

УДК 303.732.4

МЕТОДОЛОГИЯ, МОДЕЛИ И АЛГОРИТМЫ ОБРАБОТКИ ИЗОБРАЖЕНИЙ ЛИМФАНГИОГЕНЕЗА ДЛЯ СРАВНЕНИЯ И ПРОГНОЗА РАЗВИТИЯ ЗЛОКАЧЕСТВЕННЫХ ОПУХОЛЕЙ

© М. В. Спринджук, Ю. Е. Демидчик, С. Е. Шелкович, А. П. Кончиц

Объединенный Институт Проблем Информатики Национальной академии наук Беларуси

лаборатория математической кибернетики

ул. Сурганова, 6, г. Минск, 220012, Беларусь

e-mail: sprindzuk@yahoo.com

**METHODOLOGY, MODELS AND ALGORITHMS OF IMAGE PROCESSING OF THE
LYMPHANGIOGENESIS PATHOLOGY FOR THE GOALS OF MALIGNANCY EXPANSION
COMPARISON AND PROGNOSIS.**

Sprindjuk, M. V., Demidchik, Yu. E., Shelkovich, S. E and Konchiz, A. P.

Abstract. Authors report their 5-year experience in lymphangiogenesis image processing and tumor expansion research. Texture and the novel for that research field fractal image features can be an adjunctive to the conventional microvessel density computation. Practical data processing of clinical data in the sphere of lymphangiogenesis discovery has elucidated the questions of data verification, appropriate organization, computability and the adequacy of prognostic algorithm application.

ВВЕДЕНИЕ

С появлением новых математических инструментов и развитого алгоритмического аппарата для прикладной статистики прогнозирование клинического состояния больных стало популярной темой в современных научных публикаций ведущих научных коллективов мира. Поиск прогностических биомаркеров исхода злокачественных опухолей — трудоемкое занятие, процесс, требующий получения и обработки сложных данных. Уникальность и длительный период времени получения и накопления необходимого клинического и патоморфологического материала делает эксперимент неповторимым и усложняет организацию данных и возможность сравнения результатов с аналогичными исследованиями. Задачи обработки данных, включающие работу со сложными программными продуктами и написание оригинального программного обеспечения под силу только специалистам с техническим образованием и соответствующим опытом. Целью настоящего исследования было разработать и апробировать модели и алгоритмы обработки изображений опухолевого лимфангиогенеза и для конкретного примера вычислить сравнительные веса предикторов-параметров, полученных с изображений.

1. МАТЕРИАЛЫ И МЕТОДЫ

Изображения были получены с биологического материала удаленных опухолей щитовидной железы и яичников. Оба класса препаратов были окрашены маркером D2-40. Статистическим материалом была информация, включающая номер-идентификатор пациента и результат развития опухолевого процесса: для карциномы щитовидной железы — метастазирование в лимфатические узлы, для рака яичников — рецидив болезни. Характеристики изображений были получены оригинальным программным обеспечением. Под разработанными алгоритмами понимались шаги действий, совершенных над изображениями для получения видимых и невидимых человеческому глазу численных характеристик, полноценно описывающих патоморфологическую картину заболевания. Моделью в данном исследовании была системно-аналитическая концепция организации и обработки данных, направленная на достижения исходной цели научного исследования (см. рисунки 1–3). Для разработки программного обеспечения применялись стандартные языки программирования: Дельфи и Питон. Их выбор был обусловлен знанием и опытом разработчиков и наличием необходимых библиотек кода. Статистическая обработка данных выполнялась и была перепроверена различным программным обеспечением: R, SPSS20, STATA11, Minitab, STATISTICA, StatsDirect, JMP SAS, MATHEMATICA.

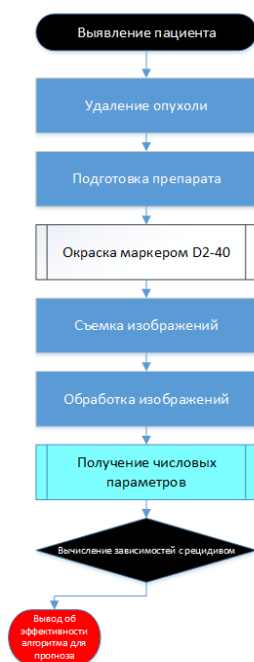


Рис. 1. Общая схема выполненного исследования.



Рис. 2. Системно-аналитическая модель принятия решения о включении пациента в группу риска по прогнозируемому событию.

