

УДК 004.932

ЭКСПЕРТНО-АНАЛИТИЧЕСКАЯ СИСТЕМА ОБРАБОТКИ И АНАЛИЗА МАММОГРАММ

Статья поступила в редакцию 02.12.2013, в окончательном варианте 15.02.2014.

Буланова Юлия Анатольевна, аспирант, Муромский институт (филиал) Владимирского государственного университета имени А. Г. и Н. Г. Столетовых», 602264, Россия, Владимирская область, г. Муром, ул. Орловская, 23, e-mail: yuliyabulanova@yandex.ru

Актуальность темы статьи обоснована с позиций обеспечения доступности и качества проведения маммографического скрининга в большинстве лечебных учреждений России. В статье проведен обзор существующих систем обработки и анализа маммограмм, показаны их особенности. Подробно рассмотрены и обоснованы цели разработки экспертно-аналитической системы, а также сформулированы основные требования к ней. Автором описана структура разработанной им экспертно-аналитической системы (ЭАС) для обработки и анализа маммограмм. Эта ЭАС включает несколько модулей, которые в статье представлены в виде блок-схемы. Результаты проведенных экспериментов по обработке и анализу маммографических снимков с использованием ЭАС и без нее (в ручном режиме), показали целесообразность применения ЭАС для диагностики новообразований. Преимущество ЭАС – высокая точность распознавания новообразований, сравнительно малый процент ложноположительных результатов.

Ключевые слова: маммограмма, патологии молочной железы, киста, фиброаденома, рак молочной железы, система обработки и анализа маммограмм, САД системы, экспертно-аналитическая система, предварительная обработка, текстурная сегментация, распознавание

EXPERT ANALYTICAL SYSTEM OF PROCESSING AND ANALYSIS OF MAMMOGRAMS

Bulanova Yuliya A., post-graduate student, Murom Institute (branch) of Vladimir State University named after Alexander and Nicholay Stoletovs, 23 Orlovskaya St., Murom, Vladimir region, 602264, Russian Federation, e-mail: yuliyabulanova@yandex.ru

Actuality of theme of Article justified in terms of ensuring access to and quality of mammographic screening in most hospitals in Russia. The article provides an overview of existing systems for processing and analysis of mammograms showing their features. The goal of developing of expert-analytical system discussed in detail and justified and the main requirements to it. The author describes the structure developed by him expert-analytical system (EAS) of processing and analysis of mammograms. This expert-analytical system includes several modules, that the article shown in block diagram form. The results of experiments on the processing and analysis of mammography images using of the EAS and without (in manual mode) showed the feasibility of the EAS for diagnosis of tumors. Advantages of EAS are high recognition accuracy neoplasms, a relatively small percentage of false positive results.

Keywords: mammogram, breast pathology, cyst, fibroadenoma, breast cancer, the system processing and analysis of mammograms, CAD systems, expert-analytical system, preprocessing, texture segmentation, recognition

Развитие медицинских систем для маммографического исследования приводит к снижению цены ошибок, связанных с человеческим фактором. Именно человек является основным источником возникновения ошибок при расшифровке маммографического снимка, а следовательно, и причиной пропуска опасных заболеваний на стадиях, поддающихся эффективному лечению. Снизить субъективность оценки маммограмм возможно путем автоматизации процесса обработки и анализа рентгенологических изображений на маммограммах.

Целью данной работы было создание экспертно-аналитической системы (ЭАС) для обработки и анализа маммограмм.

Исходя из цели работы, были сформулированы следующие задачи:

- 1) выполнить обзор и анализ существующих систем обработки/анализа маммограмм;
- 2) сформулировать основные требования к разрабатываемой ЭАС;
- 3) проанализировать целесообразную структуру этой ЭАС;
- 4) описать основные этапы реализации ЭАС;
- 5) протестировать работу ЭАС и сделать выводы о качестве получаемых результатов.

Обзор существующих систем обработки маммограмм. Диагностика заболеваний молочной железы является приоритетным направлением в современной медицине – в силу высокой частоты заболеваний и тяжести их последствий при несвоевременном лечении. Одним из неинвазивных методов выявления этих патологий является маммография [4, 7–9, 19].

На мировом рынке существует несколько фирм, предлагающих компьютерные системы анализа маммограмм (от англ. Computer Aided Mamogramm Analysis (CAMA)). Самые распространенные САМА системы: AccuDetect, Second Look Digital, The MAMMEX MammoCAD, TotalLook MammoAdvantage, Syngo MammoCAD, Fujifilm's Digital Mammography System. Рассмотрим их подробнее.

