

УДК 004.932.2

DOI 10.18413/2411-3808-2019-46-2-305-310

**ВЫБОР ОПТИМАЛЬНЫХ ПАРАМЕТРОВ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ
АНАЛИЗОВ МОКРОТЫ, ОКРАШЕННОЙ ПО МЕТОДУ ЦИЛЯ – НИЛЬСЕНА
НЕЙРО-НЕЧЕТКОЙ СИСТЕМОЙ ANFIS****SELECTION OF OPTIMAL PARAMETERS OF THE NEURO-FUZZY ANFIS SYSTEM
FOR CLASSIFICATION OF ZIEHL-NIELSEN STAINED SPUTUM SMEAR IMAGES****И.Г. Шеломенцева^{1,2}, С.В. Ченцов², А.Н. Наркевич¹
I.G. Shelomentseva^{1,2}, S.V. Chentsov², A.N. Narkevich¹**

¹Красноярский государственный медицинский университет им. проф. В.Ф. Войно-Ясенецкого,
Россия, 660022, г. Красноярск, ул. Партизана Железняка, д. 1

²Сибирский федеральный университет,
Россия, 660041, г. Красноярск, Свободный пр., д. 79

¹ Krasnoyarsk State Medical University n.a. prof. V.F. Voyno-Yasenetsky,
1 Partizan Zheleznyaka St, Krasnoyarsk, 660022, Russia

² Siberian Federal University,
79 Svobodny Av., Krasnoyarsk, 660041, Russia

E-mail: inga.shell@yandex.ru

Аннотация

Туберкулез по-прежнему является значимым заболеванием мирового уровня. Одним из широко используемых методов диагностики туберкулеза является метод микроскопии Циля – Нильсена. В статье приведены результаты вычислительного эксперимента, направленного на оценивание оптимальных параметров нейро-нечеткой системы классификации изображений анализов мокроты, окрашенных по методу Циля – Нильсена. Авторами описаны используемые методы предобработки и сегментации исследуемых изображений. Экспериментальная выборка построена на основе цветовых характеристик и характеристик формы регионов интереса. В качестве варьируемых значений исследуемых функций рассматривались количество и параметры входной и выходной функции принадлежности для классической нейро-нечеткой системы на основе модели Сугено, размер радиуса действия одного кластера для системы субтрактивной кластеризации. В качестве критериев сравнения использовались значения среднеквадратичной ошибки, регрессии и значение точности классификации. Полученные результаты будут использованы для построения оптимального классификатора исследуемых цифровых изображений мокроты, окрашенной по методу Циля – Нильсена.

Abstract

Tuberculosis (TB) is an important public health issue in this world. The Ziehl – Nielsen method of microscopy is the one of the widely used methods for the diagnosis of tuberculosis. In this paper, we present the results of experiment for calculation of optimal parameters of the neuro-fuzzy classification system for ZN (Ziehl – Nielsen) stained images of sputum smear samples obtained using a light microscope. The authors describe the methods for preprocessing and segmentation of the images. We use the color and shape characteristics of the regions of interest for experimental sample. We use the quantity and parameters of the input and output membership functions for the classical neuro-fuzzy system based on the Sugeno model and size of the range of a single cluster for the subtractive clustering. Authors use the values of mean-square error, regression and accuracy as comparison criteria. The results will be used to build the optimal classifier of ZN (Ziehl – Nielsen) stained images of sputum smear samples obtained using a light microscope.

Ключевые слова: распознавание образов, микобактерия туберкулеза, микроскопия, метод Циля – Нильсена, ANFIS, genfis 1, genfis 2, среднеквадратичная ошибка, регрессия, точность.

Keywords: image processing, tuberculosis bacteria, microscopic, method Ziehl-Nielsen, ANFIS, genfis 1, genfis 2, root-mean-square error, regression, accuracy.

