

УДК 681.327.12.001.362

М. М. ЛУКАШЕВИЧ,¹ В. В. СТАРОВОЙТОВ.²

МЕТОДИКА ПОДСЧЕТА ЧИСЛА ЯДЕР КЛЕТОК НА МЕДИЦИНСКИХ ГИСТОЛОГИЧЕСКИХ ИЗОБРАЖЕНИЯХ

¹Белорусский государственный университет информатики и радиоэлектроники²Объединенный институт проблем информатики Национальной академии наук
Беларуси

В статье исследуется методика автоматического подсчета числа ядер клеток на гистологических изображениях. Эта операция широко применяется при диагностике различных заболеваний и морфологическом анализе клеток. В связи с этим, процедура автоматического подсчета числа ядер клеток является ключевым этапом в системах микроскопического анализа медицинских изображений гистологических препаратов. Основной целью работы была разработка эффективной схемы автоматического подсчета ядер клеток на основе современных методов обработки изображений: направленной фильтрации, адаптивной бинаризации изображений и математической морфологии. В отличие от известных исследований, представленный подход не предусматривает сегментацию ядер клеток на изображении, а лишь предполагает их обнаружение и подсчет их количества. Это позволяет избежать сложных алгоритмических вычислений и обеспечивает хорошую точность подсчета ядер клеток.

В работе описан ряд экспериментов, выполненных для оценки эффективности предложенной методики с использованием доступной в интернете тестовой базы медицинских гистологических изображений. Определены критичные параметры алгоритмов, настраиваемые на каждом этапе анализа изображений. Для каждого параметра определен интервал тестируемых значений, а затем реализована процедура выбора не только оптимальных значений каждого параметра, но их из взаимная комбинация, на основе общепринятых количественных оценок точности (Precision) и полноты (Recall). Полученные результаты сравнивались с последними достижениями в данной области и показали приемлемый уровень точности предложенной методики. Прототип программного обеспечения, разработанного в рамках проведенного исследования, можно рассматривать как автоматический инструмент для анализа ядер клеток. Разработанный подход может быть адаптирован к различным задачам анализа ядер клеток различных органов.

Ключевые слова: изображения гистологических препаратов, обработка и анализ цифровых изображений, бинаризация

Введение

Разработка автоматизированных систем обработки медицинских изображений является наиболее важным и быстро развивающимся направлением в области микроскопических исследований гистологических препаратов. Компьютерный анализ микроскопических изображений клеток широко применяется в ряде прикладных областей, включая диагностику заболеваний и морфологический анализ клеток, что особенно важно в медико-биологических исследованиях и биоинформатике. Подход, основанный на микроскопических исследованиях, играет важную роль при решении следующих задач: обнаружение злокачественных и раковых клеток, определение естествен-

ных морфологических изменений клеток, исследование характеристик клеток в динамике (например, во время процедуры терапии). В связи с этим, процедура автоматического подсчета числа ядер клеток является ключевым блоком в системах микроскопического анализа медицинских изображений гистологических препаратов. С 1970-х годов ведутся активные работы по разработке методов и алгоритмов для решения данной проблемы [1, 2], а обнаружение клеточных структур и подсчет числа ядер клеток являются общими задачами для многих исследований.

Одним из наиболее простых, быстрых и очевидных методов подсчета ядер клеток является пороговая сегментация изображений, что по-

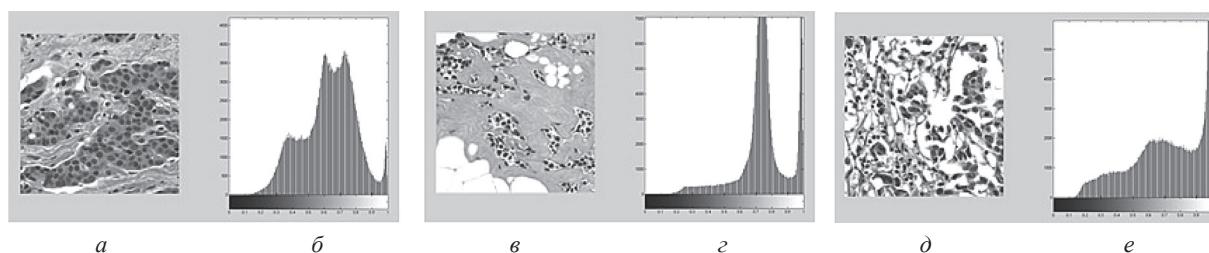


Рис. 1. Примеры гистологических изображений *a, в, д* и их гистограммы яркости соответственно *б, г, е*

