

КЛАССИФИКАЦИЯ КЛЕТОК ГИППОКАМПА С ПОМОЩЬЮ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

Деева Ольга Кирилловна¹, Чосич Марко², Колесникова Инна Александровна³, Стрельцова Оксана Ивановна⁴

¹Студент;

Национальный исследовательский ядерный университет «МИФИ»;
Россия, 115409, г. Москва, Каширское шоссе, 31;
e-mail: deeva.olga.2020@gmail.com.

²Ассоциированный профессор;

Белградский университет;
Сербия, 11000, г. Белград, Студенческий проезд, 1;
e-mail: markocotic@gmail.com

³Младший научный сотрудник;

Объединенный институт ядерных исследований;
Россия, 141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
Ассистент;
Государственный университет «Дубна»;
Россия, 141980, Московская обл., г. Дубна, ул. Университетская, 19;
e-mail: innakolesnikova@jinr.ru

⁴Начальник сектора;

Объединенный институт ядерных исследований;
Россия, 141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
Доцент;
Государственный университет «Дубна»;
Россия, 141980, Московская обл., г. Дубна, ул. Университетская, 19;
e-mail: strel@jinr.ru.

Автоматическая идентификация и классификация нейронов в микропрепаратах нервной ткани имеет важное значение при изучении воздействия ионизирующего излучения на головной мозг. Оценка состояния клеток ЦНС специалистом вручную является трудоемким и субъективным процессом, в то время как алгоритмы машинного обучения показали потенциал в автоматизации этой задачи. В данной работе были использованы 81 фотоизображение препаратов гиппокампа мышей, на которых выделяли клетки без видимых повреждений, легко-измененные и дистрофические. Для каждой клетки вычислялись следующие параметры: Площадь, Округлость и Структурная сложность ядра. Данные параметры использовались для обучения классификатора RandomForestClassifier с использованием библиотеки scikit learn. Точность классификации составила 68%, при этом наиболее значимым признаком оказалась структурная сложность ядра. Предложенный классификатор может служить основой для автоматической системы анализа нейронов в микропрепаратах головного мозга.

Ключевые слова: нервная ткань, нейрон, дистрофия, морфометрия, классификатор, машинное обучение.

Для цитирования:

Классификация клеток гиппокампа с помощью алгоритмов машинного обучения / О. К. Деева, М. Чосич, М. А. Колесникова, О. И. Стрельцова // Системный анализ в науке и образовании: сетевое научное издание. 2024. № 2. С. 1-8. EDN: BSBAOL. URL : <https://sanse.ru/index.php/sanse/article/view/614>.



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/deed.ru>

CLASSIFICATION OF HIPPOCAMPAL CELLS USING MACHINE LEARNING ALGORITHMS

Deeva Olga K.¹, Ćosić Marko², Kolesnikova Inna A.³, Streltsova Oksana I.⁴

¹Student;

National Research Nuclear University «MEPhI»;
31 Kashirskoe shosse, Moscow, 115409, Russia;
email: deeva.olga.2020@gmail.com

²Associate Professor,

University of Belgrade;
1 Studentski trg, Belgrade, 11000, Serbia;
email: markocosic@gmail.com.

³Junior Researcher,

Joint Institute for Nuclear Research;
6 Joliot-Curie Str., Dubna, Moscow region, 141980, Russia;
Assistant;
Dubna State University;
19 Universitetskaya Str., Dubna, Moscow region, 141980, Russia;
email: innakolesnikova@jinr.ru

⁴Head of the Sector;

Joint Institute for Nuclear Research;
6 Joliot-Curie Str., Dubna, Moscow region, 141980, Russia;
Associate Professor;
Dubna State University;
19 Universitetskaya Str., Dubna, Moscow region, 141980, Russia;
email: strel@jinr.ru.

