

В.Н. Павлов¹, А.Р. Билялов¹, Р.Ф. Гильманова¹, Р.Р. Якупов¹,
А.С. Ковтуненко², А.Т. Бикмеев², С.Г. Карчевский³, Р.Р. Халилов³
**ИСПОЛЬЗОВАНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ МЕТОДОВ
ОБРАБОТКИ ДАННЫХ РАМАН-СПЕКТРОСКОПИИ
ДЛЯ ДИАГНОСТИКИ ЗЛОКАЧЕСТВЕННЫХ ОПУХОЛЕЙ**

¹ФГБОУ ВО «Башкирский государственный медицинский университет»
Минздрава России, г. Уфа

²ФГБОУ ВО «Уфимский государственный авиационный технический университет», г. Уфа

³ГУП «Институт нефтехимпереработки Республики Башкортостан», г. Уфа

Целью представленного исследования является разработка универсального алгоритма анализа спектрограмм комбинационного рассеяния света (раман-спектроскопии) при помощи интеллектуальных методов обработки данных.

Представленный алгоритм реализован авторами на языке программирования Python и представляет собой последовательность математических преобразований раман-спектров с целью автоматического распознавания нормальных и опухолевых тканей.

Выявлено, что оптимальным алгоритмом является вычисление базовой линии спектра итерационной полиномиальной регрессией, последующим частотным анализом спектра, определением главных компонент спектра и машинным обучением. Достоверность идентификации и классификации нормальной ткани и ткани злокачественной опухоли составила от 97,5 до 98%.

Ключевые слова: комбинационное рассеяние света, раман-спектроскопия, злокачественная опухоль, частотный анализ, deep learning.

V.N. Pavlov, A.R. Bilyalov, R.F. Gilmanova, R.R. Yakupov,
A.S. Kovtunencko, A.T. Bikmееv, S.G. Karchevsky, R.R. Khalilov

**THE USE OF INTELLIGENT DATA PROCESSING TECHNIQUES OF RAMAN
SPECTROSCOPY FOR THE DIAGNOSIS OF MALIGNANT TUMORS**

The purpose of this study is to develop a universal algorithm for analyzing the spectrograms of Raman scattering (Raman-spectroscopy) using intelligent data processing methods.

The presented algorithm is realized by the authors in the programming language Python and represents a sequence of mathematical transformations of Raman spectra for the purpose of automatic recognition of normal and tumor tissues.

It is found that the optimal algorithm is the calculation of the baseline spectrum by iterative polynomial regression, subsequent frequency analysis of the spectrum, determination of the main components of the spectrum, and machine learning. The reliability of identification and classification of normal tissue and tissue of a malignant tumor was 97.5 to 98%.

Key words: Raman scattering, Raman spectroscopy, malignant tumor, frequency analysis, deep learning.

Ранняя диагностика рака имеет решающее значение для своевременного, эффективного и в конечном итоге успешного лечения. Изменения структуры и концентрации основных биохимических веществ клеток и тканей начинаются задолго до появления клинических симптомов злокачественной опухоли. В связи с этим спектроскопия, позволяющая обнаруживать изменения химических связей в молекулах, является потенциальным инструментом ранней диагностики опухолей. В качестве метода молекулярной спектроскопии спектроскопия комбинационного рассеяния света (раман-спектроскопия) может обнаруживать индуцированные раком изменения молекулярной структуры и состава ткани [3].

Спектроскопия комбинационного рассеяния света (раман-спектроскопия) представляет собой оптический метод, который сегодня рассматривается для характеристики множества заболеваний, в том числе в приложениях *in vivo*, демонстрирующих различия между доброкачественными и злокачественными опухолевыми тканями [1]. В связи с этим становится актуальной проблема разра-

ботки высокоточной методики регистрации спектрограмм, ведущей к разработке точных моделей, которые впоследствии могут быть использованы вместе с инструментами *in vivo* для проведения оценки состояния тканей и определения границ опухоли.

Интерес к спектроскопии биологических тканей быстро растет, поскольку как клинические, так и неклинические исследователи признают, что вибрационные спектроскопические методы, инфракрасные (ИК) и спектроскопические методы комбинационного рассеяния потенциально могут стать неинвазивными инструментами для диагностики заболеваний. Однако существует значительный пробел в разработке методов анализа спектрограмм, поскольку детали характерных пиковых частот и их определения, которые могут быть отнесены к конкретным функциональным химическим группам, присутствующим в биологических тканях, не полностью поняты. Кроме того, на сегодняшний день не существует единого источника, который учитывал бы как инфракрасные (ИК), так и комбинаторные спектроскопические исследования биологических тканей,

поскольку исследователи должны полагаться на ряд источников исследований, и в большинстве случаев интерпретация спектральных данных существенно различается [5].