Введение

На текущий момент туберкулез является социально значимым заболеванием. Согласно докладу Всемирной организации здравоохранения в 2016 году, по оценкам ВОЗ умерло от туберкулеза 1,4 миллиона человек [Глобальный ..., 2017]. Большую часть смертей можно было предотвратить, если бы диагностика туберкулеза была выполнена на ранней стадии, в том числе и средствами вычислительной техники.

Искусственная нейронная сеть является одним из наиболее важных инструментов, который используется сейчас в диагностике и оценке состояния здоровья. Существует большое разнообразие в моделях, методах и программах, поддерживающих работу с искусственными нейронными сетями [Волчек и др., 2017; Тетерин и др., 2018]. Нейронно-нечеткие сети ориентированы на поиск эффективного и оптимального решения для задачи прикладного исследования. Поэтому вопросы, связанные с выбором параметров нейронно-нечеткой системы, являются важным этапом ее адаптации под решение конкретной задачи [Лубенцова, Пиотровский, 2017].

Текущее исследование направлено на определение параметров использования системы нейро-нечеткой классификации ANFIS для задачи распознавания микроскопических изображений мокроты, окрашенной по методу Циля – Нильсена. Выбор этой системы обусловлен ее доступностью и применимостью для задач классификации медицинских изображений, которые обычно характеризуются чувствительностью к морфологическим изменениям, шуму, режиму получения снимков и физическим параметрам приборов [Lorenz et al., 1997, Das, Bhattacharya, 2008, Das, Bhattacharya, 2011, Huang et al., 2007].

Материалы и методы

Материалы исследования представляют собой набор микроскопических изображений мокроты пациентов противотуберкулезного диспансера, окрашенных по методу Циля – Нильсена и полученных при помощи тринокулярного микроскопа Микромед 1 вар. 3-20 при увеличении 10×60 с установленной цифровой камерой TourCam UCMOS01300KPA с разрешением 0,3 МР. Обработка мокроты методом Циля – Нильсена подразумевает обработку карболовым фуксином с дальнейшим обесцвечиванием 5% раствором серной кислоты или 3% солянокислым спиртом и докрасиванием 0,25% раствором метиленового синего. Тем самым красный краситель удаляется из присутствующей микрофлоры, кроме кислотоустойчивых микобактерий, к которым относится и микобактерия туберкулеза (рисунок 1) [Приказ ..., 2003].

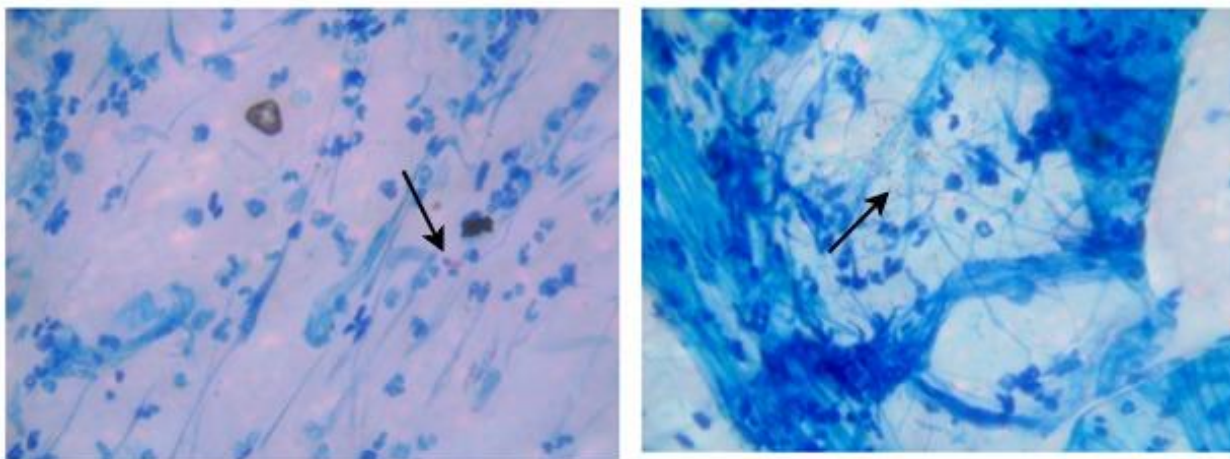


Рис. 1. Пример изображений анализов мокроты, окрашенной по методу Циля – Нильсена
Fig. 1. ZN (Ziehl – Nielsen) stained images of sputum smear samples

