопия, видовая принадлежность, метод главных компонент, дискриминантный анализ, хемометрические алгоритмы, судебная экспертиза

Для цитирования: Хох А.Н., Звягинцев В.Б. Установление породы древесины по спектрам диффузного отражения в ближней инфракрасной области с применением линейного дискриминантного анализа // Теория и практика судебной экспертизы. 2022. Т. 17. № 1. С. 50–57. <https://doi.org/10.30764/1819-2785-2022-1-50-57>

Identifying Timber Species by Diffuse Reflection Spectra in the Near-Infrared Region Using a Linear Discriminant Analysis

Anna N. Khokh¹, Vyacheslav B. Zvyagintsev²

¹ State Institution “Scientific and Practical Center of the State Forensic Examination Committee of the Republic of Belarus”, Minsk 220114, Belarus

² Educational Institution “Belarusian State Technological University”, Minsk 220006, Belarus

Abstract. The article presents an algorithm of analysis of coniferous timber materials (*Pinus sylvestris* L., *Pinus sibirica* Du Tour, *Larix decidua* Mill., *Abies alba* L. and *Picea abies* (L.) Karst.) using the spectroscopy method in the near-infrared region. Timber is a complex organic material consisting of cellulose, hemicellulose, lignin, and extractive substances. These compounds generate absorption bands in the near-infrared region which are mainly overtones and combination bands of the O-H, N-H and C-H functional groups. The authors suggest the best ways to correct scattering considering the specific features of the samples under examination. Based on the application of linear discriminant analysis, a method is proposed for the automatic identification of the timber species by diffuse reflection spectra in the near-infrared region, as well as key parameters sufficient for the classification procedure are determined. The obtained LDA-models showed high predictive capability. The overall average classification accuracy reached 97,6%. Based on

the results obtained, it can be stated that the NIR spectroscopy is suitable for solving tasks related to the identification of the timber species; it has proven to be an effective tool for distinguishing wood species.

Keywords: *timber, NIR spectroscopy, species identification, principal component analysis, discriminant analysis, chemometric algorithms, forensic examination*

For citation: Khokh A.N., Zvyagintsev V.B. Identifying Timber Species by Diffuse Reflection Spectra in the Near-Infrared Region Using a Linear Discriminant Analysis. *Theory and Practice of Forensic Science*. 2022. Vol. 17. No. 1. P. 50–57. (In Russ.). <https://doi.org/10.30764/1819-2785-2022-1-50-57>

Введение

Целью экспертизы древесины, как правило, является установление ее видовой принадлежности, что чаще всего не представляет сложностей и решается с помощью классических методов макродиагностики. В ряде случаев древесина для исследования поступает в виде опилок, стружек или небольших частиц в количестве, недостаточном для идентификации породы на основе только макроскопических характеристик [1, р. 48]. При этом возникает необходимость изучения ее анатомического строения, а иногда – проведения молекулярно-генетических анализов [2, с. 84], которые трудоемки и требуют применения специальных знаний. В связи с этим в последние годы востребованы и другие методы [3–5], позволяющие проводить быструю и эффективную экспресс-диагностику.

Один из таких методов автоматизированной идентификации породы древесины – спектроскопия в ближней инфракрасной области (БИК-спектроскопия) инструментальный неразрушающий метод количественного и качественного анализа, требующий обязательной статистической обработки [6, с. 100].

Ориентировочно БИК-область расположена между $13\,300\text{ см}^{-1}$ и $4\,000\text{ см}^{-1}$ (от 740–780 до 2500 нм) [7, с. 826], что делает возможным использование так называемого диффузного отражения, при котором свет, проходящий через образец, рассеивается внутри него, отражается и затем детектируется. Таким образом можно получать информацию не только о химическом составе, но и о физической структуре образца. В этом диапазоне лежат многочисленные обертоны и комбинационные полосы, обусловленные в основном колебаниями связей атома водорода с атомами других химических элементов, например, С-Н, О-Н и Н-Н.