зволяет разделить объекты интереса и фон. Другой подход заключается в использовании различных детекторов краев (LoG фильтр, фильтры Лапласа и т. д.). Тем не менее, обнаружение и сегментации ядер клеток является сложной задачей в силу того, что клетки имеют неоднородную структуру и высокую вариативность. Еще одной особенностью гистологического изображения является неоднородная подсветка регистрируемого препарата, что также накладывает ограничения на использование традиционных подходов.

В последние годы было предложено большое число методов и алгоритмов сегментации, разделения и классификации ядер клеток при проведении процедуры гистологического анализа. Более сложные решения состоят из нескольких этапов анализа изображений и включают в себя использование нескольких независимых алгоритмов. Эти решения основаны на традиционных алгоритмах анализа и обработки изображений, таких как метод активных контуров, метод морфологического водораздела, морфологические операции, алгоритм k -средних, метод опорных векторов и т. д. [3–12]. В последних работах для повышения эффективности исследователи предлагают комбинировать различные подходы. Обзор методов и алгоритмов, используемых для решения задач обнаружения, сегментации и классификации клеток медицинских изображениях можно найти в работах [1, 2]. В данной работе решалась задача автоматического подсчета ядер клеток на гистологических изображениях независимо от их размера.

В зависимости от качества подготовленного для исследования препарата и оптического оборудования, микроскопические изображения могут быть как полутоновыми, так и цветными. Большое число различных объектов на гистологических изображениях является серьезным недостатком, влияющим на процедуру анализа изображений данного класса. Ана-

лиз клеточных структур является трудной задачей из-за сложности природных объектов (зашумленность, неоднородный фон, вариативность характеристик объекта). Большинство ядер клеток являются частью гистологических структур со сложными и нерегулярными визуальными признаками [13]. Примеры медицинских гистологических изображений и их гистограммы яркости представлены на рис. 1. Форма гистограмм доказывает тот факт, что задачу автоматического обнаружения и сегментации ядер клеток сложно решить с помощью только лишь пороговых методов.

2. Методы и данные

2.1. Данные. Исследуемые гистологические препараты окрашивались гематоксилином и эозином (один из самых распространенных методов окрашивания в гистологии), после чего были получены цифровые изображения (H&E изображения), вошедшие в общедоступную базу медицинских гистологических изображений [14, 15]. Гистологические срезы были оцифрованы с помощью сканера Zeiss Mirax Scan, оснащенного Zeiss Plan-Apochromat 20x (числовая апертура = 0,8) и AVT Marlin F-146C Firewire 1/2" CCD камерой, размер пикселя 4,65×4,65 мкм. Комбинация 20x увеличения и 1x C-Mount адаптера позволили получить разрешение изображения 0,23×0,23 мкм. Все слайды были отсканированы в 20x увеличении и при полном разрешении. Полученные изображения были преобразованы в файлы изображений JPEG с 85% качеством JPEG. Размер изображения – 600×600 пикселей. Тестовая база состоит из 36 цветных изображений, которые содержат 7931 ядер клеток. В базе изображений присутствуют снимки различных органов (молочной железы, почек, слизистой оболочки желудка, соединительной ткани, тонкой кишки и т. д.). Координаты центров ядер клеток были отмечены тремя экспертами-патологоанатомами и представлены в виде 36 xml-фай-