AccuDetect – система обнаружения подозрительных областей как на аналоговой, так и на цифровой маммограммах [18]. Аналитические алгоритмы данной системы позволяют выделить микрокальцинаты и определить их характеристики: плотность кластера (количество кальцинатов в кластере), форма микрокальцинатов и другие особенности. При анализе объекта, подозрительного на опухоль, с помощью системы определяется степень всеобщих периостальных разрастаний (спикул), форма опухоли, контраст по отношению к окружающим тканям, текстуры внутреннего содержания опухоли, текстура края опухоли и другие важные характеристики в отношении определения границ и классификации опухоли. AccuDetect автоматически идентифицирует и отмечает подозрительные области, не затеняя «область интереса», позволяет более точно интерпретировать маммограмму и увеличить вероятность обнаружения микрокальцинатов и опухолей.

MAMMEX MammoCAD – фирма-производитель SCANIS (США) – компьютерная система для выявления рака молочной железы, помогает радиологам в поиске и выявлении патологий на маммограмме. Эти патологии включают микрокальцинаты и опухоли [16]. MammoCAD предоставляет возможность ведения базы данных по пациентам. Система предусматривает корректировку яркости и контраста, просмотр снимков в нескольких проекциях, увеличение и выделение «области интереса» (ROI), вычисление характеристик ROI, быстро выделяет микрокальцинаты и опухоли.

Second Look Digital – компьютерная система распознавания и анализа маммограмм. Она разработана для помощи врачам-рентгенологам в обнаружении рака молочной железы на ранней стадии. Каждая маммограмма обрабатывается по отдельности на наличие любых подозрительных областей. Результаты обработки объединяются в DICOM отчет. Если найдена опухоль или микрокальцинаты, то в отчет записываются координаты патологии [15].

TotalLook MammoAdvantage преобразует аналоговые снимки в цифровые маммограммы для сравнительного обзора на одной рабочей станции. TotalLook MammoAdvantage предлагает: оцифровку аналоговых снимков с настраиваемыми характеристиками выходного изображения, динамическую регулировку качества изображения, интеллектуальную систему сжатия изображения с целью уменьшения требований к пространству и оптимизации скорости передачи данных по сети [14].

Точное воспроизведение изображений TotalLook MammoAdvantage обеспечивается технологией высокой четкости считывания данных на основе датчиков ПЗС (прибор с зарядовой связью) – от англ. Charge-Coupled Device (CCD технология). Это гарантирует точ-

ность считывания деталей изображения. Также в функционал данной системы входят возможности увеличения яркости и контрастности, увеличения четкости края изображения, визуализации тканей молочной железы.

Syngo MammoCAD – система для обнаружения рака молочной железы на самых ранних и наиболее поддающихся лечению стадиях. Есть возможность обработки сразу 4 снимков за 90 секунд. Примерная производительность системы – 40 человек в час (у каждого человека по 4 снимка) [11].

Fujifilm's Digital Mammography System – многофункциональная консоль, которая позволяет обрабатывать изображения (яркость/контрастность, четкость), содержит базу данных о пациентах, позволяет выполнять предварительный просмотр изображения для обработки и печати, работать с DICOM интерфейсом. Система маркирует все подозрительные области (микрокальцинаты, злокачественные и доброкачественные опухоли) [12] (табл.

Таблица 1

Сравнительная характеристика перечисленных систем

Название системы	Производитель (страна)	Ключевые возможности системы	Чувствительность обнаружения опухолей (%)	Доля ложноположительных результатов	Пропускная способность (человек/час)	Совместимость с цифровыми маммографическими комплексами (производители)
AccuDetect	Parascript (США)	Предобработка и распознавание	90,7 %	0,412 %		Agfa Healthcare, Carestream, FujiFilm, GE Healthcare, Hologic, IMS, Philips Healthcare, Planmed, Siemens Medical Solutions
MAMMEX MammoCAD	SCANIS (США)	Предобработка и распознавание	-	-	-	-
Second Look Digital	iCAD (США)	Предобработка и распознавание	90–96 %	2,0–2,9	30	GE Healthcare, Siemens Medical, Fujifilm, Sectra, Agfa HealthCare, Planmed, Philips, IMS Giotto, Hologic
TotalLook Mammo-Advantage	iCAD (США)	Предобработка	-	-	-	GE Healthcare, Siemens Medical, Fujifilm, Sectra, Agfa HealthCare, Planmed, Philips, IMS Giotto, Hologic
Syngo MammoCAD	Siemens (Германия)	Предобработка и распознавание	-	-	40	MAMMOMAT Novation, MAMMOMAT Inspiration
Fujifilm's Digital Mammography System	Fujifilm (Япония)	Предобработка и распознавание	90,6 %	2,5	-	-

Все вышеперечисленные системы имеют закрытый код программного обеспечения (ПО); минимум функций, которые можно узнать только после составления договора на по-

ставку ПО. Дорогостоящие зарубежные системы [17] для большинства лечебных учреждений России малодоступны. Следовательно, актуальным для России является разработка относительно дешевой компьютерной системы диагностики патологий в молочной железе, так как базовая комплектация ПО указанных в правой колонке зарубежных систем получения маммограмм позволяет только изменить качество маммографических снимков.