Перекрывающиеся БИК-спектры являются неспецифичными из-за низкой интенсивности и поэтому плохо различимы. Однако они содержат важную информацию о молекулярной структуре образца, которую можно извлечь с использованием современных хемометрических алгоритмов анализа.

Цель работы заключалась в разработке алгоритма идентификации хвойных пород древесины с применением линейного дискриминантного анализа (LDA).

Ранее было выполнено формальное независимое моделирование аналогий классов (SIMCA) для девяти видов хвойных [8]. Результаты последующих исследований показали, что SIMCA целесообразно использовать при бинарной классификации, а в случаях многоклассовой классификации следует применять LDA.

Объекты и методы

Объектами исследования служили образцы сосны обыкновенной (*Pinus sylvestris* L.), сосны сибирской кедровой (*Pinus sibirica* Du Tour), лиственницы европейской (*Larix decidua* Mill.), пихты белой (*Abies alba* L.), ели обыкновенной (*Picea abies* (L.) Karst.) размером 20 60 150 мм (вдоль волокон) из ксилотеки УО «Белорусский государственный технологический университет» и натурной коллекции НПЦ «Государственный комитет судебных экспертиз Республики Беларусь», а также реальные объекты экспертных исследований в виде древесной муки с размером частиц около 1 мм и опилок с размером частиц 2–5 мм.

От 50 до 70 исследованных образцов для каждой породы из ксилотеки было использовано в качестве обучающей выборки (набора обучения). При формировании выборки важно было исследовать максимум имеющихся образцов для получения максимума информации относительно допусти-

мых вариаций. 50 объектов экспертных исследований (по 10 каждой породы) использовали в качестве тестовой выборки.

Перед исследованием все образцы подвергались естественной сушке (влажность не превышала 12 %).

БИК-спектры диффузного отражения регистрировали на спектрометре MicroNIR OnSite (VIAMI, США) в диапазоне 950–1650 нм без предварительной специальной пробоподготовки. Время интегрирования – 10,4 мс, число счетов – 200, разрешение – 2 см⁻¹, количество сканов – 64. Для каждого образца проводили по 20 последовательных измерений на продольных разрезах древесины, которые в дальнейшем усреднялись до одного спектра. Полученные спектры статистически обрабатывали с использованием пакета программ The Unscrambler X v.10.4.1 (CAMO, США). Для оценки полученных результатов применяли метод главных компонент, для построения классификационных моделей – LDA.

Любой БИК-спектр содержит большое количество избыточной информации («шума»), что требует перед построением классификационных моделей его предварительной обработки. Это позволяет свести к минимуму вариации, связанные с изменением базовой линии вследствие аддитивного и мультипликативного рассеяния, и сохранить полезную информацию для идентификации видов.

Результаты и их обсуждение

Результаты ранее проведенных экспериментальных исследований показали, что для решения экспертных задач по установлению видовой принадлежности древесных пород необходимо вычисление производных второго порядка по Савицкому-Голею (полином второго порядка, 7 сглаживающих точек) [8].

На рисунке 1 показан внешний вид БИК-спектров пяти исследованных древесных пород без обработки, на рисунке 2 – спектры после вычисления производных.

Полученные спектроскопические данные обучающего набора анализировали с помощью PCA [9], в соответствии с которым при получении многомерной выборки (в нашем случае 270 наблюдений по 125 признакам (длинам волн) каждый из 270 образцов представляется точкой в 125-мерном пространстве признаков. Это пространство преобразуется таким образом, что основные особенности образцов наглядно представляются на плоскости, координатами которой являются первые, главные компоненты: обычно главная компонента 1, главная компонента 2 и главная компонента 3. При этом происходит свертывание пространства признаков от 125-ти до 3-мерного, что позволяет проводить четкую дифференциацию образцов в координатах главных компонент. Нахождение этих координат и составляет сущность метода.