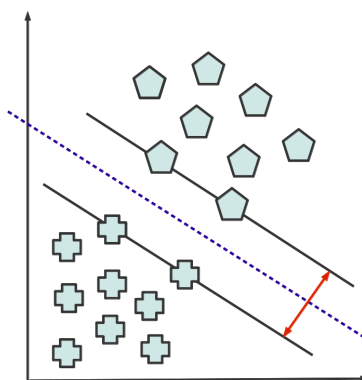


Рис. 3. Бинарная SVM-модель (support vector machines — опорные машинные вектора, сокращенно — ОВ). Синим пунктиром представлена оптимальная плоскость классификации, красной линией — оптимальная граница разделения классов. Объекты на сплошной линии — опорные вектора.

2. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

Наиболее распространенный маркер-дескриптор ангиогенеза — площадь прокрашенных сосудов. В данной работе, помимо этой характеристики, которая была необходима для пополнения коллекции аналогичных исследований, мы впервые для данной предметной области изучили прогностические возможности применения текстурных характеристик изображений, таких, в своем роде фрактальных параметров, как однородности распределения сосудов на плоскости изображения, долей крупных и мелких объектов сосудов. На рисунках 4–10 представлены результаты применения различных прогностических алгоритмов для прогноза исхода рака на основе полученных характеристик изображений.

Метод	Общее число выборки по классам "0/1"	Число неправильно классифицированных для класса "0"	Число неправильно классифицированных для класса "1"	Процент неправильно классифицированных для класса "0"	Процент неправильно классифицированных для класса "1"
ОВ линейное ядро	7/6	5	0	71	0
ОВ сигмовидное ядро	7/6	7	0	100	0
Ближайших соседей	3/10	1	5	33,3	50
Линейный дискриминантный анализ	15/33	0	33	0	100
Бинарная логистическая регрессия	15/33	3	7	20	21

Рис. 4. Сводка результатов алгоритмов машинного обучения для классификации метастазирования по группе характеристик изображений.

Математическая основа алгоритма вычисления однородности распределения сосудов на изображении представлена в формуле (1), где H — энтропия, t — верхний предел, p_i — вероятность попадания чёрного, представляющего микрососуды, пиксела в i -й фрагмент.

$$H = - \sum_{i=0}^t p_i \ln(p_i) \quad (1)$$

Метод	Общее число выборки по классам "0/1"	Число неправильно классифицированных для класса "0"	Число неправильно классифицированных для класса "1"	Процент неправильно классифицированных для класса "0"	Процент неправильно классифицированных для класса "1"
ОБ линейное ядро	14/7	0	7	0	100
ОБ сигмоидное ядро	14/7	9	2	64,3	28,6
Ближайших соседей	14/7	4	4	28,6	57,1
Линейный дискриминантный анализ	51/30	22	11	43	36,7
Бинарная логистическая регрессия	51/30	19	7	37	23

Рис. 5. Сводка результатов алгоритмов машинного обучения для классификации рецидива по группе характеристик изображений.

Параметр	СтОшибка	ChiSquare	Вероятность > ChiSquare
Однородность	5.64597	5.09	0.0241
Площадь	0.106462	7.7	0.0055
ОДКОС	7.05029	2.32	0.1279
ОДМОС	14.8346	2.44	0.1182

Рис. 6. Данные расчета бинарной логистической регрессии для метастазирования в лимфатические узлы рака щитовидной железы на основе отдельных значений характеристик изображений.

Параметр	СтОшибка	ChiSquare	Вероятность > ChiSquare
Однородность	4.73458	1.01	0.0241
Площадь	0.444415	7.7	0.1263
ОДКОС	15.3748	5.18	0.0229
ОДМОС	33.5237	4.72	0.0298

Рис. 7. Данные расчета бинарной логистической регрессии для метастазирования в лимфатические узлы рака щитовидной железы на основе отдельных значений характеристик изображений.

Этот результат исследования уже сообщался нами в других публикациях как и описание других применявшихся алгоритмов для получения характеристик изображений [1-14], поэтому подробно в данной статье не рассматривается.