Для организации задачи автоматического распознавания к исследуемому набору изображений были применены операции фильтрации и сегментации, процедура выделения признаков и формирования классификаторов.

Проведенное исследование [Наркевич и др., 2018] позволило сформировать набор признаков с точки зрения их информативности – исследовался вектор признаков, состоящий из трех характеристик цвета (значение красного и зеленого цветов цветовой модели RGB и значение оттенка цветовой модели HSV) и одной характеристики формы (компактность). Под компактностью подразумевается отношение сторон ограничивающего исследуемый объект прямоугольника.

Для задачи классификации был использован аппарат нейро-нечеткой логики на примере системы ANFIS. ANFIS является редактором для работы с нейро-нечеткими сетями и является надстройкой над пакетом Matlab. Преимуществом использования данного пакета является автоматическая генерация нейро-нечетких сетей из набора экспериментальных данных, также интуитивная понятность и высокая скорость работы системы [Omisore et al., 2017].

Система ANFIS поддерживает 8 функций принадлежности – группа треугольных функций, которая используется для задания неопределенностей типа «среднее значение», «расположен в интервале», «подобен объекту», «похож на предмет» и группа П-образных функций, которая используется для задания неопределенностей типа «приблизительно в пределах от и до», «примерно равно», «около» [Edoras, 2019]. К треугольным функциям относятся трапецеидальная (*trapmf*) и треугольная (*trimf*), а к П-образным функциям – колоколообразная (*gbellmf*), гауссиана и ее комбинация (*gaussmf* и *gauss2mf*), П-образная функция (*psigmf*), функции разности и произведения между значениями сигмоидальных функций (*dsigmf* и *psigmf*) [Штовба, 2007].

Кроме этого, система ANFIS содержит два выходных значения функции принадлежности – в виде константы или в виде функции. Такое ограничение связано с тем, что система ANFIS основана на модели типа Sugeno, в которой заключения правил задаются линейной функцией от входного вектора [Terano et al., 1989].

ANFIS предлагает исследователям модели нечеткого логического вывода – *genfis 1* и *genfis 2* [Babazadeh et al., 2012]. Модель *genfis 1* создает нечеткую систему с фиксированным количеством функций принадлежности, что может вызвать проблему размерности базы правил в случае большого количества входных переменных. Именно поэтому исследуемый вектор признаков не содержит большого количества переменных. Варьируемые аргументы данной функции включают в себя вектор количества термов входных значений, векторы типов функций принадлежности входных и выходных переменных [Куделина, 2017].

Модель *genfis 2* использует субтрактивную кластеризацию и имеет только один варьируемый параметр – коэффициент подавления, который используется для определения радиуса действия одного кластера [Жилияков, Лихошерстный, 2011]. Работа функции *genfis 2* происходит в два этапа – вначале используется встроенная функция *subclust* для настройки параметров нечеткой системы (мощности терм-множеств и количества правил), затем при помощи метода наименьших квадратов определяется постфиксная часть каждого правила, в результате чего формируется окончательная база правил нечеткой системы.

Вычислительный эксперимент

Вычислительный эксперимент строился на базе классификатора, состоящего из признаков формы и цвета ROI (регионов интереса). Предварительно ROI были выделены из базовых изображений с использованием операций сегментации и фильтрации.