Automatic identification and classification of neurons in microslides of nervous tissue is important for analyzing the effects of ionizing radiation on the brain. Manually assessing the condition of CNS cells by a specialist is a time-consuming and subjective process, while machine learning algorithms have shown potential in automating this task. In this work, 81 photographic images of mouse hippocampal samples were used, on which cells without visible damage or critical damage and degenerative cells were divided. The following parameters were calculated for each cell: Area, Roundness and Structural complexity of the nucleus. These parameters were used to train the RandomForestClassifier classifier using the scikit-learn library. The classification accuracy was 68%, with the most significant feature being the structural complexity of the nucleus. The proposed classifier can serve as the basis for an automatic system for analyzing neurons in microslides of the brain.

Keywords: nervous tissue, neuron, degenerative cells, morphometry, classifier, machine learning.

For citation:

Deeva O. K., Cosic M., Kolesnikova I. A., Streltsova O. I. Classification of hippocampal cells using machine learning algorithms. *System analysis in science and education*, 2024;(4):1-8 (in Russ). EDN: BSBAOL. Available from: <https://sanse.ru/index.php/sanse/article/view/614>.

Введение

Идентификация и классификация нейронов является одной из важнейших задач в области гистологического анализа изображений микропрепаратов. Изучение воздействия ионизирующего излучения (ИИ) на ткани и органы важно для оценки рисков, связанных с применением радиации в медико-биологической практике, научно-исследовательской и атомно-промышленной отраслях. Исследования, проведенные в России и за рубежом, показывают, что воздействие тяжелых заряженных частиц и высокоэнергетических протонов на структуры головного мозга могут вызывать когнитивные нарушения [1, 2]. Такие повреждения могут приводить, например, к частичной или полной потере операторских

функций космонавтов при полетах за пределы околоземной орбиты [3]. Поэтому точная оценка биологических эффектов, вызываемых ИИ с различными физическими характеристиками, является ключевой задачей радиационной биологии. Морфологический анализ нейронов — это нетривиальная задача, требующая компетенций опытного специалиста. Более того, это трудоемкий процесс. Качественный анализ клеток головного мозга и соотношение нейронов без видимых нарушений по отношению к клеткам с критическими изменениям, позволяют сделать выводы о влиянии облучения на центральную нервную систему (ЦНС). Алгоритмы машинного и глубокого обучения показали хорошие результаты при решении задач идентификации и классификации [4, 5, 6]. В связи с чем предпринята попытка применения машинного обучения в качестве инструмента по автоматизации распознавания клеток головного мозга и их классификации в зависимости от степени нарушений. Предполагается вычленивать критически значимые признаки оценки тяжести изменений в клетке для специалиста гистолога, которые возможно описать математически и положить в основу будущего классификатора.

В области машинного обучения, особенно глубокого обучения, произошёл значительный прогресс в идентификации, классификации и количественной оценке образцов микропрепаратов. Вычислительное моделирование для анализа медицинских изображений оказывает значительное влияние на клинические применения и научные исследования. За последние десятилетия возросла значимость медицинской визуализации, такой как компьютерная томография, магнитно-резонансная томография, позитронно-эмиссионная томография, маммография, ультразвук, рентгенография и другие. В клинических условиях интерпретация медицинских изображений обычно осуществляется специалистами, такими как рентгенологи и врачи. Однако из-за потенциальной усталости экспертов, исследователи и врачи начали использовать компьютерные технологии [7, 8, 9, 10].

Разработка и применение современного математического аппарата для автоматизации извлечения информации из изображений для принятия интеллектуальных решений является одной из ключевых задач информатики. Новые методы анализа изображений необходимы для создания прорывных стандартизированных информационных технологий для автоматического анализа широкого спектра биомедицинских изображений и поддержки этих технологий.