Требования к создаваемой ЭАС. Первым требованием, предъявляемым к ЭАС, является наличие подсистемы предварительной обработки, которая должна подавлять аппаратный шум маммографа, удалять артефакты, увеличивать яркость и контраст снимков. Маммограмма состоит из трех областей: фон, грудная мышца и область молочной железы. Обработка всего снимка не целесообразна, поэтому подсистема предварительной обработки должна уметь выделять область молочной железы, учитывая индивидуальные особенности каждой женщины.

По статистическим данным, собранным автором, самыми распространенными заболеваниями молочной железы являются киста, фиброаденома и рак. Следовательно, ЭАС должна содержать подсистему выделения областей для трех этих новообразований. После выделения областей для классификации объектов необходим расчет их текстурных характеристик.

Неотъемлемой частью ЭАС является подсистема распознавания (классификации) выделенных областей новообразований.

Кроме того, ЭАС должна позволять в ручном режиме врачу-рентгенологу выделять подозрительные области, проводить обработку и расчет численных характеристик этих областей.

Составной частью ЭАС должен быть модуль работы с базой данных о пациентах. Эта база должна содержать основную информацию о пациенте, результаты медицинского обследования, рентгенограммы молочной железы.

Необходимым требованием для врачей-рентгенологов к ЭАС является формирование отчетов по результатам обследования каждого пациента и статистических отчетов по всем исследованиям.

Кроме того, необходимо предусмотреть блок оцифровки аналоговых маммограмм – на случай использования создаваемой ЭАС совместно с пленочным маммографическим комплексом. В этом блоке необходимо использовать функцию импорта оцифрованного изображения из различных форматов (bmp, png, tiff и т.д.) в формат DICOM-3.0 (это последняя версия протокола DICOM [3], используемого для передачи данных между медоборудованием и медицинскими информационными системами).

Оцифрованные изображения маммограмм могут обрабатываться/анализироваться не только в местах их получения, но также и в специальных медико-аналитических центрах [1].

Если такие центры будут осуществлять обработку данных с ряда маммографов, то это даст возможность значительно снизить суммарную стоимость приобретения специализированного ПО для анализа маммограмм (конечно, при условии заключения соответствующих лицензионных соглашений с разработчиками ПО) [6].

Одновременно может быть обеспечено и повышение качества результатов анализа маммограмм – за счет высокой квалификации врачей-аналитиков, работающих в таких центрах. При обеспечении должной оперативности и мер информационной безопасности при передаче данных о пациентах это предложение вполне укладывается в направления развития телемедицинских технологий в России [3, 6].

Структура системы обработки и анализа маммограмм (СОиАМ) По результатам анализа сформулированных требований к ЭАС, автором была разработана структура СОиАМ, которая отображена на рисунке.

В данной блок-схеме выделены основные блоки операций, связанные с получением, обработкой, анализом и хранением маммограмм.