Необходимость фильтрации обусловлена тем, что разные образцы мокроты окрашиваются красителями разной интенсивности, и их изображения получаются неоднородными по цветовой гамме и нередко зашумлены. Распознавание изображений, не прошедших этап предобработки, чревато большим значением погрешности результатов. Ранее проведенные исследования показали, что применение линейной фильтрации при помощи свертки ядром размерности 3×3 к изображениям анализа мокроты при диагностике тубер-



кулеза по методу Циля – Нильсена повышает эффективность использования выбранных методов распознавания образов [Шеломенцева, 2017].

Процедура сегментации используется для выделения на подготовленном изображении областей интереса для последующей операции классификации. Исследования [Наркевич и др., 2017] показали, что среди рассмотренных авторами алгоритмов сегментации, наиболее подходящей для задач исследования, явилась сегментация при помощи искусственной нейронной сети и вейвлет преобразования Мексиканская шляпа. Сравнение было выполнено с такими методами, как детекторы границ и углов (алгоритмы FAST, Канны, Собеля, Робертса, Робинсона, Харриса), пороговая бинарная сегментация, метод Отцу, контурный анализ.

Начальные параметры вычислительного эксперимента приведены в таблице 1. Выборочные результаты вычислительного эксперимента представлены в таблицах 2 и 3.

Таблица 1
Table 1

Значения параметров для работы ANFIS
Parameter values for ANFIS

Параметр	Значение	Параметр	Значение
Количество эпох	200	Величина уменьшения шага	0.9
Величина ошибки	0.0	Величина увеличения размера шага	1.1
Размер шага	0.01		

Таблица 2
Table 2

Значения параметров для работы функции genfis 1
Parameter values for genfis 1

Варьируемые параметры (количество, тип входной функции, тип выходной функции)	MSE	R	Точность, %
3, gaussmf, constant	0,095	0,7860	87,3
3, gaussmf, linear	0,182	0,6345	87,6
3, trimf, constant	0,096	0,7834	87,3
3, trimf, linear	0,098	0,7779	87,5
3, trapmf, constant	0,104	0,7637	86,3
3, trapmf, linear	0,113	0,7816	87,3
3, gbellmf, constant	0,097	0,7803	87,2
2, gbellmf, constant	0,096	0,78	87,3
3, dsigmf, constant	0,096	0,7829	87,3
2, dsigmf, constant	0,100	0,7745	86,5
3, pimf, constant	0,106	0,7571	86,1

При работе системы ANFIS размер шага сначала увеличивается до максимального значения, затем начинает уменьшаться с целью достижения оптимального значения. В качестве минимального и максимального параметра были выбраны значения по умолчанию 0,9 и 1,1. Количество эпох (итераций обучения) было взято равным 200.

В процессе вычислительного эксперимента происходило поэтапное изменение одного из параметров исследуемой модели с целью оценки его влияния на результат обучения. В качестве критериев для сравнения нейро-нечетких моделей были выбраны значение среднеквадратичной ошибки (MSE), значение регрессии (R) и точность классификации (accuracy). В нейро-нечетких системах критерии среднеквадратичной ошибки и регрессии характеризуют качество построенной модели [Gliwa, Byrski, 2011]. Точность accuracy рассчитывается как доля правильных ответов алгоритма классификации, построенного на базе нейро-нечеткой системы, а именно как отношение суммы долей правильно распознанных микобактерий туберкулеза и правильно распознанных регионов интереса, не являющихся микобактериями, к общему числу выделенных объектов.

Таблица 3
Table 3

Значения параметров для работы функции *genfis 2*
Parameter values for *genfis 2*

Радиус	MSE	R	Точность
0,2	0,093	0,79	87,3
0,3	0,094	0,78	87,1
0,4	0,095	0,78	87,2
0,5	0,097	0,78	86,8
0,6	0,097	0,78	86,8
0,7	0,097	0,78	86,85
0,8	0,097	0,78	86,9
0,9	0,097	0,78	86,9

Выводы

В случае использования функции *genfis 1* наилучшие показатели точности система ANFIS продемонстрировала при гауссовской функции принадлежности, количеством переменных, равных 3, и выходом в виде линейной функции. Наилучшие показатели регрессии и значения MSE система показала на П-образных функциях принадлежности, количеством переменных, равных 3, и выходом в виде константы.