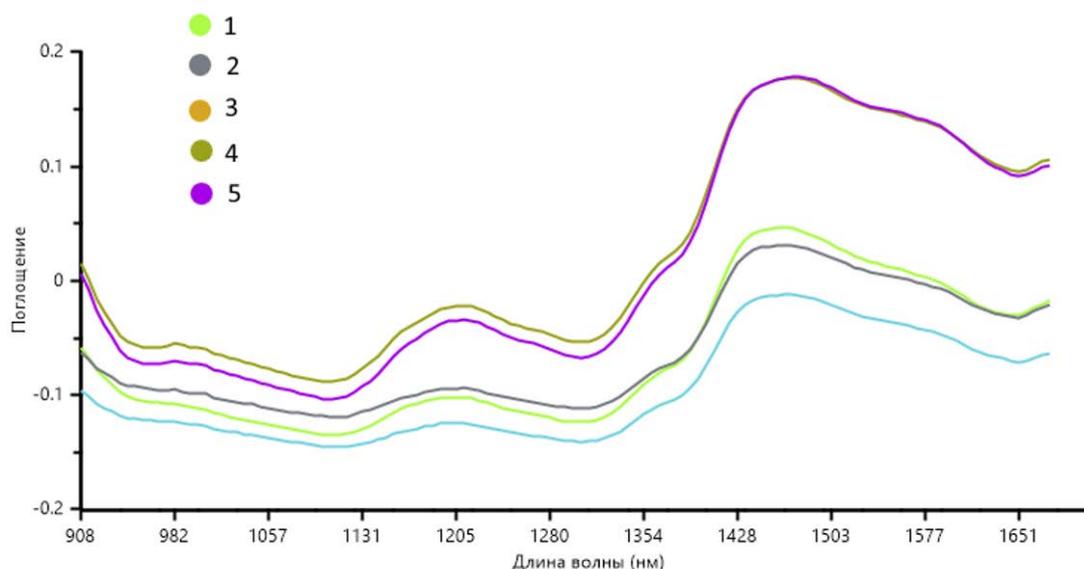


Рис. 1. Необработанные средние БИК-спектры для каждого вида: 1 – сосна обыкновенная; 2 – сосна сибирская кедровая; 3 – лиственница европейская; 4 – пихта белая; 5 – ель обыкновенная

Fig. 1. Raw average NIR-spectra for each species: 1 – Scots pine; 2 – Siberian cedar pine; 3 – European larch; 4 – Silver fir; 5 – Norway spruce