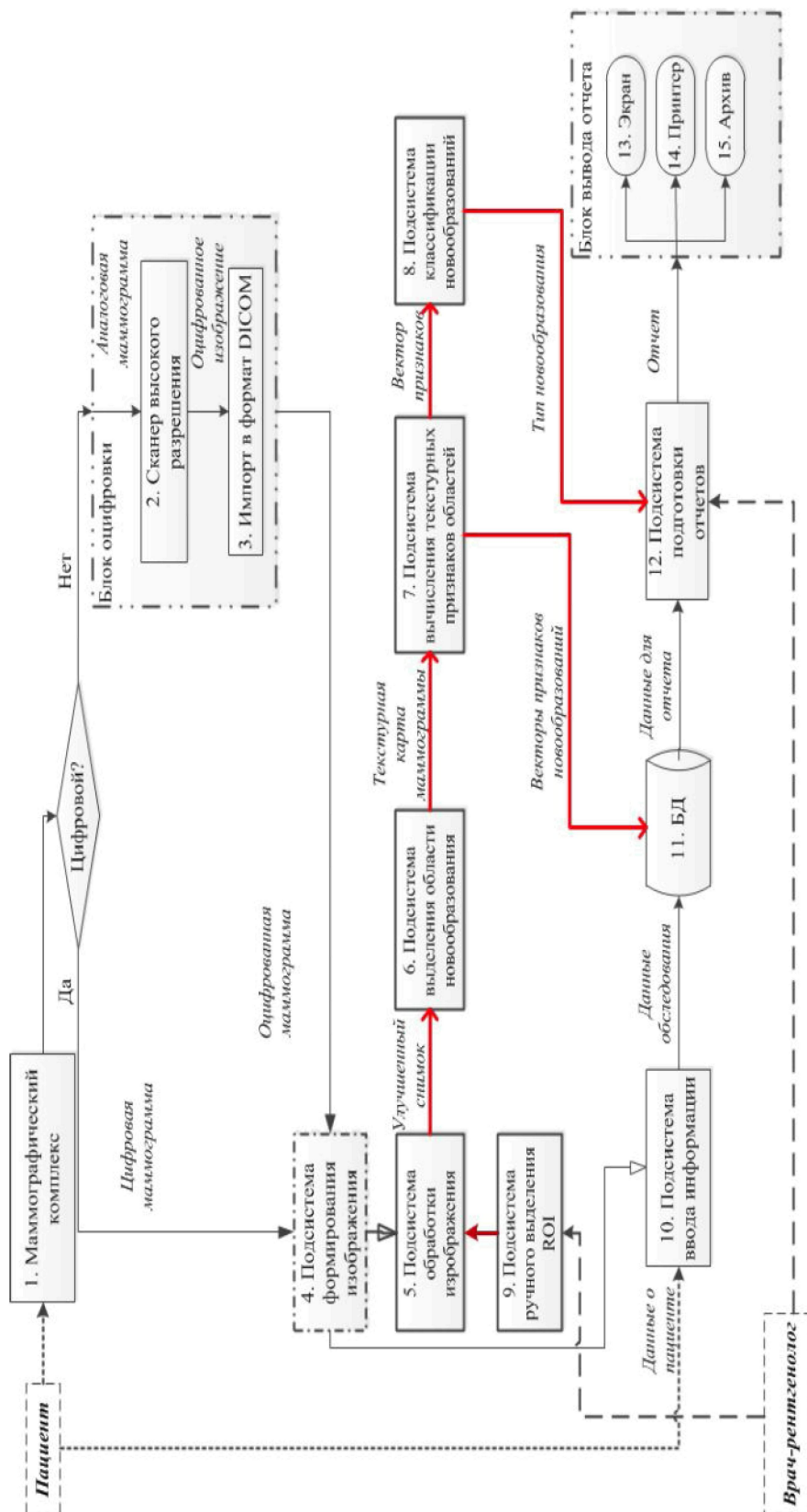


Рис. Блок-схема взаимосвязей программных модулей СОиАМ
Обозначения: ROI (с англ. Region of Interest) – область интереса; БД – база данных

- Данные, поступающие от пациента
- Ручное управление врачом-рентгенологом
- Переходы между этапами работы ЭАС
- Стандартные операции работы с БД
- Способ получения маммограммы для обработки

Основные функции СОиАМ:

- регистрация маммографических снимков;
- автоматическое определение области молочной железы на снимке;
- предварительная обработка снимков, содержащих три типа новообразований;
- выделение областей новообразований трех типов (киста, фиброаденома, рак молочной железы);
- идентификация выделенных областей новообразований;
- визуальный контроль результатов распознавания с возможностью редактирования и выполнения других этапов обработки на любом из шагов алгоритмов распознавания;
- выполнение стандартных операций с базой данных (БД) пациентов;
- хранение всех изменений, внесенных пользователем в процесс редактирования данных;
- генерация файла формата DICOM, содержащего основные данные об исследовании;
- формирование и печать отчета (протокола) маммографического исследования.

Отличительные особенности ЭАС:

- архитектура СОиАМ включает следующие подсистемы: предварительной обработки изображений; выделения новообразований; распознавания новообразований; работы с БД, формирования статистики;
- использование открытого интерфейса подсистемы формирования печатных отчетов маммографического исследования.

Реализация системы. При создании универсальной СОиАМ была предусмотрена возможность ее интеграции как с цифровым, так и с аналоговым маммографическим комплексом. Если результатом первого этапа является пленочная маммограмма, то с помощью блока оцифровки изображения переводятся в цифровой формат. Для оцифровки (этап 2) снимков используется сканер, описанный в [4, 10].

Оцифрованный снимок молочной железы поступает в подсистему предварительной обработки, где сначала формируется модель молочной железы с помощью случайных марковских полей [2].

Предполагаем, что распределение яркостей на изображении принадлежит нормальному закону распределения, тогда оно разбивается на блоки 8×8 пикселей. Формируются компоненты Марковской модели для набора блоков $\mathbf{D} = \{1, \dots, y\}$, где $\mathbf{D}$ – множество блоков, y – количество блоков, на которое разбито изображение.

Для каждой окрестности рассчитывается два показателя: средняя яркость и среднеквадратическое отклонение (СКО) яркости от среднего значения. Определяются достоверные интервалы для каждой области, затем осуществляется объединение одинаковых блоков. Исследования показали, что средняя яркость изображения области молочной железы колеблется от 120 до 150, а СКО – от 12 до 30. Дальнейшая обработка проводится только внутри контура, который очерчивает молочную железу. Использование такой модели для обнаружения области молочной железы дает возможность автоматически ограничивать область для последующей обработки. Кроме того, учитываются индивидуальные особенности строения молочных желез у каждой пациентки. На следующем шаге в этой подсистеме врач, исходя из плотности молочной железы пациентки, улучшает маммографический снимок на основе алгоритмов гистограммного преобразования изображения и фильтрации [9].