Целью данного исследования является разработка методики анализа спектрограмм комбинационного рассеяния света (раман-спектроскопии) при помощи интеллектуальных методов обработки данных.

Материал и методы

В работе использовали образцы опухолевой и нормальной тканей, полученные после оперативного удаления злокачественных и доброкачественных опухолей у 12 пациентов отделений урологии и онкологии Клиники БГМУ. Все пациенты находились на стационарном лечении. В процессе проведения работы от всех пациентов, включенных в исследование, было получено информированное согласие. Патологическая постановка диагноза первичной опухоли (рТ) определялась из образца опухолевой ткани в соответствии с системой классификации TNM (7-е издание) на основании клинических данных.

В операционном отделении проводилась секция послеоперационного препарата и его маркировка в соответствии со стандартами Клиники БГМУ. Образцы тканей вытирали стерильной тканью и переносили в раствор 0,9% NaCl. В лабораторию оптической спектроскопии образцы доставляли в течение 2 часов после окончания операции.

Для получения рамановских спектров использовался аппарат Horiba XploRA plus, Model BX 41 TF (Horiba, Ltd., Япония). Для исследования биологических тканей использовался лазер длиной волны 785 нм, мощностью до 100 мВт. Всего было получено 170 спектрограмм комбинационного рассеяния света.

Данные со спектрографа обрабатывали на программном комплексе – пакет LabSpec v.6.4.4. В данном программном пакете выполнялись графический анализ данных и конвертация полученных измерений в форматы SPC (файл двоичного формата для хранения данных спектроскопии) и CSV для дальнейшей обработки с помощью разработанного алгоритма анализа. Последующий графический анализ проводился с применением программного обеспечения Spectragryph v. 1.2.8.

Результаты и обсуждение

На сегодняшний день имеются попытки использования спектроскопии комбинационного рассеяния Фурье, конфокальной рамановской микроспектроскопии, резонансной спектроскопии комбинационного рассеяния, спектроскопии комбинационного рассеяния

на поверхности для ранней диагностики опухолевой трансформации клеток [7].

Рамановская спектроскопия представляет собой вибрационно-спектроскопический метод, который используется для оптического зондирования молекулярных изменений, связанных с пораженными тканями.

Спектры комбинационного рассеяния получают путем указания монохроматического лазерного луча на образец и представляют собой графики заданной интенсивности в зависимости от разности энергий падающего и рассеянного фотонов. Когда свет взаимодействует с молекулой, большая часть света рассеивается с той же частотой, что и падающий свет (упругое рассеяние). Как уже упоминалось, только небольшая фракция рассеивается на другой длине волны (неупругое или комбинационное рассеяние) из-за энергии света, изменяющей колебательное состояние молекулы. Потери (или усиления) в энергии фотонов соответствуют разности в окончательном и начальном уровнях колебательной энергии молекул, участвующих во взаимодействии. Полученные спектры характеризуются сдвигами в волновых числах (обратных длине волны в см^{-1}) от частоты падающего излучения. Разность частот между рассеянным и рамановским рассеянием называется рамановским сдвигом, который уникален для отдельных молекул, измеряется детектором и представлен как $\frac{1}{\text{см}}$. Рамановские пики спектрально узкие и во многих случаях могут быть связаны с вибрацией определенной химической связи (или одной функциональной группы) в молекуле. Поскольку эти колебательные переходы связаны с соответствующими молекулярными связями, они уникальны для молекулы и генерируют четкие спектры комбинационного рассеяния (как отпечатки пальцев) [2,4].

В спектроскопических исследованиях точное определение пиков может оказать большое влияние на надежность результатов. В процессе изучения современных публикаций по данной тематике стало очевидным, что большинство ученых в основном использовали для расчетов данные спектров. Однако отсутствие надежной и подробной базы данных, охватывающей большинство известных спектральных пиков, точная идентификация химического состава биологической ткани является чересчур трудоемкой задачей.