При использовании функции *genfis2* наилучшие показатели критериев сравнения система показала на значении радиуса области правил в диапазоне от 0,2 до 0,4.

В контексте проводимых исследований система *genfis2* использовалась для выявления кластерной структуры параметров исходного материала путем разбиения входных данных на группы, что поможет создать исходную систему для обучения посредством *genfis1*. Наличие оптимальных параметров функций *genfis 1* и *genfis 2* системы ANFIS позволит построить модель нечеткого вывода, которая ляжет в основу высокопроизводительного классификатора изображений анализов мокроты, окрашенной по методу Циля – Нильсена.

Список литературы References

1. Глобальный доклад о туберкулезе ВОЗ за 2017 год. URL: http://www.who.int/tb/publications/global_report/ru/ (дата обращения: 27 февраля 2019).
Global'nyy doklad o tuberkuleze VOZ za 2017 god [Global tuberculosis report 2017]. Available at: http://www.who.int/tb/publications/global_report/en/ (accessed 27 February 2019)
2. Волчек Ю.А., Шишко О.Н., Спиридонова О.С., Мохорт Т.В. 2017. Положение модели искусственной нейронной сети в медицинских экспертных системах. *Juvenis scientia*. 9: 4–9.
Volchek Ju.A., Shishko O.N., Spiridonova O.S., Mohort T.V. Polozheniye modeli iskusstvennoy neyronnoy seti v meditsinskikh ekspertnykh sistemakh [Position of the artificial neural network model in medical expert systems]. *Juvenis scientia*. 9: 4–9.
3. Жилияков Е.Г., Лихошерстный А.Ю. 2011. Метод нейросетевого распознавания объектов на аэрокосмических изображениях земной поверхности. *Научные ведомости БелГУ. История. Политология. Экономика. Информатика*. 13(108): 115–120.
Zhylyakov E.G. Likhosherstnyy A.U. 2011. Neural network method for recognition of objects in the picture is of the earth's surface. *Belgorod State University Scientific Bulletin. History. Political science. Economics. Information technologies*. 13(108): 115–120.
4. Куделина Д.В. 2017. Нечеткая система контроля состояния изоляции. *Научные ведомости БелГУ. Экономика. Информатика*. 2(251): 102–107.
Kudelina D.V. 2017. Fuzzy control system of the insulation state. *Belgorod State University Scientific Bulletin. Economics. Information technologies*. 2(251): 102–107.
5. Лубенцова Е.В., Пиотровский Д.Л. 2017. Исследование алгоритмов обучения нейро-нечеткой системы управления биотехнологическим процессом. *Научный журнал КубГАУ*. 128. 1–11.
Lubencova E. V., Piotrovskij D. L. The study of learning algorithms of neuro-fuzzy systems control of biotechnological process. *Scientific Journal of KubSAU*. 128. 1–11.