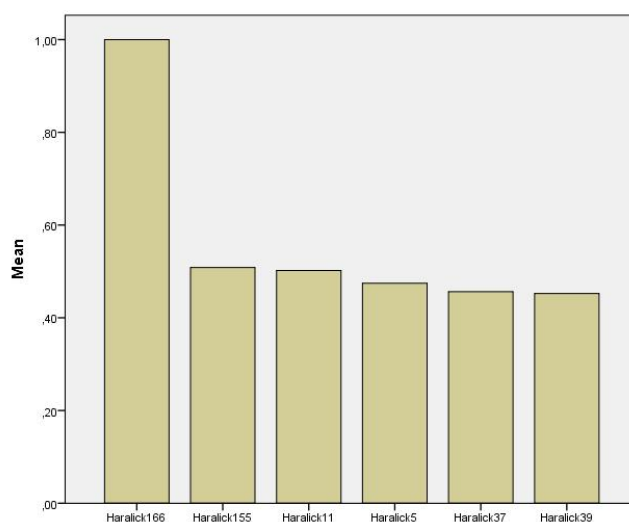


Рис. 8. Веса значений текстурных характеристики изображений карциномы щитовидной железы (идентификатор соответствует ряду значений). На первом месте — 13-ая математическая функция Харалика в 12-ом векторном пространстве.

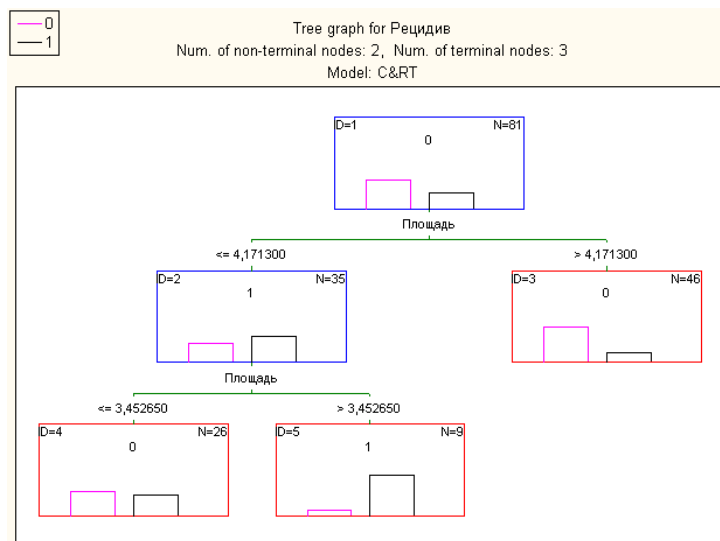


Рис. 9. Дерево, автоматически построенное на основе дискриминирующих значений площади сосудов опухоли. Показаны точные значения площади сечения лимфатических капиляров опухоли, позволяющие судить о прогнозе рецидива болезни [<http://www.statistica.ru/glossary/general/derevya-klassifikatsii/>].