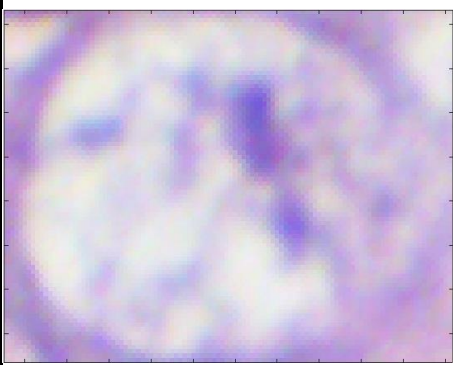
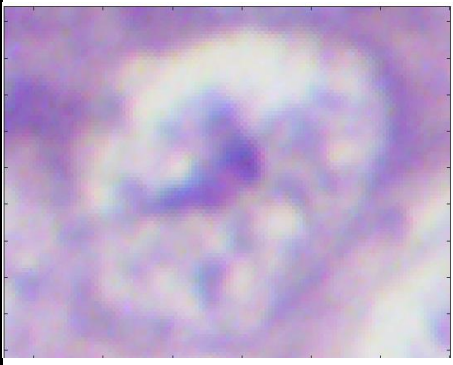

Внедрение и использование достижений машинного и глубокого обучения в нашей области может сократить время, затрачиваемое на анализ данных, и снизить влияние человеческого фактора.

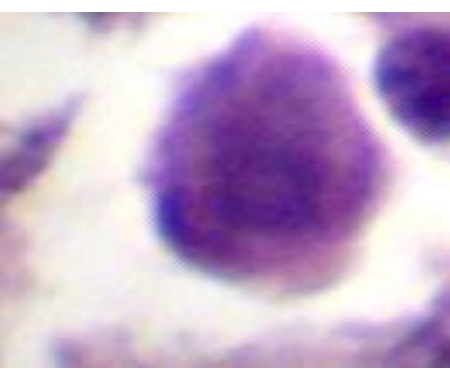
Описание работы

В работе используется набор данных, состоящий из 81 фотоизображения микропрепаратов гиппокампа мышей. Данные предоставлены сектором радиационной физиологии ЛРБ ОИЯИ и были получены в эксперименте, описание которого вместе с полученными результаты представлены в статье [11]. Фотоизображения микропрепаратов были получены с помощью оптического микроскопа (световой микроскоп *BiOptic* B200 и промышленная цифровая камера 5,1МП 1/2,5 с цветным *CMOS*-сенсором *Aptina*).

Биологический материал фиксировали в формалине. Гистологические препараты были приготовлены по стандартной гистологической технике и окрашены гематоксилин-эозином, в результате чего ядра клеток окрашены в фиолетовый цвет, а цитоплазма – в розовый. Нейроны на полученных изображениях были классифицированы вручную на две группы: первая группа включала нейроны без видимых изменений или легко изменённые, вторая – дистрофические клетки. Количество проанализированных нейронов, соответствующих каждому типу, составило по 56 клеток. При анализе нейрона (при ручной разметке) патоморфологу важно оценить форму ядра, внутреннюю структуру ядра, состояние элементов хроматофильной субстанции, структуру ядрышка, наличие перирядового отека, наличие глиоцитов вокруг клетки, сближенность нейронов, соотношение объема цитоплазмы и клеточного ядра и многое другое. Считается, что для нейронов без видимых повреждений (табл. 1) при визуальном-оптическом методе исследования характерно чередование окрашенных и неокрашенных компартментов цитоплазмы, четко очерченные границы ядра клетки, его правильная форма, хорошо видимые ядрышки и ряд других признаков [5]. Именно поэтому параметры, характеризующие форму и внутреннюю структуру ядра, могут быть полезны для классификации дистрофических и нормальных клеток.

Табл. 1. Примеры клеток и соответствующие им параметры

Ядро клетки	Площадь, мкм	Округлость, %	Сложность
 Нейрон без видимых изменений или легко изменённая клетка	6185.52	74.09	0.0794
 Нейрон без видимых изменений или легко изменённая клетка	4480.43	73.32	0.0391
 Дистрофический нейрон	5174.50	40.67	0.0610

	4519.55	55.53	0.1025
Дистрофический нейрон			

С целью формирования группы признаков для автоматизации морфологического анализа, а именно, построения классификатора, в работе предлагается рассмотреть следующие параметры клеток: площадь, индекс округлости [12] и структурная сложность ее ядра [13].

Площадь и округлость были выбраны для характеристики ядра клетки с точки зрения ее формы, а структурная сложность была выбрана, как функция, отражающая внутреннее строение ядра.