Из литературы известно, что эффективность анализа спектрограмм и обнаружения типовых для опухолевой ткани изменений, достигаемая с помощью различных методов обработки сигналов, различается. Можно за-

метить, что метод главных компонент (PCA) в сочетании с методом опорных векторов (SVM) дает наивысшую точность (99,9%), за которым следуют комбинация PCA и искусственных нейронных сетей (98%) и логистический регрессионный анализ (LRA) – 97%. Метод обработки PCA + LDA (латентное размещение Дирихле) является самым популярным (45%), который также может сопровождаться PCA-ANN (33%) и SVM (22%).

В настоящей работе предлагается новый алгоритм обработки сигналов для автоматического распознавания раман-спектров включающий в себя следующие этапы: предварительная обработка сигнала, представление сигнала в требуемом виде, выявление особенностей, удаление ошибок, отбор и классификация особенностей. Алгоритм реализован на языке Python в виде оригинального программного обеспечения.

Предварительная обработка спектра проводилась методом корректировки базовой линии итерационной полиномиальной аппроксимацией (рис. 1). Это один из наиболее часто используемых методов для корректировки базовой линии в спектроскопии.

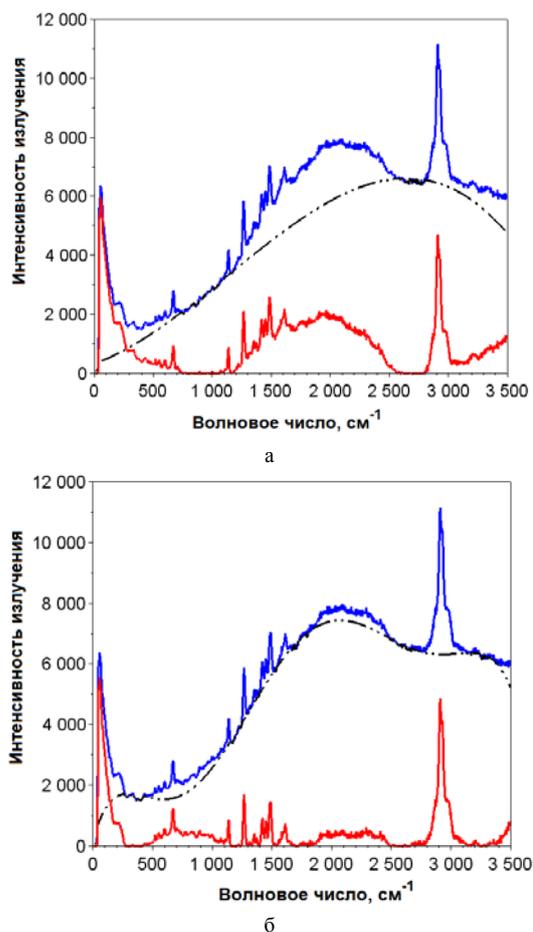


Рис. 1. Исходная спектрограмма (верхняя сплошная кривая) и найденная базовая линия (пунктирная кривая в середине): а – аппроксимация 3-го порядка, б – аппроксимация 6-го порядка и скорректированный спектр (серая нижняя линия)

На рис.1а показано использование аппроксимации третьего порядка, а на рис.1б – шестого порядка. Видно, что второй вариант дает более точный результат.

Первичное изображение на основе спектра строилось с использованием интегральных преобразований. Для выявления особенностей спектра, в качестве которого рассматриваются пики большой высоты, и устранения шума в экспериментальных данных мы воспользовались разложением скорректированного спектра в ограниченный функциональный ряд по функциям Добеши четвертого порядка. Данное разложение позволяет проводить многоуровневый частотный анализ спектра, абстрагируясь от физического смысла пиков на спектре.

На рис. 2 дано графическое представление разложения спектрограмм по функциям Добеши после корректировки базовой линии. Низкочастотные составляющие (расположены в самом низу рис. 2) имеют значительные отличия, тогда как в средней части спектра можно наблюдать повторяющиеся картины. Таким образом, можно сделать вывод о теоретической возможности классификации различных тканей на основе более глубокого анализа частотного разложения их раман-спектров. Результатом разложения каждого спектра является набор коэффициентов разложения, совокупность которых представляет собой образ в многомерном признаковом пространстве.

Для снижения размерности признакового пространства и выявления наиболее характерных абстрактных свойств спектров на основе корреляционного анализа всей выборки производились выбор главных компонент и проецирующее линейное преобразование образов из исходного признакового пространства в редуцированное. Метод анализа главных компонент (PCA) состоит из следующих основных этапов. Во-первых, PCA преобразует данные входных функций в ортогональное пространство с использованием ортогонального линейного преобразования. Результатом являются ортогональные компоненты, известные как главные компоненты. Во-вторых, главные компоненты организованы в соответствии с их дисперсией. Разница является мерой вариабельности распределения выборок и выражается как среднее квадратическое отклонение каждого образца от его среднего значения.