6. Минздрав России. О совершенствовании противотуберкулезных мероприятий в Российской Федерации. Приказ. М., 2003, 109.
- Minzdrav Rossii [Ministry of Health of the Russian Federation]. O sovershenstvovanii protivotuberkuleznykh meropriyatiy v Rossiyskoy Federatsii [On improving tuberculosis measures in the Russian Federation]. Order. M., 2003, 109 (in Russian).
7. Наркевич А.Н., Плотников Д.В., Виноградов К.А., Катаева А.В. 2018. Сравнение методов отбора признаков для идентификации объектов на цифровых изображениях микроскопических препаратов. Инженерный вестник Дона. 2: 1–10.
- Narkevich A.N., Plotnikov D.V., Vinogradov K.A., Kataeva A.V. 2018. Comparison of methods for the selection of signs for the identification of objects in digital images of microscopic preparations. Engineering journal of Don. 2: 1–10.
8. Наркевич А.Н., Шеломенцева И.Г., Виноградов К.А., Сысоев С.А. 2017 Сравнение методов сегментации цифровых микроскопических изображений мокроты, окрашенных по методу Циля – Нильсена. Инженерный вестник Дона. 4: 1–11.
- Narkevich A.N., Shelomentseva I.G., Vinogradov K.A., Sysoev S.A. Comparison of segmentation methods for digital microscopic images of sputum stained by the Ziehl – Nielsen method. Engineering journal of Don. 4: 1–11.
9. Тетерин Д.А., Хабибулин Р.Ш., Гудин С.В. 2018. Обзор применения искусственных нейронных сетей в управлении социальными и экономическими системами. Научные ведомости БелГУ. Сер. Экономика. Информатика. 45(3): 574–583.
- Teterin D.A., Habibulin R.Sh., Gudim S.V. 2018. Overview of the use of artificial neural networks in the management of social and economic systems. Belgorod State University Scientific Bulletin. Economics. Information technologies. 45(3): 574–583.
10. Тэрано Т., Асаи К., Сугэно М. 1993. Прикладные нечеткие системы. М., Мир, 368. (Terano T., Asai K., Sugeno M. 1989. Applied Fuzzy Systems. Academic Press, 314).
- Terano T., Asai K., Sugeno M. 1993. Applied Fuzzy Systems. M., Mir, 368. (Terano T., Asai K., Sugeno M. 1989. Applied Fuzzy Systems. Academic Press, 314).
11. Шеломенцева И.Г. 2017. Результаты фильтрации и сегментации изображений анализа мокроты, окрашенной по методу Циля – Нильсена. International journal of advanced studies. 7(4–2): 110–114.
- Shelomentseva I.G. 2017. Results of filtration and segmentation of the images of analysis of sputum stained by Ziehl-Nielsen method. International journal of advanced studies. 7(4–2): 110–114.
12. Штовба С.Д. 2007. Проектирование нечетких систем средствами MATLAB. М., Горячая линия – Телеком, 288.
- Shtovba S.D. 2007. Proektyrovanye nechetkyh system sredstvamy MATLAB [Design of fuzzy systems using MATLAB]. M., Gorjachaja lynuja – Telekom, 288.
13. Babazadeh K.N., Arabalibeik H., Salehian P., Setayeshi S. 2012. Abnormal red blood cells detection using adaptive neuro-fuzzy system. Studies in health technology and informatics. 173: 30–34.
14. Das A., Bhattacharya M. 2008. GA Based Neuro Fuzzy Techniques for Breast Cancer Identification. International Machine Vision and Image Processing Conference: 136–141.
15. Das A., Bhattacharya M. 2011. Computerized decision support system for mass identification in breast using digital mammogram: A study on GA based neuro-fuzzy approaches. Advanced in experimental medicine and biology. 696: 523–533.
16. Edoras. CentOS server at San Diego State University. URL: <https://edoras.sdsu.edu/doc/matlab/> (accessed 27 February 2019).
17. Gliwa B., Byrski A. 2011. Hybrid neuro-fuzzy classifier based on NEFCLASS model. Computer Science. 12: 115–135.
18. Huang, M.L., Chen H.Y., Huang J.J. 2007. Glaucoma detection using adaptive neuro-fuzzy inference system. Expert Systems with Applications. Vol. 32: 458–468.
19. Lorenz A., Blum M., Ermert H., Senge T. 1997. Comparison of Different Neuro-Fuzzy Classification Systems for the Detection of Prostate Cancer in Ultrasonic Images. IEEE Ultrasonics Symposium Proceedings. An International Symposium. Vol.2: 1201–1204.
20. Omisore M.O., Samuel O.W., Atajeromavwo E.J. 2017. A genetic-neuro-fuzzy inferential model for diagnosis of tuberculosis. Applied Computing and Informatics. 13(1): 27–37.