Площадь была определена как произведение количества пикселей, относящихся к выделенной области ядра клетки, и масштаба изображения.

Округлость была вычислена с помощью программного обеспечения для обработки изображений *ICY (Institut Pasteur and France-Bioimaging)*. Значение округлости клетки определялось, как отношение между радиусом наименьшей описанной окружности и радиусом максимальной вписанной в ядро клетки окружности. Данное значение выражалось в процентах. В случае идеального круга значение индекса округлости равно 100%:

$$R = \frac{R_{in}}{R_{out}} \cdot 100\%$$

Структурная сложность была вычислена, как сложность объектов, находящихся в минимальном прямоугольном окне, содержащем ядро клетки. Для каждого такого окна, находилась последовательность I_0, I_1, \dots (I_0 - исходное изображение с максимальным разрешением) изображений все меньшего и меньшего разрешения с помощью процедуры понижающей выборки. Структурная сложность в заданном масштабе C_k ($k = 1 \dots N-1$) определяется следующим выражением:

$$C_k = \sum_{j=0}^n \sum_{i=0}^m (I_{k+1}(i, j) - I_k(i, j))^2$$

где индексы i, j указывают вертикальное и горизонтальное положение пикселя изображения с большим разрешением. Общая структурная сложность C определяется как сумма:

$$C = \sum_{k=1}^{N-1} C_k$$

Для многоканальных изображений структурная сложность определялась как среднее арифметическое значений сложности каждого канала (подробное описание расчетов можно найти в работе Багрова и др. [13]).

По вычисленным признакам был сформирован набор данных в виде таблицы с последним столбцом — метками классов: к первому классу относили клетки без видимых изменений или легко-измененные, ко второму - дистрофические.

В качестве алгоритма классификации был выбран алгоритм случайного леса (*RandomForestClassifier* из библиотеки *scikit-learn*).

RandomForestClassifier был выбран как классификатор, так как с его помощью удобно обрабатывать небольшое количество данных. Также его преимуществами являются способность обрабатывать шумные и нелинейные данные и устойчивость к переобучению. Для обучения использовались примерно 75% данных (84 клетки).

Оставшиеся 25% (28 клеток) были использованы для тестирования работы классификатора. При проведении данного эксперимента правильно классифицированы были 19 клеток, из которых 9 были нормальными и 10 являлись дистрофическими. По итогу была получена точность классификации равная 68 процентов, которая была вычислена, как отношение правильно идентифицированных изображений к их общему количеству.

Также были построены парные диаграммы для сравнения распределения пар числовых переменных, содержащие гистограмму для каждой функции в диагональных прямоугольниках. Данные диаграммы представлены на рис. 1. Рыжими точками отмечены дистрофические нейроны, синими — нормальные или легко-измененные нейроны. Как видно из диаграмм, лучше всего дифференцировка нейронов происходит при использовании в паре таких признаков, как индекс округлости и структурная сложность ядра клетки.