Например, улучшение характеристик снимка, содержащего изображение рака молочной железы, требует вмешательства врача-рентгенолога, т.е. совместной работы ПЭВМ и врача, так как для диагностики данного типа новообразования необходимо учитывать характеристики ткани молочной железы пациентки (жировая инволюция, фиброзно-кистозная болезнь (ФКБ), аденоз). Автором были разработано три алгоритма предварительной обработки маммографических снимков [7, 8, 19]. Они учитывают свойства ткани молочной желе-

зы и позволяют обнаруживать опухоли, находящиеся под мастопатией, которые врач-рентгенолог не способен увидеть непосредственно на рентгеновском снимке.

На следующем этапе улучшенный снимок поступает на вход подсистемы выделения областей новообразований, где изображение подвергается текстурной сегментации, зависящей от предыдущих алгоритмов предварительной обработки. Серия вычислительных экспериментов показала, что для локализации всех трех типов патологий молочных желез «беспроектным» вариантом является использование алгоритма маркерного водораздела [7].

При этом для обнаружения областей новообразований кисты, фиброаденомы и РМЖ наилучшим образом подходят алгоритмы маркерного водораздела [8] и Fuzzy C-mean [19].

Прежде чем выделить контур на изображении, текстурная карта [7] снимка подвергается бинаризации. Алгоритмы бинаризации для каждого вида новообразований свои. Например, выделение области рака молочной железы на фоне аденоза описано в [7].

Выделение контура обнаруженной области осуществляется с помощью алгоритма из [19].

В подсистеме вычисления признаков определяются текстурные признаки каждой выделенной области по матрице смежности [5, 13] с помощью признаков Харалика второго порядка. В данном исследовании для формирования матрицы смежности использовалось расстояние, равное «1» по четырем направлениям (0° , 45° , 90° , 135°). Таким образом, общее количество признаков, характеризующих каждую выделенную область, равно 56.

Далее формируется вектор признаков выделенной области $V[v_1, v_2, \dots, v_{56}]$. На следующем этапе полученный вектор признаков сравнивается с эталонными векторами признаков ($V_3[v_{3,1}, v_{3,2}, \dots, v_{3,56}]$) из базы знаний. Для этой цели вычисляются минимумы средних сумм квадратов отклонений [7]. Тем самым определяется тип новообразования, т.е. осуществляется машинная диагностика вида болезни молочной железы (подсистема классификации новообразований).

Подсистемы работы с БД обеспечивают стандартные операции (добавление, изменение, удаление записей о проведенных исследованиях), формирование статистики по исследованиям.

Обработка и анализ маммограмм с помощью СОиАМ. С помощью маммографа, описанного в [7], были получены маммограммы (около 500 штук) в виде рентгеновских снимков молочной железы на пленке с размером 18×24 см. Подробно характеристики пленки и экранов рентгеновского оборудования указаны в [7, 11]. В итоге было сформировано 2 выборки. В обучающую выборку вошли снимки с подтвержденным диагнозом из базы MIAS [2] (322 снимка). Для контрольной выборки из 500 снимков были отобраны 100 снимков, диагнозы на которых были подтверждены биопсией. Размер оцифрованных снимков составлял 1024×1024 пикселей.

Обучающая выборка под контролем врача-рентгенолога была обработана по рассмотренной выше технологии для обнаружения новообразований. Для каждой области выделенного новообразования был сформирован вектор эталонных признаков и занесен в базу знаний.

Результаты распознавания новообразований для обучающей выборки с помощью комплексной системы, построенной на основе разработанной технологии, и врачом-рентгенологом были объединены в табл. 2.

Результаты экспериментов показали, что врач на фоне ФКБ выявил не более 50 % как доброкачественных (киста, фиброаденома), так и злокачественных новообразований. В то же время СОиАМ правильно выделила более 80 % новообразований.

Таблица 2

Результаты диагностики-распознавания новообразований

Параметры		Количество снимков	Врач-рентгенолог		Система	
Снимки с «нормой»	Жировая инволюция	50	50	100 %	50	100 %
	ФКБ	105	105	100 %	90	85,7 %
	Аденоз	28	28	100 %	22	78,5 %
Снимки с кистами	Жировая инволюция	39	39	100 %	35	89,7 %
	ФКБ	24	11	45,8 %	20	83 %
	Аденоз	6	3	50 %	4	67 %
Снимки с фиброаденомами	Жировая инволюция	7	7	100 %	6	85,7 %
	ФКБ	9	6	66,7 %	8	88,9 %
	Аденоз	6	3	50 %	5	83,3 %
Снимки с РМЖ	Жировая инволюция	23	23	100 %	22	95,6 %
	ФКБ	20	8	40 %	17	85 %
	Аденоз	5	2	40 %	4	80 %

Таким образом, наличие мастопатии любого типа уменьшает вероятность постановки правильного диагноза, так как накладывает тень практически на всю область молочной железы. При этом увеличивается вероятность пропуска рентгенологом новообразований на ранней стадии при скрининге. Указанный недостаток в значительной мере устраняет основанная на предложенной технологии комплексная ЭАС (СОиАМ). Благодаря ее применению при обработке маммограмм, осложненных мастопатией, «убирается» верхний слой ткани и становится видно новообразование. Кроме того, учитывая возможность выделения контуров подозрительных областей, можно достаточно уверенно определить, к какому виду эти новообразования относятся (доброкачественные или злокачественные).