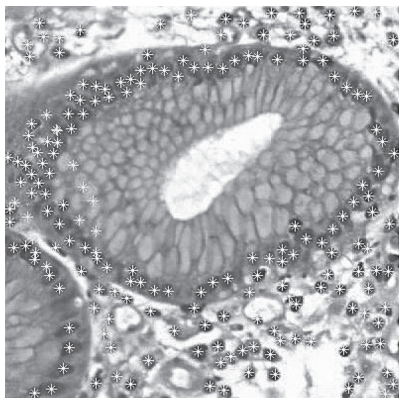


Рис. 2. Пример H&E изображения с отмеченными центрами ядер клеток

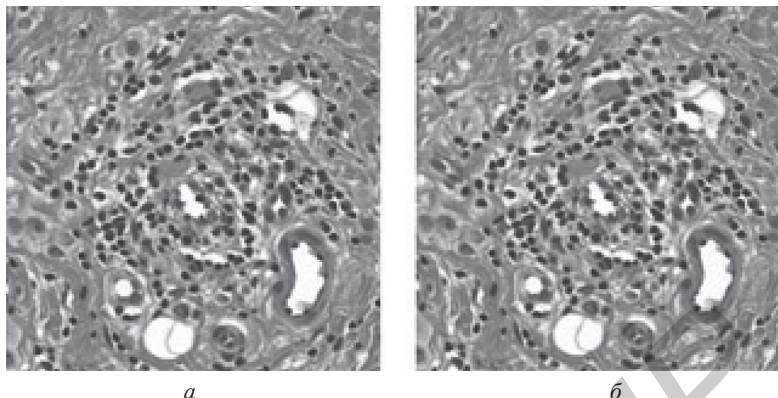


Рис. 3. *a* – исходное изображение, *б* – изображение после уменьшения числа цветов в цветовом пространстве RGB

лов (для каждого гистологического изображения соответственно), рис. 2.

2.2. Основная идея. Предложенная методика подсчета числа ядер клеток на гистологических изображениях состоит из следующих этапов: 1. уменьшение числа цветов в цветовом пространстве RGB; 2. изменение контраста изображения; 3. применение направленной фильтрации [17]; 4. пороговая сегментация; 5. морфологической обработки.

Этап предварительной обработки включает в себя несколько этапов. Исходные изображения из тестовой базы содержат ограниченное число основных цветов (розовый, фиолетовый и оттенки синего). Предлагается уменьшить количество цветов в исходной палитре изображения с использованием алгоритма квантования цвета методом минимизации дисперсии [18]. Результаты реализации данного алгоритма практически не заметны, но играют решающее значение на последующих этапах анализа, рис. 3.

На следующем этапе происходит увеличение контрастности изображений путем растяжения значений интенсивностей динамическо-

го диапазона. Далее предлагается использовать направленный фильтр, который обрабатывает изображение, руководствуясь дополнительной информацией о каждом пикселе. При удачном выборе дополнительной информации направленный фильтр превосходит по качеству билатеральную фильтрацию. Бинаризация реализуется с помощью алгоритма Sauvola [19], который анализирует обрабатываемую картинку и адаптивно вычисляет индивидуальный порог бинаризации для каждого пикселя. Найденный порог используется для бинаризации значения текущего пикселя [19–21], рис. 4.

Затем применяются операции математической морфологии – открытие и закрытие с использованием структурного элемента небольшого размера. Операция закрытия «закрывает» небольшие внутренние «дырки» в изображении, и убирает углубления по краям области. Открытие позволяет избавиться от небольших кусочков изображения, выходящих за границу области. После этого определяется число областей на изображении, соответствующих ядрам клеток.

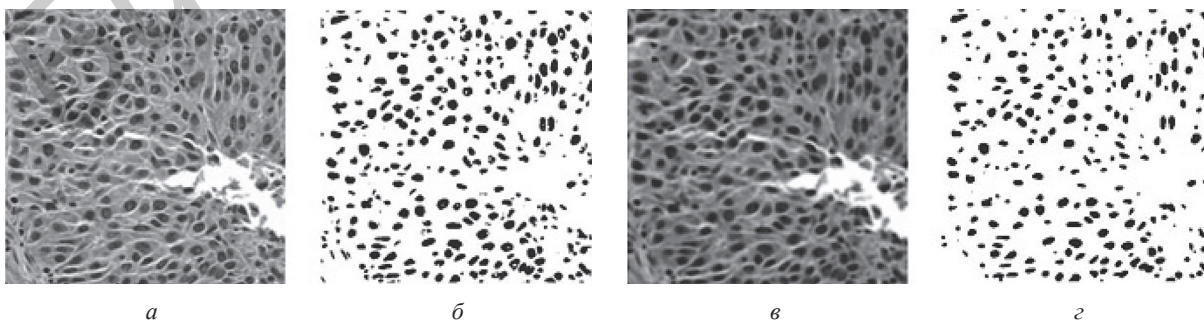


Рис. 4. Улучшение результатов бинаризации с помощью направленного фильтра: *a* – полутоновое изображение без применения направленного фильтра, *б* – результат пороговой бинаризации; *в* – полутоновое изображение с применением направленного фильтра; *г* – результат пороговой бинаризации