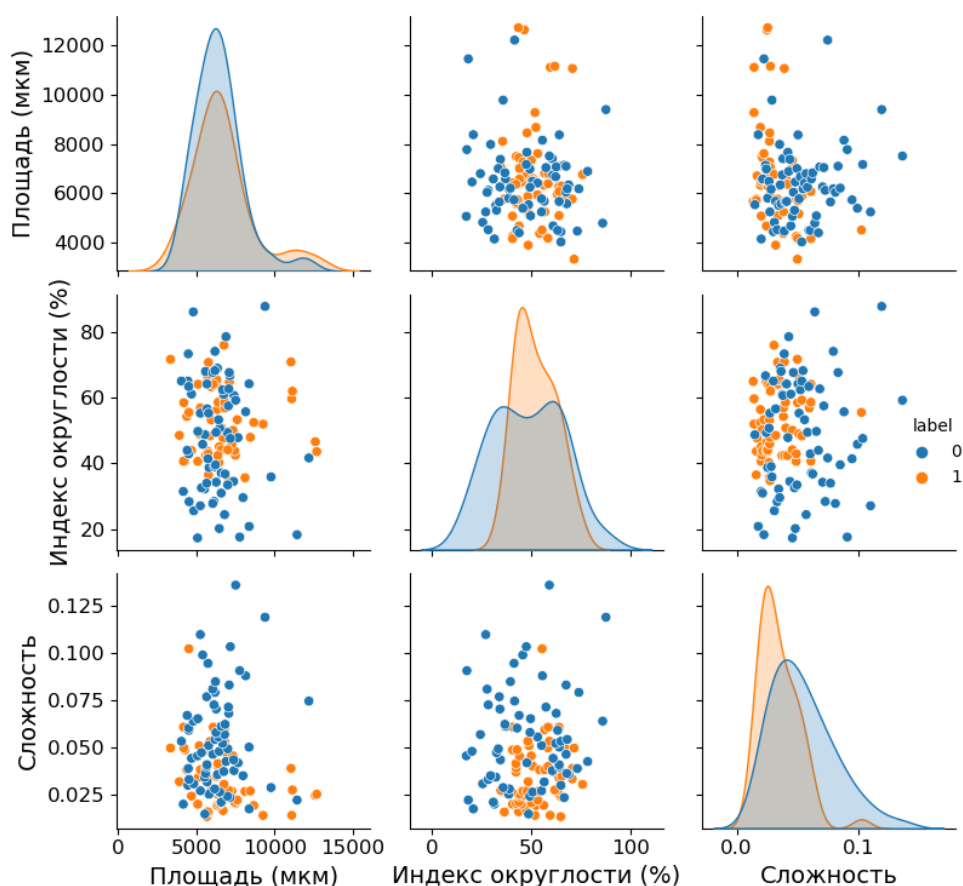


Рис. 1. Парные диаграммы рассеяния нейронов по их признакам

На рис. 2 представлено распределение значимости признаков классификации, рассчитанное благодаря функции классификатора «*forest.feature_importances_*». Как видно из диаграммы самым значимым признаком является структурная сложность клетки.

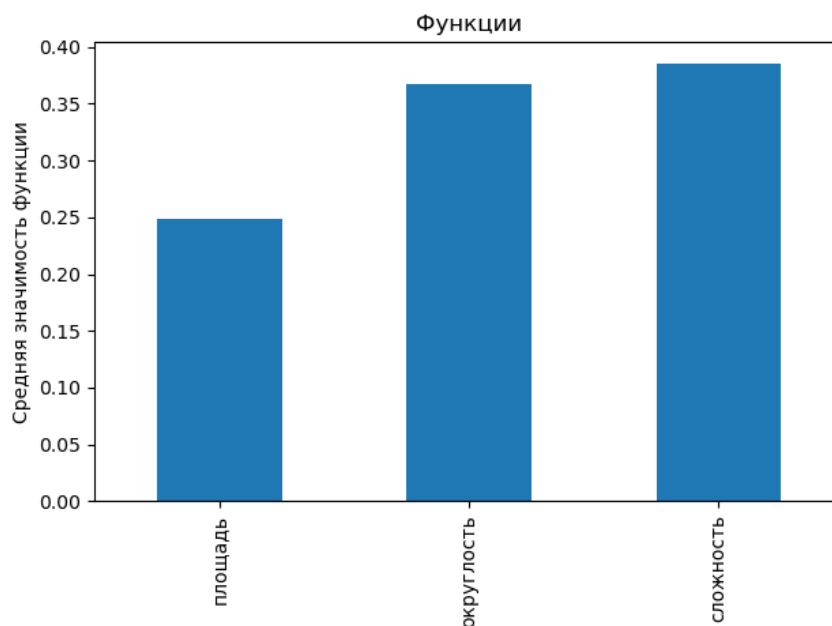


Рис. 2. Диаграмма распределения значимости признаков для классификации

Заключение

Настоящая работа посвящена разработке различных функций для классификации дистрофических нейронов и нормальных в совокупности с легко-измененными, а также изучению признаков клеток, которые могут послужить критериями при решении задачи автоматизации распознавания и классификации.