Заключение. Анализ возможностей существующих САМА-систем показал их высокую чувствительность обнаружения новообразований и низкий уровень ложноположительных срабатываний. Однако при сравнении некоторых таких систем в лаборатории Общества радиологов Северной Америки (RSNA) [17] было выявлено большее количество ложных срабатываний, чем указывают разработчики этих систем. Кроме того, не было найдено ни одной САМА системы, разработанной в России.

По результатам анализа существующих САМА-систем были сформулированы основные требования к создаваемой ЭАС по маммограммам.

Структура СОиАМ показывает универсальность разработанной ЭАС, а также отражает основные этапы ее работы.

Предлагаемая СОиАМ обеспечивает комплексное решение задачи автоматизации процесса маммографического скрининга, а именно, ввода информации о пациенте, регистрации маммограмм, предварительной обработки снимков, выделения подозрительных областей, идентификации выделенных новообразований, формирования бланка протокола маммографического исследования.

Результаты экспериментальных исследований показали, что и ЭАС, и врач с высокой вероятностью выделяют все новообразования на фоне жировой инволюции. Однако на фоне ФКБ и аденоза у врача возникают сложности с постановкой диагноза – в этом случае качество результатов у СОиАМ выше.

Список литературы

1. Абраменко Д. Ю. Анализ некоторых направлений использования информационно-телекоммуникационных технологий для повышения качества предоставления медицинских услуг в регионах / Д. Ю. Абраменко, К. И. Квятковский, А. В. Петраев, И. В. Сибикина // Прикаспийский журнал: управление и высокие технологии. – 2013. – № 4 (24). – С. 183–190.
2. Ахметшин А. М. Сегментация низкоконтрастных изображений с применением иерархической модели Марковского случайного поля / А. М. Ахметшин, А. Е. Федоренко // Искусственный интеллект. – 2002. – № 4. – С. 414–421.
3. Брумштейн Ю. М. Системный анализ направлений и особенностей информатизации сферы здравоохранения России / Ю. М. Брумштейн, Е. В. Скляренко, А. С. Мальвина, Ю. Ю. Аксенова, А. Б. Кузьмина // Прикаспийский журнал: управление и высокие технологии. – 2013. – № 4 (24). – С. 73–86.
4. Буланова Ю. А. Использование информационных технологий для локализации области рака молочной железы на маммограммах с преобладанием железистого компонента / Ю. А. Буланова // Прикаспийский журнал: управление и высокие технологии. – 2013. – № 3 (23). – С. 100–111.
5. Гайдель А. В. Исследование текстурных признаков для диагностики заболеваний костной ткани по рентгеновским изображениям / А. В. Гайдель, С. С. Первушкин // Компьютерная оптика. – 2013. – Т. 37, № 1. – С. 113–119.
6. Захаров Д. А. Комплексное применение информационно-коммуникационных технологий в сфере телемедицины / Д. А. Захаров, Ю. М. Брумштейн. – Астрахань : Астраханский государственный университет, Издательский дом «Астраханский университет», 2013 – 133 с.
7. Садыков С. С. Исследование маркерного водораздела для выделения области рака молочной железы / С. С. Садыков, Ю. А. Буланова, Е. А. Захарова, В. С. Яшков // Алгоритмы, методы и системы обработки данных. – 2013. – № 23. – С. 56–64.
8. Садыков С. С. Предварительная обработка маммографических снимков / С. С. Садыков, Ю. А. Буланова, В. С. Яшков // Надежность и качество : тр. междунар. симп. – 2013. – Т. 1. – С. 340–343.
9. Садыков С. С. Технология выделения кисты на маммограмме / С. С. Садыков, Е. А. Захарова, Ю. А. Буланова // Вестник РГРТУ. – 2013. – № 1. – С. 7–11.
10. Щербаков С. На стыке двух эпох: Планшетный сканер Epson Perfection V750 Pro / С. Щербаков // Foto&Video. – 2006. – № 7. – С. 60–65.
11. Computer-Aided Detection for Digital Mammography syngo MammoCAD. Siemens. – 2008. – P. 1–8. – Available at: http://www.medical.siemens.com/siemens/en_INT/gg_sps_FBAs/files/brochures/cad/finalpdfMammoCAD.pdf (accessed 25.11.2013)
12. FUJIFILM Digital Mammography CAD. – Available at: <http://www.fujifilm.eu/eu/products/medical-systems/products/p/fujifilm-digital-mammography-cad/> (accessed 25.11.2013).
13. Haralick R. M. Textural Features for Image Classification / R. M. Haralick, K. Shanmugan, I. Dinstein // IEEE Transactions on systems, man and cybernetics. – 1973. – Vol. SMC-3, № 6. – P. 610–621.
14. iCAD. Mammography CAD. TotalLook MammoAdvantage. – Available at: <http://www.icadmed.com/products/mammography/totallook.htm> (accessed 25.11.2013).
15. iCAD. Mammography CAD. SecondLook Digital. – Available at: <http://www.icadmed.com/products/mammography/secondlookdigital.htm> (accessed 25.11.2013).
16. MedWOWGlobal. Компании медицинского оборудования. Scanis. – Available at: <http://ru.medwowglobal.com/company/scanis/91190> (accessed 25.11.2013).
17. New Technologies to Help Improve Mammography. – Available at: <http://www.imaginis.com/mammography/new-technologies-to-help-improve-mammography> (accessed 17.01.2014).
18. Parascript. Medical Imaging. AccuDetect. – Available at: <http://www.parascript.com/medical-imaging/> (accessed 25.11.2013).
19. Sadykov S. S. Algorithm of localization of breast cancer in the background of mastopathy / S. S. Sadykov, Y. A. Bulanova // 11th International Conference on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-11-2013). – 2013. – Vol. 2. – P. 717–721.