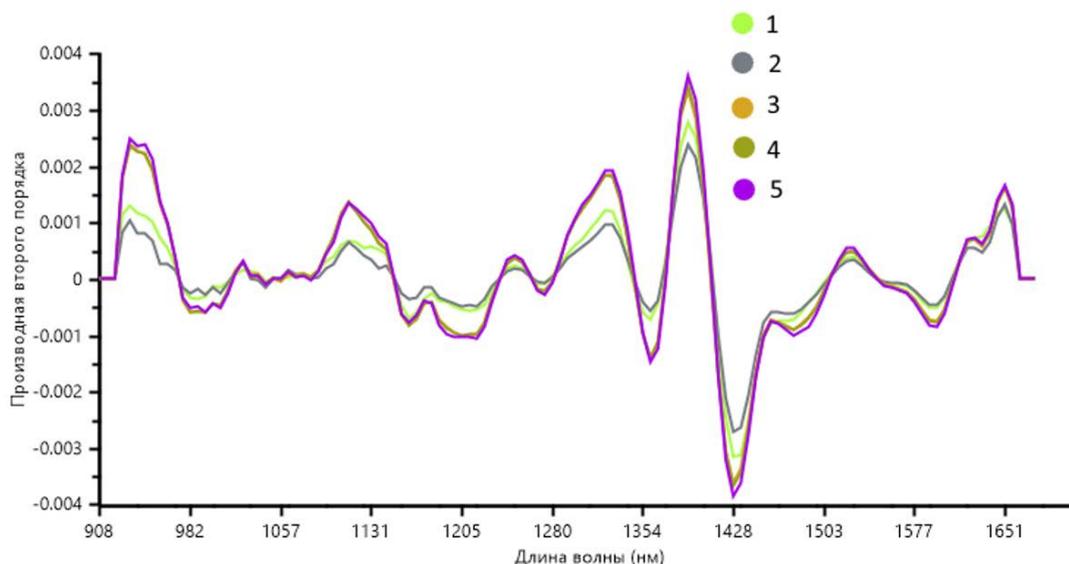


Рис. 2. Средние БИК-спектры после вычисления производных второго порядка (здесь и далее пункты 1–5 соответствуют обозначениям на рис. 1)

Fig. 2. Average near-infrared spectra after calculation of the second-order derivatives (points 1–5 henceforth correspond to the symbols in fig. 1)

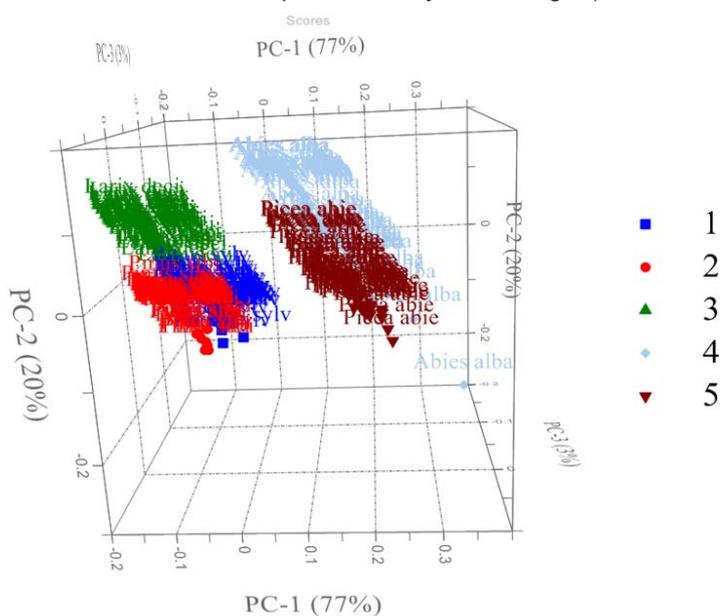


Рис. 3. График МГК-счетов для 270 исследованных образцов

Fig. 3. The PCA score plot for 270 tested samples

На основании проведенного анализа все исследованные образцы объединили в 2 класса: 1) образцы сосны обыкновенной, сосны сибирской кедровой и лиственницы европейской; 2) образцы пихты белой и ели обыкновенной.

В каждом из классов отчетливо выделяются подклассы, соответствующие исследованным древесным породам. Это дало возможность провести линейный дискриминантный анализ и классифицировать исследуемые объекты из обучающего набора по принципу максимального сходства. Следует отметить, что количество образцов,

необходимое для разработки LDA-модели, может варьироваться в зависимости от типа анализируемого объекта и составлять от 30–40 до 150 и более образцов [10, 11]. В проводимых исследованиях минимальное количество образцов составило 50, максимальное – 70.

На рисунке 4 показаны результаты LDA с использованием только двух главных компонент: PCA-1 и следующей за ней PCA-2. Точность классификации обучающей выборки составила 99,2 %, что позволяет использовать полученную модель для классификации новых образцов, порода которых