Для решения задачи распознавания в редуцированном признаковом пространстве использовались современные вычислитель-

ные технологии и технологии искусственного интеллекта deep learning, основанные на методах машинного обучения и глубокого анализа обучающих выборок. На рис. 3 показаны

результаты обучения при 10000 эпохах (50 эпох на точку графика). При этом оценка корректности распознавания на наборе данных для обучения составила 97,5%.

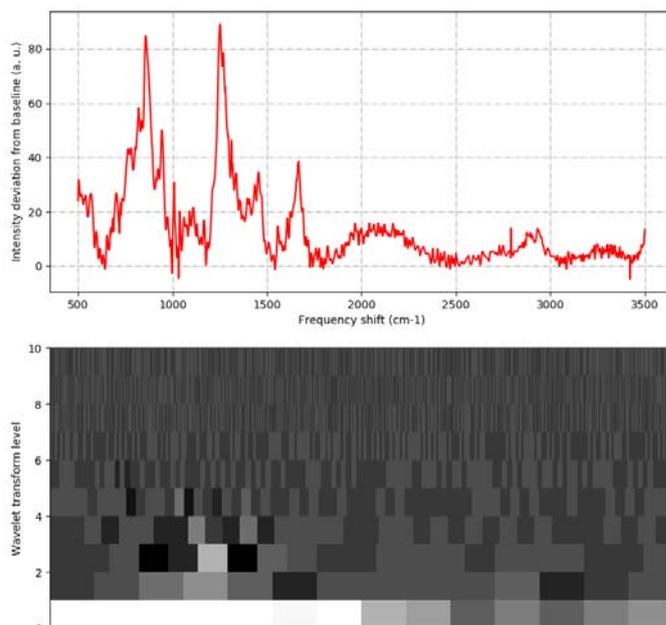


Рис. 2. Графическое представление вейвлет-преобразования раман-спектрограмм

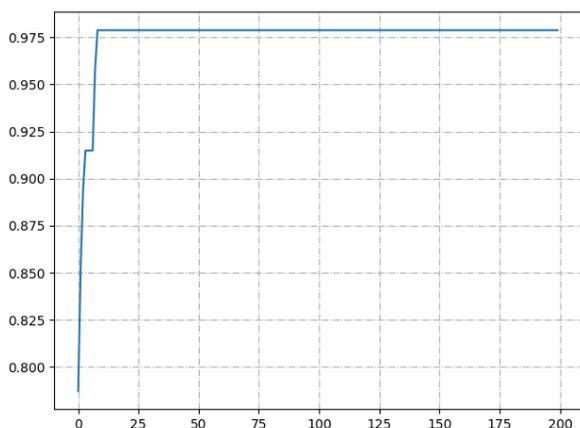


Рис. 3. Изменение ошибки в процессе обучения ($\delta=0,00976077$)

Наилучшие результаты обучения системы были $\delta=0,00968478$ при обучении 5000 эпох. Корректность распознавания злокачественной ткани при кросс-валидации системы составила 98%.

Заключение

В данной работе предложен интеллектуальный алгоритм распознавания злокачественной ткани на основе глубокого анализа спектров комбинационного рассеяния света. Алгоритм был реализован на языке Python с использованием технологий Deep Learning. Достоверность идентификации злокачественной опухлевой ткани составила от 97,5 до 98%.

Сведения об авторах статьи:

Павлов Валентин Николаевич – д.м.н., профессор, чл.-корр. РАН, зав. кафедрой урологии с курсом ИДПО, ректор ФГБОУ ВО БГМУ Минздрава России. Адрес: 450008, г. Уфа, ул. Ленина, 3. E-mail: pavlov@bashgmu.ru.

Биялялов Азат Ринатович – к.м.н., доцент кафедры травматологии и ортопедии с курсом ИДПО ФГБОУ ВО БГМУ Минздрава России. Адрес: 450008, г. Уфа, ул. Ленина, 3. E-mail: bilyalov@bashgmu.ru.