References

1. Abramenko D. Yu., Kvyatkovskiy K. I., Petraev A. V., Sibikina I. V. Analiz nekotorykh napravleniy ispolzovaniya informatsionno-telekommunikatsionnykh tekhnologiy dlya povysheniya kachestva pre-dostavleniya meditsinskikh uslug v regionakh [Analysis of some directions of use of Information and telecommunication technologies to improve the quality of medical services in the regions]. *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Management and High Technologies], 2013, no. 4 (24), pp. 183–190.
2. Akhmetshin A. M., Fedorenko A. Ye. Segmentatsiya nizkokontrastnykh izobrazheniy s primeneniem ierarkhicheskoy modeli Markovskogo sluchaynogo polya [Segmentation of low contrast images using a hierarchical markov random field model]. *Iskusstvennyy intellekt* [Artificial Intelligence], 2002, no. 4, pp. 414–421.
3. Brumshteyn Yu. M., Sklyarenko Ye. V., Malvina A. S., Aksenova Yu. Yu., Kuzmina A. B. Sistemnyy analiz napravleniy i osobennostey informatizatsii sfery zdravookhraneniya Rossii [The system analysis of trends and characteristics of healthcare informatization of Russia]. *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Management and High Technologies], 2013, no. 4 (24), pp. 73–86.
4. Bulanova Yu. A. Ispolzovanie informatsionnykh tekhnologiy dlya lokalizatsii oblasti raka molochnoy zhelezy na mammogrammakh s preobladaniem zhelezistogo komponenta [Using information technologies for localization of breast cancer on mammograms with a predominance of glandular component]. *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Management and High Technologies], 2013, no. 3 (23), pp. 100–111.
5. Gaydel A. V., Pervushkin S. S. Issledovanie teksturnykh priznakov dlya diagnostiki zabolevaniy kostnoy tkani po rentgenovskim izobrazheniyam [Research of textural features for the diagnosis of bone diseases by X-ray images]. *Kompyuternaya optika* [Computer Optics], 2013, no. 1 (37), pp. 113–119.
6. Zakharov D. A., Brumshteyn Yu. M. *Kompleksnoe primeneniye informatsionno-kommunikatsionnykh tekhnologiy v sfere telemeditsiny* [Integrated application of information and telecommunication technologies in the field of telemedicine]. Astrakhan, 2013. 133 p.
7. Sadykov S. S., Bulanova Yu. A., Zakharova Ye. A., Yashkov V. S. Issledovanie markernogo vodorazдела dlya vydeleniya oblasti raka molochnoy zhelezy [Watershed research to highlight the area of breast cancer]. *Algoritmy, metody i sistemy obrabotki dannykh* [Algorithms, Methods and Data Processing Systems], 2013, no. 23, pp. 56–64.
8. Sadykov S. S., Bulanova Yu. A., Yashkov V. S. Predvaritelnaya obrabotka mammograficheskikh snimkov [Preprocessing of mammographic images]. *Nadezhnost i kachestvo: trudy mezhdunarodnogo simpoziuma* [Reliability and Quality: Proceedings of the International Symposium], 2013, vol. 1, pp. 340–343.
9. Sadykov S. S., Zakharova Ye. A., Bulanova Yu. A. Tekhnologiya vydeleniya kisty na mammo-gramme [Technology of Allocation of Cysts on Mammogram]. *Vestnik Ryazanskogo gosudarstvennogo radio-tekhnicheskogo universiteta* [Bulletin of Ryazan State Radioengineering University], 2013, no. 1, pp. 7–11.
10. Shcherbakov S. Na styke dvukh epokh: Planshetnyy skaner Epson Perfection V750 Pro [At the junction of two eras: Flatbed scanner Epson Perfection V750 Pro]. *Foto&Video*, 2006, no. 7, pp. 60–65.
11. Computer-Aided Detection for Digital Mammography syngo MammoCAD. Siemens, 2008, p. 1–8. Available at: http://www.medical.siemens.com/siemens/en_INT/gg_sps_FBAs/files/brochures/cad/finalpdfMammoCAD.pdf (accessed 25 November 2013).
12. FUJIFILM Digital Mammography CAD Available at: <http://www.fujifilm.eu/eu/products/medical-systems/products/p/fujifilm-digital-mammography-cad/> (accessed 25 November 2013).
13. Haralick R. M., Shanmugan K., Dinstein I. Textural Features for Image Classification. *IEEE Transactions on systems, man and cybernetics*, 1973, vol. SMC-3, no. 6, pp. 610–621.
14. iCAD. Mammography CAD. TotalLook MammoAdvantage. Available at: <http://www.icadmed.com/products/mammography/totallook.htm> (accessed 25 November 2013).
15. iCAD. Mammography CAD. SecondLook Digital. Available at: <http://www.icadmed.com/products/mammography/secondlookdigital.htm> (accessed 25 November 2013).
16. MedWOWGlobal. Kompanii meditsinskogo oborudovaniya. Scanis. Available at: <http://ru.medwowglobal.com/company/scanis/91190> (accessed 25 November 2013).
17. New Technologies to Help Improve Mammography. Available at: <http://www.imaginis.com/mammography/new-technologies-to-help-improve-mammography> (accessed 17 January 2014).