По итогам проведенной работы была достигнута точность классификатора в 68%, что указывает на то, что предложенный классификатор может быть использован в качестве одного из этапов автоматической системы идентификации клеток гиппокампа на изображениях микропрепаратов, окрашенных рутинным методом. Для повышения точности заданного алгоритма вместо классификатора *Random Forest* может быть использована специально разработанная нейросеть, на вход которой будут поданы описанные выше признаки.

Кроме того, увеличение репрезентативных данных, используемых для обучения и тестирования классификатора, позволит повысить его точность и устойчивость к различным вариациям структур клеток.

Список источников

1. Григорьев А. И., Красавин Е. А., Островский М. А. К вопросу о радиационном барьере при пилотируемых межпланетных полётах // Вестник Российской академии наук. – 2017. – Т. 87, №. 1. – С. 65-69. – DOI: 10.7868/S0869587317010030.
2. Long-term deficits in behavior performances caused by low- and high-linear energy transfer radiation / R. Patel, H. Arakawa, T. Radivoyevitch [et al.] // Radiation Research. – 2017. – Vol. 188(6). – P. 672–680. – DOI: 10.1667/RR14795.1
3. Действие факторов космического полета на центральную нервную систему : Структурно-функциональные аспекты радиомодифицирующего влияния / В. В. Антипов, Б. И. Давыдов, И. Б. Ушаков, В. П. Федоров. – Ленинград : Издательство Наука, 1989. – 328 с.
4. Komura D., Ishikawa S. Machine learning methods for histopathological image analysis // Computational and structural biotechnology journal. – 2018. – Vol. 16. – P. 34-42.

5. Коржевский, Д. Э. Молекулярная нейроморфология. Нейродегенерация и оценка реакции нервных клеток на повреждение / Д. Э. Коржевский, И. П. Григорьев, Е. А. Колос. — Санкт-Петербург : СпецЛит, 2015. — 110 с.
6. Machine learning methods for histopathological image analysis: A review / J. De Matos, S.T. Ataky, A.D. Britto [et al.] // *Electronics*. – 2021. – Vol. 10, No 5. – P. 562.
7. Роль информационных технологий в медицине / Н. С. Самакин, А. Е. Иванова, Е. Б. Матвеева, А. А. Бабаева // Лучшая студенческая статья 2017. – 2017. – С. 246-249.
8. Дороничева А. В., Савин С. З. Методы распознавания медицинских изображений для задач компьютерной автоматизированной диагностики // *Современные проблемы науки и образования*. – 2014. – №. 4. – С. 623-623.
9. Блинов Н. Н. Биомедицинские изображения в современной медицине // *Медицинская техника*. – 2010. – №. 5. – С. 5-9.
10. Лукашевич М. М., Старовойтов В. В. Методика подсчета числа ядер клеток на медицинских гистологических изображениях // *Системный анализ и прикладная информатика*. – 2016. – №. 2. – С. 37-42.
11. The Effects of Whole Body Gamma Irradiation on Mice, Age-Related Behavioral, and Pathophysiological Changes / I. A. Kolesnikova, M Lalkovičova, Yu S Severyukhin [et al.] // *Cellular and Molecular Neurobiology*. – 2023. – Vol. 43, No. 7. – P. 3723-3741. – DOI: 10.1007/s10571-023-01381-1.
12. Оценка различных морфологических характеристик астроцитов в микроколонке коры головного мозга интактных мышей / А. О. Карпова, Н. Н. Проданец, О. С. Баскина [и др.] // *Биосистемы: организация, поведение, управление : Тезисы докладов 73-й Всероссийской с международным участием школы-конференции молодых ученых, Нижний Новгород, 28–30 октября 2020 года*. – Нижний Новгород: Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, 2020. – С. 93.
13. Multiscale structural complexity of natural patterns / A. A. Bagrov, I. A. Iakovlev, M. I. Katsnelson [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2020. – Vol. 117, No. 48. – P. 30241-30251. – DOI 10.1073/pnas.2004976117.