Гильманова Рита Фларидовна – клинический ординатор кафедры онкологии с курсом онкологии и патологической анатомии ИДПО ФГБОУ ВО БГМУ Минздрава России. Адрес: 450008, г. Уфа, ул. Ленина, 3. E-mail: margaritagilmanova@mail.ru.

Якупов Расуль Радикович – д.м.н., профессор кафедры травматологии и ортопедии с курсом ИДПО ФГБОУ ВО БГМУ Минздрава России. Адрес: 450008, г. Уфа, ул. Ленина, 3. E-mail: rasulr@mail.ru

Ковтуненко Алексей Сергеевич – к.т.н., доцент кафедры информатики ФГБОУ ВО УГАТУ. Адрес: 450008, г. Уфа, ул. К. Маркса, 12. E-mail: askovtunenکو@mail.ru.

Бикмеев Александр Тимерзянович – к. физ.-мат. н., доцент кафедры высокопроизводительных вычислительных технологий и систем ФГБОУ ВО УГАТУ. Адрес: 450008, г. Уфа, ул. К. Маркса, 12. E-mail: bikmeyer@gmail.com.

Карчевский Станислав Геннадиевич – начальник отдела физико-химических методов исследования департамента химии и технологии ГУП «Институт нефтехимпереработки Республики Башкортостан». Адрес: 450065, Уфа, ул. Инициативная, 12. E-mail: st_karchevsky@mail.ru.

Халилов Радик Равилович – с.н.с. ГУП «Институт нефтехимпереработки Республики Башкортостан». Адрес: 450065, Уфа, Инициативная, 12. E-mail: radik_khalilov@mail.ru.

ЛИТЕРАТУРА

1. A review of Raman spectroscopy advances with an emphasis on clinical translation challenges in oncology / M. Jermyn, J. Desroches, K. Aubertin [et al.] // Phys. Med. Biol. – 2016. – Vol. 61, № 23. – P. R370-R400.

2. Cancer imaging using surface-enhanced resonance Raman scattering nanoparticles / S. Harmsen, M.A. Wall, R. Huang, M.F. Kircher // Nat. Protoc. – 2017. – Vol. 12. – P. 1400-14.
3. Discrimination of breast cancer from normal tissue with Raman spectroscopy and chemometrics / Q.-B. Li, W. Wang, Ch.-H. Liu, G.-J. Zhanga // J. Applied Spectroscopy. – 2015. – Vol. 82. – P. 450-455.
4. Potential for Raman spectroscopy to provide cancer screening using a peripheral blood sample / A.T. Harris, A. Lungari, C.J. Needham [et al.] // Head Neck Oncol. – 2009. – Vol. 1, № 1. – Art. 34.
5. Rehman ur, I. Vibrational spectroscopy for tissue analysis / I. ur Rehman, Z. Movasaghi, Sh. Rehman. – CRC Press, 2012. – 356 p.
6. Stone, N. Advanced transmission Raman spectroscopy: a promising tool for breast disease diagnosis / N. Stone, P. Matousek // Cancer Res. – 2008. – Vol. 68. – P. 4424-30.
7. Surface-enhanced Raman spectroscopy investigation on human breast cancer cells / J. Zhu, J. Zhou, J. Guo [et al.] // Chemistry Central Journal. – 2013. – P. 7-37.
8. Vibrational spectroscopy for medical diagnosis / ed. by M. Diem, P.R. Griffiths, J.M. Chalmers. – London: Wiley, 2008. – 358 p.
9. Wachsmann-Hogiu, S. Chemical analysis in vivo and in vitro by Raman spectroscopy: from single cells to humans / S. Wachsmann-Hogiu, T. Weeks, T. Huser // Curr. Opin. Biotechnol. – 2009. – Vol. 20. – P. 63-73.