18. Parascript. Medical Imaging. AccuDetect. Available at: <http://www.parascript.com/medical-imaging/> (accessed 25 November 2013).

19. Sadykov S. S., Bulanov Y. A. Algorithm of localization of breast cancer in the background of mastopathy. *11th International Conference on Pattern Recognition and Image Analysis: New Information Technologies (PRLA-11-2013)*, 2013, vol. 2, pp. 717–721.

УДК 004.93'12

МЕТОД КОРРЕКЦИИ ОШИБОК КЛАССИФИКАЦИИ РАСПОЗНАННЫХ СИМВОЛОВ

Статья поступила в редакцию 30.01.2014, в окончательном варианте 16.02.2014.

Брейман Александр Давидович, кандидат технических наук, доцент, Национальный исследовательский университет «Высшая школа экономики», 101000, Российская Федерация, г. Москва, ул. Мясницкая, 20, e-mail: abreyman@hse.ru

Яковлев Илья Александрович, аспирант, Московский государственный университет приборостроения и информатики, 107996, Российская Федерация, г. Москва, ул. Стромынка, 20, e-mail: krofes@gmail.com

Процесс распознавания текстовых документов неизбежно связан с возникновением ошибок распознавания, для выявления и исправления которых используют методы пост-обработки, как правило, опирающиеся на словарный поиск. Использование словарей позволяет достичь приемлемого качества распознавания для латиницы, кириллицы и других фонетических алфавитов, однако мало-пригодно для языков, в которых выделение отдельных слов в письме нехарактерно или обязательно (китайский, японский, корейский, вьетнамский и прочие языки). В статье рассмотрены существующие методы, направленные на решение данной проблемы, а также описан новый подход к исправлению некоторых видов ошибок, основанный на применении ансамблей нейронных сетей (по нейронной сети на каждый возможный символ), позволяющий сократить количество ошибок в результате распознавания иероглифического письма, а для фонетических алфавитов – снизить зависимость от качества словарей.

Ключевые слова: оптическое распознавание символов, ошибки распознавания символов, пост-обработка ошибок распознавания, система верификации распознавания, система коррекции ошибок распознавания без словаря, распознавание иероглифов, нейронные сети, нейросетевые ансамбли

OPTICAL CHARACTER RECOGNITION ERRORS CORRECTION METHOD

Breyman Aleksandr D., Ph.D. (Engineering), Associate Professor, National Research University “Higher School of Economics”, 20 Myasnitskaya St., Moscow 101000, Russian Federation, e-mail: abreyman@hse.ru

Yakovlev Ilya A., post-graduate student, Moscow State University of Instrument Engineering and Computer Science, 20 Stromynka St., Moscow, 107996, Russian Federation, e-mail: krofes@gmail.com

Optical recognition of text documents is inevitably error-prone process. To identify and correct that errors systems use post-processing techniques that are usually based on dictionary search. Using dictionaries can bring an acceptable quality of recognition for Latin, Cyrillic and other phonetic alphabets, but of little use for the languages in which the selection of individual words is untypical or optional (Chinese, Japanese, Korean, Vietnamese and other languages). This paper discusses known methods to address this problem, and proposes a new approach to correcting certain types of errors, based on the application of neural networks ensembles (containing distinct neural network for each possible character), which allows to reduce the num-