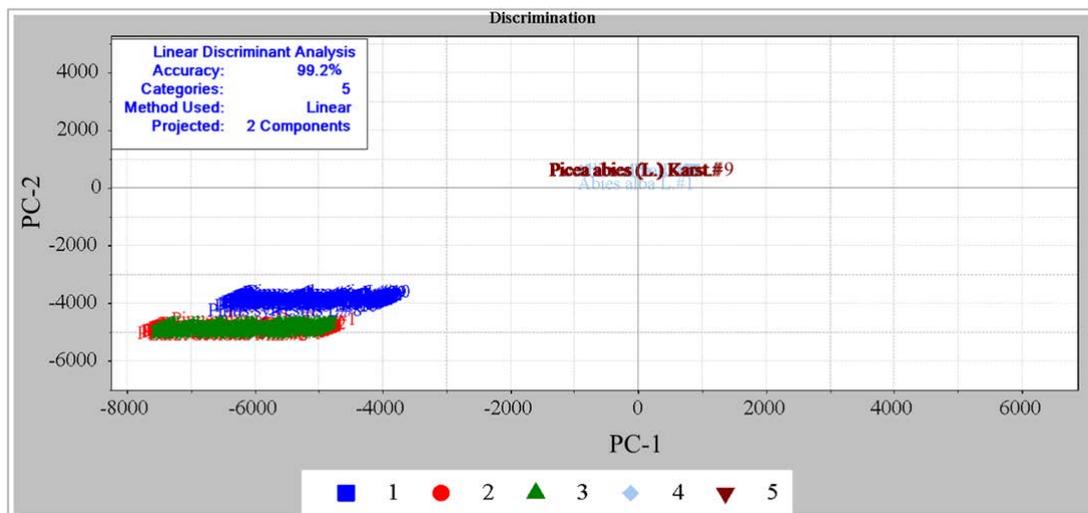


Рис. 4. Двумерное пространство (PC1, PC2) с пятью древесными породами
Fig. 4. Two-dimensional space (PC1, PC2) with 5 timber species

неизвестна. Результат свидетельствует о принципиальной возможности использования двух главных компонент для проведения линейного дискриминантного анализа БИК-спектров хвойных древесных пород. Построенная калибровочная зависимость имеет линейный характер с коэффициентом корреляции $r = 0,9999$ и коэффициентом детерминации $R^2 = 99,98$.

По результатам дискриминации установлено, что для сосны обыкновенной, сосны сибирской кедровой, лиственницы европейской и ели обыкновенной было верно классифицировано 220 (100,0 %) наблюдений, а для пихты белой – 48 (96 %) (табл. 1).

Для каждой древесной породы рассчитаны значения пяти функций дискриминации (D1–D5). Показано, что отнесение объектов к классу определяется значениями функций D3–D5, т. е. именно они обладают большими дискриминирующими возможностями,

а к классу 2 – D1 и D2 соответственно. В свою очередь, на разделение объектов на подклассы оказывают влияние оставшиеся функции. Так, для сосны обыкновенной значение D3 составляет $400,8 \pm 16$, в то время как для сосны сибирской кедровой – $50,1 \pm 7$, лиственницы европейской – $-2,4 \pm 1,6$ (рис. 5).

Полученную LDA-модель использовали для классификации реальных экспертных образцов. При этом некорректно отнесенные образцы из обучающей выборки были предварительно исключены из модели. С учетом рассчитанных значений канонических дискриминантных функций экспертные образцы отнесены к конкретной породе (табл. 2).

Видовую принадлежность экспертных образцов подтвердили микроскопические исследования. Точность классификации тестовой выборки составила 96 %. Постро-

Таблица 1. Результат классификации обучающей выборки
Table 1. The result of the training sample classification

Древесная порода		Предсказанная принадлежность к группе				
		1	2	3	4	5
Частота	1	70	0	0	0	0
	2	0	50	0	0	0
	3	0	0	50	0	0
	4	0	0	0	48	0
	5	0	0	0	2	50
%	1	100	0	0	0	0
	2	0	100	0	0	0
	3	0	0	100	0	0
	4	0	0	0	96	0
	5	0	0	0	4	100