REFERENCES

1. Jermyn M., Desroches J., Aubertin K. et al. A review of Raman spectroscopy advances with an emphasis on clinical translation challenges in oncology. Phys. Med. Biol., 2016, Vol. 61, № 23, P. 370-400. (in Eng)
2. Harmsen S., Wall M.A., Huang R., Kircher M.F. Cancer imaging using surface-enhanced resonance Raman scattering nanoparticles. Nat. Protoc. 2017. Vol. 12. P. 1400-14. (in Eng)
3. Li Q.-B., Wang W., Liu Ch.-H., Zhanga G.-J. Discrimination of breast cancer from normal tissue with Raman spectroscopy and chemometrics. J. Applied Spectroscopy. 2015, Vol. 82, P. 450-455. (in Eng)
4. Harris A.T., Lungari A., Needham C.J. et al. Potential for Raman spectroscopy to provide cancer screening using a peripheral blood sample. Head Neck Oncol. 2009. Vol. 1, № 1. Art. 34. (in Eng)
5. Rehman ur I., Movasaghi Z., Rehman Sh. Vibrational spectroscopy for tissue analysis. CRC Press, 2012. 356 p. (in Eng)
6. Stone N., Matousek P. Advanced transmission Raman spectroscopy: a promising tool for breast disease diagnosis. Cancer Res. 2008. Vol. 68. P. 4424-30. (in Eng)
7. Zhu J., Zhou J., Guo J. et al. Surface-enhanced Raman spectroscopy investigation on human breast cancer cells. Chemistry Central Journal. 2013. P. 7-37. (in Eng)
8. Diem M., Griffiths P.R., Chalmers J.M. Vibrational spectroscopy for medical diagnosis. London: Wiley, 2008, 358 p. (in Eng)
9. Wachsmann-Hogiu, S., Weeks T., Huser T. Chemical analysis in vivo and in vitro by Raman spectroscopy: from single cells to humans. Curr. Opin. Biotechnol., 2009, Vol. 20, P. 63-73. (in Eng)

УДК 615.22/322.074:581.

© И.Р. Баймухаметов, С.Р. Хасанова, Н.В. Кудашкина, 2018

И.Р. Баймухаметов, С.Р. Хасанова, Н.В. Кудашкина
**РАЗРАБОТКА МЕТОДИКИ КОЛИЧЕСТВЕННОГО ОПРЕДЕЛЕНИЯ
 ФЛАВОНОИДОВ В ГЕПАТОПРОТЕКТОРНОМ СБОРЕ**
*ФГБОУ ВО «Башкирский государственный медицинский университет»
 Минздрава России, г. Уфа*

В связи с неуклонным ростом числа заболеваний печени поиск новых фитопрепаратов, обладающих гепатопротекторным действием, является актуальным. Для внедрения в практику новых лекарственных растительных препаратов необходимо разработать нормативный документ, регламентирующий качество фитопрепарата. В данной статье рассматриваются вопросы разработки методики количественного определения флавоноидов в новом гепатопротекторном сборе. Для разработки методики использован спектрофотометрический метод. Исследованы различные параметры экстракции флавоноидов. Подобраны оптимальные условия извлечения флавоноидов (измельченность сырья, концентрация этилового спирта, время экстракции), количество комплексообразователя, время образования комплекса флавоноидов с хлоридом алюминия. Согласно проведенным исследованиям разработана спектрофотометрическая методика количественного определения суммы флавоноидов в пересчете на гиперозид. Содержание флавоноидов в гепатопротекторном сборе в среднем составило $1,53 \pm 0,03\%$. На основании полученных данных разработан числовой показатель для включения в проект фармакопейной статьи для гепатопротекторного сбора – содержание флавоноидов не менее 1,2%.

Ключевые слова: спектрофотометрия, флавоноиды, гиперозид, гепатопротекторный сбор, количественное определение.

I.R. Baimukhametov, S.R. Khasanova, N.V. Kudashkina
**DEVELOPMENT OF THE METHOD FOR QUANTITATIVE DETERMINATION
 OF FLAVONOIDS IN HEPATOPROTECTIVE COLLECTION**

In connection with a steady increase in the number of liver diseases, the search for new herbal formulations that have a hepatoprotective effect is topical. To introduce new medicinal herbal preparations into practice, it is necessary to develop a regulatory document regulating the quality of herbal formulations. The article deals with the development of a methodology for the quantitative determination of flavonoids in a new hepatoprotective preparation. To develop the technique, a spectrophotometric method was used. Various parameters of extraction of flavonoids have been studied. Optimal conditions for the extraction of flavonoids (raw materials, ethyl alcohol concentration, extraction time), the amount of complexing agent, and the formation of a complex of flavonoids with aluminum chloride were detected. According to the research, a spectrophotometric method for quantifying the amount of flavonoids in terms of hyperoside was developed. The content of flavonoids in the hepatoprotective collection averaged $1.53 \pm 0.03\%$. Based on the data obtained, a numerical indicator was developed for the inclusion in the draft of a pharmacopoeial article for hepatoprotective collection - the content of flavonoids is not less than 1.2%.

Key words: spectrophotometry, flavonoids, hyperoside, hepatoprotective collection, quantitative content.