Таблица 1. Выбор оптимальным значений параметров алгоритмов, реализуемых на отдельных этапах предложенной методики

Этап обработки	Параметр	Тестируемые диапазоны значений параметров	Значение параметров, дающее максимальные значения точности и полноты
квантование цвета методом минимизации дисперсии	количество цветов в цветовом пространстве	3, 4, 5, 7, 9	4
направленная фильтрация	размер маски степень размытия	3×3, 4×4, 5×5, 7×7 0,3, 0,4, 0,5, 0,55, 0,6, 0,7	3×3 0,55
Бинаризация методом Sauvola	размер анализируемой области при бинаризации порог	5×5, 7×7, 15×15, 30×30, 45×45, 55×55, 65×65, 75×75, 85×85, 95×95, 105×105, 115×115, 125×125, 135×135, 145×145 0,2, 0,25, 0,3, 0,35, 0,4, 0,45, 0,5, 0,7, 0,9	75×75 0,3
Морфологическая фильтрация (замыкание и отмыкание)	радиус структурного элемента форма структурного элемента	1, 2, 3, 4, 5 диск	3 диск

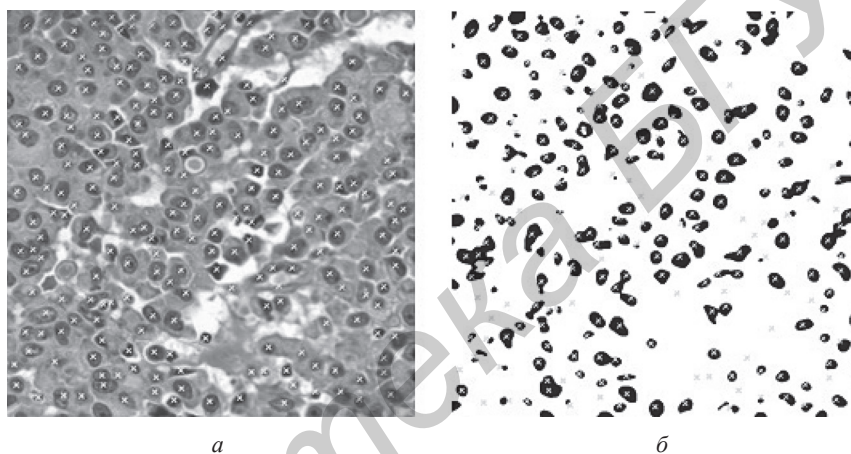


Рис. 5. Пример определения ядер клеток

2.3. Результаты экспериментов и оценка эффективности предложенной методики.

Предложенная методика реализована в системе Matlab. Был проведен ряд экспериментов по оценке ее эффективности с использованием описанной выше тестовой базы изображений. Определены критичные параметры алгоритмов, настраиваемые на каждом этапе анализа изображений. Для каждого параметра был установлен интервал тестируемых значений, а затем реализована процедура выбора не только оптимальных значений каждого параметра, но их из взаимная комбинация, на основе оценки значений точности (*Precision*) и полноты (*Recall*), табл. 1.

Пример результата определения ядер клеток приведен на рис. 5.

Эффективность подсчета ядер клеток оценивалась с использованием следующих значений: *TP* – истинно-положительное решение; *FP* – ложноположительное решение; *FN* –

ложноотрицательное решение. На основе данных параметров вычислялись точность (*Precision*) и полнота (*Recall*), которые являются метриками, используемыми при оценке большей части алгоритмов извлечения информации:

$$Precision = \frac{TP}{TP + FP}, \quad (1)$$

$$Recall = \frac{TP}{TP + FN}. \quad (2)$$

F-мера представляет собой гармоническое среднее между точностью и полнотой. Она стремится к нулю, если точность или полнота стремится к нулю:

$$F = 2 \frac{Precision \cdot Recall}{Precision + Recall} \quad (3)$$

Полученные результаты сравнивались с последними достижениями в данной области на указанной тестовой базе [7, 9–10], табл. 2.