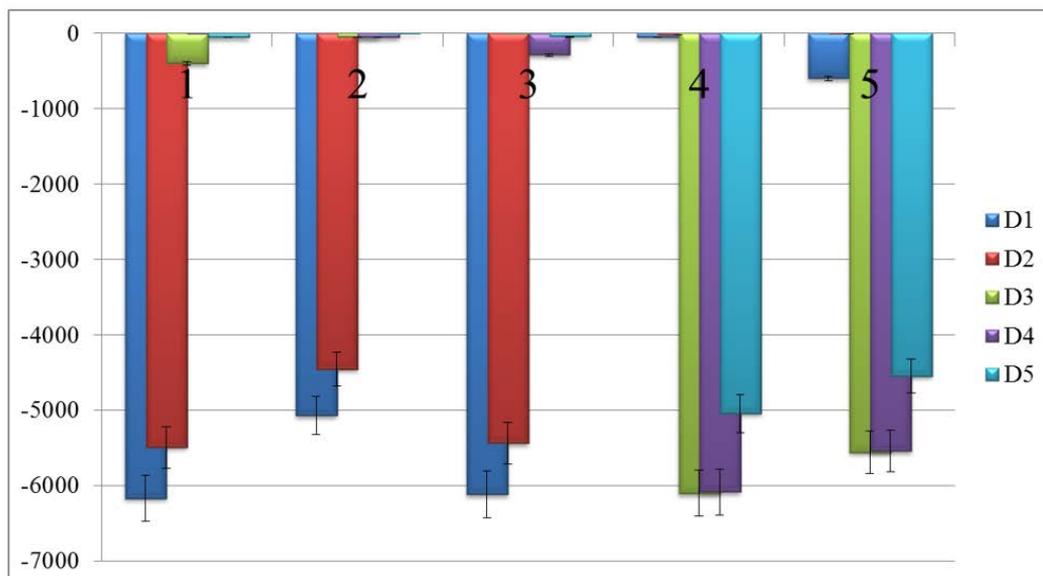


Рис. 5. Канонические дискриминантные функции для пяти древесных пород
Fig. 5. Canonical discriminant functions for 5 timber species

Таблица 2. Результат классификации тестовой выборки
Table 2. The result of the test sample classification

Древесная порода		Предсказанная принадлежность к группе				
		1	2	3	4	5
Частота	1	10	0	0	0	0
	2	0	10	0	0	0
	3	0	0	10	0	0
	4	0	0	0	9	1
	5	0	0	0	1	9
%	1	100	0	0	0	0
	2	0	100	0	0	0
	3	0	0	100	0	0
	4	0	0	0	90	10
	5	0	0	0	10	90

енная тестовая зависимость также имела линейный характер с коэффициентом корреляции $r = 0,9998$ и $R^2 = 99,6$.

Как и в случае обучающей выборки, все образцы сосны обыкновенной, сосны сибирской кедровой и лиственницы европейской были классифицированы правильно, ошибки были выявлены для пихты белой и ели обыкновенной (1 неправильно классифицированный образец на 10 исследованных). При этом один из неправильно классифицированных образцов относился к древесной муке, а другой – к опилкам. Это позволяет предположить, что фракционный состав древесины не оказывает существенного влияния на БИК-спектры и установление видовой принадлежности.

Заключение

Разработанный на основе применения линейного дискриминантного анализа БИК-спектров алгоритм позволил провести объективную идентификацию пяти хвойных древесных пород со схожими спектрометрическими профилями с достаточно высокой точностью (97,6 %). Его основным преимуществом является простота реализации, возможность автоматического принятия решения о принадлежности экспертных образцов к определенной породе. Полученные результаты открывают перспективы к использованию изложенных в работе подходов и для других древесных пород. Вместе с тем важным условием для правильной интерпретации результатов является наличие базы эталонных (референтных) БИК-

спектров различных пород древесины, что позволит сравнивать неизвестные образцы

с существующими в базе, и на этой основе проводить классификацию.