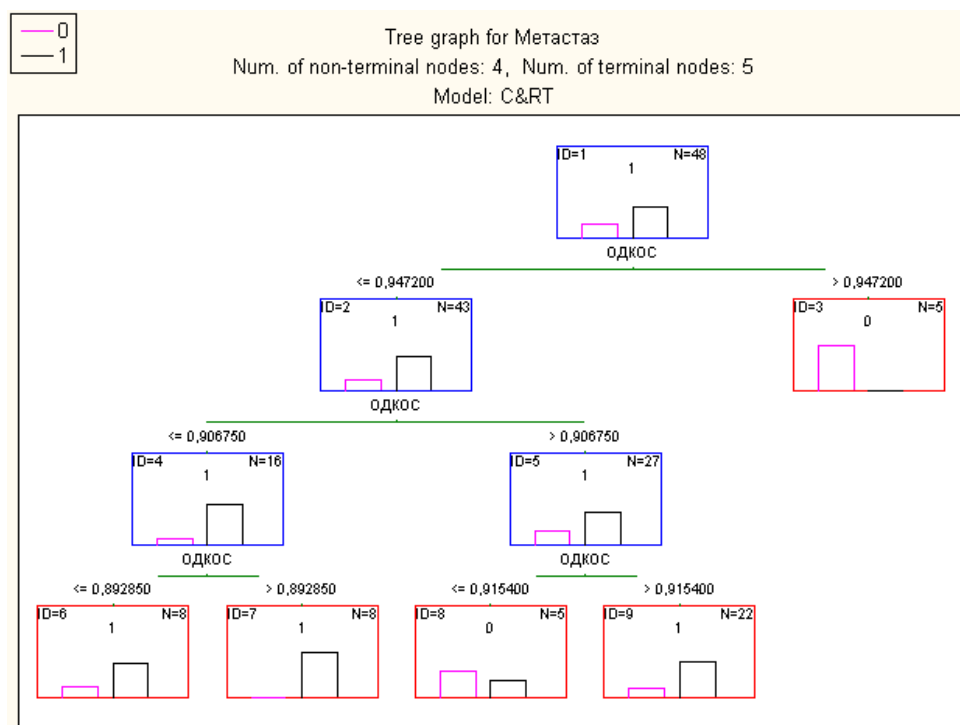


Рис. 10. Дерево, автоматически построенное на основе дискриминирующих значений ОДКОС (относительной доли крупных объектов сосудов) и ОДМОС (относительной доли мелких объектов сосудов) опухоли щитовидной железы.

“Деревья решений” (классификации) — метод, позволяющий предсказывать принадлежность наблюдений или объектов к тому или иному классу категориальной зависимой переменной в соответствии со значениями одной или нескольких предикторных переменных. Основные понятия из теории “деревьев решений”:

- Объект — пример, шаблон, наблюдение;
- Атрибут — признак, независимая переменная, свойство;
- Метка класса — зависимая переменная, целевая переменная, признак, определяющий класс объекта;
- Узел — внутренний узел дерева, узел проверки;
- Лист — конечный узел дерева, узел решения;
- Проверка — условие в узле.

Отличительная черта метода “деревьев классификации” — это присущая ему гибкость, — способность “деревьев классификации” последовательно изучать эффект влияния отдельных переменных. Есть еще целый ряд причин, делающих деревья

классификации более гибким средством, чем традиционные методы анализа. Широкая сфера применения “деревьев классификации” делает их весьма привлекательным инструментом анализа данных, но не следует полагать, что его рекомендуется использовать вместо традиционных методов статистики. Напротив, если выполнены более строгие теоретические предположения, налагаемые традиционными методами, и выборочное распределение обладает некоторыми специальными свойствами, то более результативным будет использование именно традиционных методов. Как метод разведочного анализа или последнее средство, в случае отказа всех традиционных методов.

Основными достоинствами метода "деревья решений" являются:

- быстрый процесс обучения;
- генерация правил в областях, где эксперту трудно формализовать свои знания;
- извлечение правил на естественном языке;
- понятная на интуитивном уровне классификационная модель;
- достаточно высокая точность прогноза, сопоставимая с другими методами (нейронными сетями, случайными лесами);
- построение непараметрических моделей [<http://cdo.bseu.by/stat1/lab7>].

При изучении прогностического значения комплекса характеристик изображений наилучшие в плане корректной классификации и, соответственно, прогноза, результаты были получены при применении бинарной логистической регрессии. При изучении прогностического веса индивидуальных характеристик изображений наименьшая стандартная ошибка классификации была выявлена для параметров площади и однородности. Учитывая разную природу характеристик изображений, мы не использовали текстурные параметры Харалика в одном стеке предикторов при расчете их прогностического веса.

Прогностический вес численных значений текстуры, которые дали корреляцию $<0,05$ по двустороннему критерию, представлены на рисунке 8. Учитывая небольшое число пациентов и изображений (81 карцинома яичников, 48 карцином щитовидной железы), мы не стали изучать индивидуальное прогностическое значение каждого численного значения полученного текстурного признака Харалика, оставив новизну воплощения этой идеи для других исследователей. Применяя таким образом стандартный математический функциональный аппарат, мы предложили новые морфометрические характеристики, алгоритмы, системно-аналитическую методологию для узкой области изучения морфологической картины лимфангиогенеза и прогноза опухолевой агрессии. В данном исследовании изучение текстуры и фрактальных параметров уникальной группы клинических патоморфологических изображений было

направлено не на распознавание образов и классификацию изображений, что не имело бы значительной новизны в связи с большим количеством аналогичных публикаций, а на получение с изображений двух классов численных характеристик, которые можно было использовать для сравнения эволюции опухолей, их развития и прогноза. Опыт изучения морфометрии лимфангиогенеза, позволил нам сформулировать методологию её применения:

- Морфометрию лимфангиогенеза целесообразно выполнять для фундаментального изучения морфологической картины злокачественных опухолей.
- Числовые параметры, полученные от обработанных программным обеспечением изображений, представляют собой *дополнительную информацию* об изучаемом организме и могут быть использованы для определения весовых коэффициентов прогноза клинических фактов, характеризующих состояние пациента в сравнении с другими параметрами.
- Рекомендуется формировать базы данных морфометрических параметров для возможности оценки эволюции морфометрической картины опухолей и возможности сформировать значительную по объему выборку для научных исследований.
- Рекомендуется применять площадь покрашенных маркером сосудов и значение однородности распределения сосудов на изображении как основные дескрипторы опухолевой картины лимфангиогенеза.
- Рекомендуется использовать маркер D2-40 для морфометрии лимфангиогенеза опухолей яичников и щитовидной железы.
- Рекомендуются дальнейшее изучение прогностической ценности морфометрических параметров лимфангиогенеза в сравнении с лабораторными и клиническими предикторами.
- Не рекомендуется применять морфометрию лимфангиогенеза для попыток установления и уточнения клинического диагноза у онкологических пациентов.

ЗАКЛЮЧЕНИЕ

Опыт изучения прогностических возможностей морфометрии показал, что метастазирование щитовидной железы лучше поддается прогнозу, чем рецидив рака яичников, что обусловлено большей агрессией рака яичников и, соответственно, более динамичным экспансивным процессом распространения опухолевых клеток репродуктивной системы. За 5 лет исследования данные о метастазировании щитовидной железы не изменились, но к 5-му году исследования количество случаев рецидива

рака яичников увеличилось относительно того временного момента, когда были получены данные об исходе заболевания. Стратификация данных по разным признакам и свойствам выборки для улучшения однородности данных была невозможна в связи с ограниченным количеством биологических объектов исследования. Таким образом, опыт изучения разных граней развития злокачественных опухолей позволил сформировать рекомендации для следующих экспериментов в этой области. Применение морфометрии опухолевого ангио- и лимфангиогенеза показывает, что *накопленная морфометрическая информация может служить источником новых исследований взаимоскорреляции и другой зависимости* клинических и лабораторных данных. Однако, безусловно, такая информация по своей природе не может иметь ключевого диагностического характера. Пригодность характеристик изображений для прогноза исхода карцином должна изучаться на верифицированном материале большого объема и однородной структуры по разным критериям с возможностью стратифицировать пациентов по группам различных параметров.

СПИСОК ЛИТЕРАТУРЫ

1. *Сприндзук М. В.* Цифровая обработка гистологических изображений: введение в область исследования и алгоритмы для морфометрии изображений ангиогенеза и лимфангиогенеза // *Врач и Информационные Технологии.* – 2012. – № 4. – С. 58–68.
Sprindzuk M. V. 2012. *Digital processing of histological images: introduction to the research field and the algorithms for the morphometry of angiogenesis and lymphangiogenesis.* *Vrach i Informaciyonnie Technologii*, 4, pp. 58–68.
2. *Сприндзук М. В., Ковалев В. А., Кончиц А. П., Демидчик Ю. Е., Фридман М. В., Богущ А. Л., Ляховский В. В.* / Ангиогенез: программное обеспечение для исследования феномена и вопросы количественной оценки гистологических изображений // *Врач и Информационные Технологии.* – 2010. – № 3. – С. 38–46.
Sprindzuk M.V., Konchits A.P., Bogush A.L., Liakhovski V.V. 2010. *Angiogenesis: software for the research of the phenomena and the questions of the quantitative evaluation of the histological images.* *Vrach i Informaciyonnie Technologii*, 3, pp. 38–46.
3. *Сприндзук М. В., Кончиц А.П., Богущ А. Л.* / Коррелятивный анализ плотности лимфатических капилляров и параметров, полученных при анатомическом и гистологическом исследовании образцов папиллярного рака щитовидной железы в группе 43 педиатрических пациентов // *Хирург.* – 2010. – № 11. – С. 7–14.
Sprindzuk M.V., Konchits A.P., Bogush A.L. 2010. *Correlative analysis of the lymphatic microvessel density and parameters derived on anatomical and histological exploration of the 43 juvenile papillary thyroid cancer patients.* *Hirurg*, 11, pp. 7–14.