Таблица 2. Сравнение эффективности предложенной методики с последними достижениями в исследуемой предметной области

	Предложенная методика	Wiennert	Al-Kofahi
Точность	0,8363±0,0120	0,908±0,04	0,707±0,13
Полнота	0,9570±0,0016	0,859±0,04	0,916±0,04
F-мера	0,8926	0,8828	0,7980

3. Обсуждение результатов

Табл. 2 показывает эффективность предложенной методики на описанной тестовой базе изображений. Разработанный подход может быть адаптирован к различным задачам подсчета ядер клеток. Его главное преимущество – это автоматический подход к подсчету ядер клеток. В процессе анализа на этапах фильтрации и морфологической обработки были «потеряны» ядра с небольшой площади. Однако это позволило получить более достоверные результаты для других ядер клеток, что связано с улучшением результатов бинаризации, когда исчезают области, не являющиеся ядра-

ми. Дальнейшие исследования будут направлены на повышение эффективности этапа сегментации и расширение набора тестовых баз изображений.

Заключение

Предложена методика автоматического обнаружения ядер клеток и подсчета их числа на основе направленной фильтрации, алгоритма бинаризации Sauvola и морфологических операций. В отличие от известных алгоритмов данный подход не предусматривает сегментацию ядер клеток, а только предполагает обнаружение и оценку их количество. Это позволяет избежать сложных алгоритмических вычислений и обеспечивает хорошую точность подсчета ядер клеток. Прототип программного обеспечения, разработанного в рамках проведенного исследования, можно рассматривать как автоматический инструмент для анализа ядер клеток. Работа частично выполнена в рамках гранта фонда фундаментальных исследований № Ф15ЛИТ-031.

Литература

1. **Methods** For Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review – Current Status and Future Potential / Irshad, H. [et. al] // IEEE Reviews in Biomedical Engineering, 2014. – Vol. 7. – P. 97–114.
2. **Recent** Advances in Morphological Cell Image Analysis / Chen S. [et. al] // Computational and Mathematical Methods in Medicine, 2012. – 10 p.
3. **Jung C., Kim C.** Impact of the Accuracy of Automatic Segmentation of Cell Nuclei Clusters on Classification of Cell Nuclei Clusters on Classification of Thyroid Follicular Lesions / C. Jung, C. Kim // Cytometry. Part A, 2014. – P. 709–719.
4. **Multi-resolution** Approach for Combining Visual Information using Nuclei Segmentation and Classification in Histopathological Images / Saharma, H. [et. al] // Proceedings of the 10th International Conference on Computer Vision, Theory and Applications, 2015. – P. 37–46.
5. **Alilou, M.** Segmentation of cell nuclei in heterogeneous microscopy images: A resalable templates approach / M. Alilou, V. Kovalev, V. Taimouri // Computerized Medical Imaging and Graphics, 2013. – Vol. 37. – P. 488–499.
6. **Kowal, M.** Nuclei Segmentation for Computer-Aided Diagnosis of Breast Cancer / M. Kowal, P. Filipczuk // Int. Journal Appl. Math. Comput. Science, 2014. – Vol. 24. – No. 1. – P. 19–31.
7. **Detection** and segmentation of cell nuclei in virtual microscopy images: a minimums-model approach / Wienert, S. [et al] // National Scientific Reports, 2012. – 2:503.
8. **White** Blood Cell Segmentation by Color-Space-Based K-Means Clustering / Zang, C. [et. al] // Sensors, 2014. Vol. – 14. – P. 16128–16147.
9. **Region-based** progressive localization of cell nuclei in microscopic images with data adaptive modeling / Y. Song [et. al] // BMC Bioinformatics. – 2013. Vol. 14. № 1. – P. 173–180.
10. **Coelho, L. P.** Nuclear segmentation in microscope cell images: a hand-segmented dataset and comparison of algorithms / L. P. Coelho, A. Shariff, R. F., Murphy // Proc. of the IEEE International Symposium Biomedical Imaging, 2009. – P. 518–521.
11. **Wavelet-Based** Multiscale Texture Segmentation: Application to Stromal Compartment Characterization on Virtual Slides / N. Signolle [et. al] // Signal Processing, 2010. – Vol. 90. – № 8. – P. 2412–2422.
12. **Segmentation** of cytological image using color and mathematical morphology / O. Lezoray [et. al] // Acta Stereologica, 1999. – 18. – P. 1–14.
13. **An** Image Analysis-Based Approach for Automated Counting of Cancer Cell Nuclei / Loukas, C. G. [et. al] // Cytometry. Part A, 2003. – P. 30–42.
14. **Improved** automatic detection and segmentation of cell nuclei in histopathology images / Al-Kofahi, Y. [et. al] // IEEE Trans. on Biomedical Engineering, 2010. – Vol. 57. – № 4. – P. 841–852.
15. **Cell-based** quantification of molecular biomarkers in histopathology specimens / Al-Kofahi, Y. [et. al] // Histopathology, 2011. – 59(1) – P. 40–54.