СПИСОК ЛИТЕРАТУРЫ

1. Toscano P., Iannotta N., Scalercio S. Botanical and Agricultural Aspects: Agronomic Techniques and Orchard Management. In: Muzzalupo I., Micali S. (Eds.). *Agricultural and Food Biotechnologies of Olea europaea and Stone Fruits*. Bentham, 2015. P. 3–73. <https://doi.org/10.2174/9781608059935115010004>
2. Ohyama M., Baba K., Itoh T. Wood Identification of Japanese *Cyclobalanopsis* Species (Fagaceae) Based on DNA Polymorphism of the Intergenic Spacer between *trnT* and *trnL* 5' exon // *Journal of Wood Science*. 2001. Vol. 47. No. 2. P. 81–86. <https://doi.org/10.1007/BF00780554>
3. Sharma V., Yadav J., Kumar R., Tesarova D., Ekielski A., Mishra P. On the Rapid and Non-Destructive Approach for Wood Identification Using ATR-FTIR Spectroscopy and Chemometric Methods // *Vibrational Spectroscopy*. 2020. Vol. 110. 103097. <https://doi.org/10.1016/j.vibspec.2020.103097>
4. Kanayama H., Ma T., Tsuchikawa S., Inagaki T. Cognitive Spectroscopy for Wood Species Identification: Near Infrared Hyperspectral Imaging Combined with Convolutional Neural Networks // *Analyst*. 2019. Vol. 144. No. 21. P. 6438–6446. <https://doi.org/10.1039/C9AN01180C>
5. Adedipe O.E., Dawson-Andoh B., Slahor J., Osborn L. Classification of Red Oak (*Quercus rubra*) and White Oak (*Quercus alba*) Wood Using a Near Infrared Spectrometer and Soft Independent Modelling of Class Analogies // *Journal of Near Infrared Spectroscopy*. 2008. Vol. 16. No. 1. P. 49–57. <https://doi.org/10.1255/jnirs.760>
6. Manley M., Baeten V. Spectroscopic Technique: Near Infrared (NIR) Spectroscopy. In: *Modern Techniques for Food Authentication*. 2nd ed. Academic Press, 2018. P. 51–102. <https://doi.org/10.1016/B978-0-12-814264-6.00003-7>
7. Aenugu H.P.R., Kumar D.S., Srisudharson N.P., Ghosh S., Banji D. Near Infra Red Spectroscopy – An Overview // *International Journal of ChemTech Research*. 2011. Vol. 3. No. 2. P. 825–836.
8. Хох А.Н., Звягинцев В.Б. Установление видовой принадлежности древесины хвойных пород с помощью спектроскопии в ближней инфракрасной области // *Вопросы криминологии, криминалистики и судебной экспертизы*. 2020. № 1. С. 161–167.
9. Paul L.C., Suman A.A., Sultan N. Methodological Analysis of Principal Component Analysis (PCA) Method // *International Journal of Computational Engineering & Management*. 2013. Vol. 16. No. 2. P. 32–38.