4. Спринджук М. В. Ангиогенез // Морфология. – 2010. – № 11. – С. 4–13.
Sprindzuk M. V. 2010. *Angiogenesis*. Morphologia, 11, pp. 4–13.
5. Спринджук М.В., Ковалев В.А., Богущ А.Л., Шелкович С.Е. Ангиогенез злокачественных опухолей щитовидной железы и яичников. Возможности цифровой обработки микроскопических изображений // ARS medica. Искусство медицины. – 2010.- № 3. — С. 9–24.
Sprindzuk M.V., Kovalev V.A., Bogush A.L., Shelkovich S.E. 2010. *Angiogenesis of malignant tumors of thyroid gland and ovaries. Capabilities of the digital image processing*. ARS Medica, No. 3, P. 9–24.
6. Спринджук М. В., Ковалев В. А., Фридман М. В. Ангиогенез: значение в современной медицине, ключевые аспекты патогенеза, проблемы оценки изображений, полученных при исследовании гистологических микропрепаратов // Медицинская Панорама. — 2009. — № 10. — С. 5–9.
Sprindzuk M. V., Kovalev V. A., Fridman M. V. 2010. *Angiogenesis: the role in modern medicine, key aspects of pathogenesis, questions of the histology sample image evaluation*. Meditsinskaya Panorama, 10, pp. 5–9.
7. Спринджук М.В., Кончиц А.П. Рост и развитие опухолевых лимфатических капилляров при папиллярном раке щитовидной железы /М.В. Спринджук, А.П. Кончиц // Онкохирургия. – 2012. – №3. –С. 12-23.
Sprindzuk M.V., Konchits A.P. 2012. *Growth and development of the tumoral lymphatic capillaries in papillary thyroid cancers*. Oncohirurgiya, pp. 12–23.
8. Спринджук М.В. Ангиогенез. /М.В. Спринджук // Вестник Восстановительной Медицины. – 2011. – №5. – С. 23–32.
Sprindzuk M.V. 2011. *Angiogenesis*. Vestnik Vosstanovitel'noi Meditsini, 5, pp. 23–32.
9. Спринджук М.В., Кончиц А.П. Опыт пилотного исследования патоморфологии лимфатических капилляров папиллярного рака щитовидной железы по данным обработки изображений, окрашенных маркером D2-40. /М.В. Спринджук, А.П. Кончиц // Современные технологии в медицине – 2011. – №3. – С. 48-53.
Sprindzuk M.V., Konchits A.P. 2011. *The pilot experience of the pathomorphological exploration of the lymphatic microvessels of thyroid cancer based on image processing of the images produced from D2-40 stained samples*. Sovremennie Technologii v Meditsine, 3, pp. 48–53.
10. Sprindzuk, M. V., Konchits, A. P., Shukelovich, A. V., Demidchik, Y. P., Shelkovich, S. E., Fridman, M. V. 2014. The technology and algorithms for tumoral behavior prediction based on computerized lymphangiogenesis morphometry of D2-40 stained images. *Information Control Management Systems*, 1, pp. 65–70.
11. Sprindzuk, M. V., Konchits, A. P., Shukelovich, A. V., Demidchik, Y. P., Shelkovich, S. E., Fridman, M. V. 2013. The method for ovarian cancer prediction based on computerized morphometry of D2-40 stained images. *International Conference on Digital Technologies, Slovakia, Zilina, IEEE*, 1, pp. 52–54.
12. Sprindzuk, M. V., Dmitruk, A. A., Kovalev, V. A., Bogush, A. L., Tuzikov, A. V., Liakhovski, V. V., Fridman, M. V. 2011. Computer-aided image processing of angiogenic histological samples in ovarian cancer. *Journal of Clinical Medicine Research*, 1, pp. 10–15.

13. Sprindzuk, M. V. 2010. Angiogenesis in Malignant Thyroid Tumors. *World Journal of Oncology*, 1, pp. 221–223.
14. Sprindzuk, M. V. 2010. Angiogenesis in thyroid malignant neoplasm: State of the art and advances of the modern digital pathology and nanotechnology. *Journal of Clinical Pathology and Forensic Medicine*, 1, pp. 16–34.

Статья поступила в редакцию 16.01.2014

Авторы выражают благодарность сотрудникам Минского Онкологического Диспансера и лаборатории патоморфологической диагностики БелМАПО, а также профессорам-математикам: Михаилу Яковлевичу Ковалеву, Александру Николаевичу Вальвачеву, Борису Андреевичу Залесскому, Виктору Владимировичу Краснопрошину, Александру Васильевичу Тузикову, Василию Ивановичу Бернику, Элле Ивановне Ковалевской, Виктору Абрамовичу Мильману, Юрию Назаровичу Сотскову и Валерию Васильевичу Старовойтову.

Исследование имело финансовую поддержку фондов МНТЦ и CRDF