16. Image Smoothing via L_0 Gradient Minimization / L. Xu [et. al] // ACM Transactions on Graphics. – December 2011. – Vol. 30. – No. 6. – article 174.
17. He K., Sun J., Tang X. Guided image filtering // Pattern Analysis and Machine Intelligence, IEEE Transactions on. – 2013. – T. 35. – №. 6. – С. 1397–1409.
18. Wan S. J., Prusinkiewicz P., Wong S. K. M. Variance-based color image quantization for frame buffer display // Color Research & Application. – 1990. – T. 15. – №. 1. – С. 52–58.
19. Sauvola, J. Adaptive document image binarization / J. Sauvola, M. Pietikainen // Pattern Recognition, 2000. – Vol. 33. – P. 225–236.
20. Shafait, F. Efficient implementation of local adaptive thresholding techniques using integral images / Shafait, F., Keysers, D. and Breuel, T. M. // Document Recognition and Retrieval XV. – 2008.
21. Stathis, P. An Evaluation Technique for Binarization Algorithms / Stathis, P., Kavallieratou E., Papamarkos N. // Journal of Universal Computer Science, 2008. – Vol. 14. – No. 18. – P. 3011–3030.

Поступила
15.04.2016

После доработки
30.04.2016

Принята к печати
10.05.2016

M. M. Lukashevich, V. V. Starovoitov

AN APPROACH TO CELL NUCLEI COUNTING IN HISTOLOGICAL IMAGE ANALYSIS

In the paper a method of automatic counting the number of cell nuclei in histological images is studied. This operation is commonly used in the diagnostics of various diseases and morphological analysis of cells. In this connection, the procedure of automatic count the number of cell nuclei is a key step in the systems of medical imaging microscopic analysis of histological preparations. The main aim of our work was to develop an efficient scheme of automatic counting cell nuclei based on advanced image processing methods: directional filtering, adaptive image binarization and mathematical morphology. Unlike prior research, the presented approach does not provide segmentation of cell nuclei in the image, but only requires to detect them and count their number. This avoids complex algorithmic calculations and provides good accuracy of counting cell nuclei.

The paper describes a series of experiments conducted to assess the effectiveness of the proposed method using the available online database of medical test histological images. Critical parameters defined algorithms, configurable at each stage of image analysis. For each parameter we have defined value ranges, and then realized a selection of optimal values for every parameter and a mutual combination of them. It is based on generally accepted quantitative measures of precision and recall. The results were compared with the state-of-art investigations in this field and demonstrated an acceptable level of accuracy of the proposed method. The software prototype developed during the study can be regarded as an automatic tool for analysis of cell nuclei. The presented approach can be adapted to various problems of analysis of cell nuclei of various organs.



Лукашевич Марина Михайловна, кандидат технических наук. Доцент кафедры ЭВМ учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», г. Минск, Беларусь. Сфера научных интересов: обработка изображений, распознавание образов, системы технического зрения.



Старовойтов Валерий Васильевич, доктор технических наук, профессор. Главный научный сотрудник ОИПИ НАН Беларуси. Лауреат премии Ленинского комсомола БССР и Государственной премии Республики Беларусь. С 2000г. преподает по совместительству в разных университетах курсы, связанные с обработкой и анализом цифровых изображений. Сфера научных интересов: обработка и анализ цифровых изображений, полученных в разных участках электромагнитного спектра. Опубликовал более 150 работ в научных журналах и трудах конференций, 3 монографии. Индекс Хирша по данным Академии Google равен 12. Подготовил 6 кандидатов наук.

E-mail: valerys@newman.bas-net.by