REFERENCES

1. Toscano P., Iannotta N., Scalercio S. Botanical and Agricultural Aspects: Agronomic Techniques and Orchard Management. In: Muzzalupo I., Micali S. (Eds.). *Agricultural and Food Biotechnologies of Olea europaea and Stone Fruits*. Bentham, 2015. P. 3–73. <https://doi.org/10.2174/9781608059935115010004>
2. Ohyama M., Baba K., Itoh T. Wood Identification of Japanese *Cyclobalanopsis* Species (Fagaceae) Based on DNA Polymorphism of the Intergenic Spacer between *trnT* and *trnL* 5' exon. *Journal of Wood Science*. 2001. Vol. 47. No. 2. P. 81–86. <https://doi.org/10.1007/BF00780554>
3. Sharma V., Yadav J., Kumar R., Tesarova D., Ekielski A., Mishra P. On the Rapid and Non-Destructive Approach for Wood Identification Using ATR-FTIR Spectroscopy and Chemometric Methods. *Vibrational Spectroscopy*. 2020. Vol. 110. 103097. <https://doi.org/10.1016/j.vibspec.2020.103097>
4. Kanayama H., Ma T., Tsuchikawa S., Inagaki T. Cognitive Spectroscopy for Wood Species Identification: Near Infrared Hyperspectral Imaging Combined with Convolutional Neural Networks. *Analyst*. 2019. Vol. 144. No. 21. P. 6438–6446. <https://doi.org/10.1039/C9AN01180C>
5. Adedipe O.E., Dawson-Andoh B., Slahor J., Osborn L. Classification of Red Oak (*Quercus rubra*) and White Oak (*Quercus alba*) Wood Using a Near Infrared Spectrometer and Soft Independent Modelling of Class Analogies. *Journal of Near Infrared Spectroscopy*. 2008. Vol. 16. No. 1. P. 49–57. <https://doi.org/10.1255/jnirs.760>
6. Manley M., Baeten V. Spectroscopic Technique: Near Infrared (NIR) Spectroscopy. In: *Modern Techniques for Food Authentication*. 2nd ed. Academic Press, 2018. P. 51–102. <https://doi.org/10.1016/B978-0-12-814264-6.00003-7>
7. Aenugu H.P.R., Kumar D.S., Srisudharson N.P., Ghosh S., Banji D. Near Infra Red Spectroscopy – An Overview. *International Journal of ChemTech Research*. 2011. Vol. 3. No. 2. P. 825–836.
8. Khokh A.N., Zviagintsev V.B. Implementation of Near-Infrared Spectroscopy to Identify Coniferous Wood Breeds. *Issues of Criminology, Forensics and Forensic Examination*. 2020. No. 1. P. 161–167. (In Russ.).
9. Paul L.C., Suman A.A., Sultan N. Methodological Analysis of Principal Component Analysis (PCA) Method. *International Journal of Computational Engineering & Management*. 2013. Vol. 16. No. 2. P. 32–38.

-
10. Boedeker P., Kearns N.T. Linear Discriminant Analysis for Prediction of Group Membership: A User-Friendly Primer // *Advances in Methods and Practices in Psychological Science*. 2019. Vol. 2. No. 3. P. 250–263.
<https://doi.org/10.1177/2515245919849378>
11. Tharwat A., Gaber T., Ibrahim A., Hassanien A.E. Linear Discriminant Analysis: A Detailed Tutorial // *AI Communications*. 2017. Vol. 30. No. 2. P. 169–190.
<https://doi.org/10.3233/AIC-170729>
10. Boedeker P., Kearns N.T. Linear Discriminant Analysis for Prediction of Group Membership: A User-Friendly Primer. *Advances in Methods and Practices in Psychological Science*. 2019. Vol. 2. No. 3. P. 250–263.
<https://doi.org/10.1177/2515245919849378>
11. Tharwat A., Gaber T., Ibrahim A., Hassanien A.E. Linear Discriminant Analysis: A Detailed Tutorial. *AI Communications*. 2017. Vol. 30. No. 2. P. 169–190.
<https://doi.org/10.3233/AIC-170729>

ИНФОРМАЦИЯ ОБ АВТОРАХ

Хох Анна Николаевна – заведующий лабораторией исследования материалов, веществ и изделий НПЦ «Государственный комитет судебных экспертиз Республики Беларусь»;
e-mail: 1ann1hoh@gmail.com

Звягинцев Вячеслав Борисович – к. б. н., доцент, заведующий кафедрой лесозащиты и древесиноведения УО «Белорусский государственный технологический университет»;
e-mail: mycolog@tut.by

Статья поступила: 16.06.2021

После доработки: 12.10.2021

Принята к печати: 10.01.2022

ABOUT THE AUTHORS

Khokh Anna Nikolaevna – Head of Laboratory of Materials, Substances and Products Research, Scientific and Practical Center of the State Forensic Examination Committee of the Republic of Belarus;
e-mail: 1ann1hoh@gmail.com

Zvyagintsev Vyacheslav Borisovich – Candidate of Biology, Assistant Professor, Head of Department of Forest Protection and Wood Science, Belarusian State Technological University;
e-mail: mycolog@tut.by

Received: June 16, 2021

Revised: October 12, 2021

Accepted: January 10, 2022