

Построение базы знаний для интерактивного распознавания в системах компьютерной микроскопии

Е. Ю. БЕРДНИКОВИЧ, Е. С. ЛЕБЕДЕВА, В. Г. НИКИТАЕВ, К. С. ЧИСТОВ

Национальный исследовательский ядерный университет «МИФИ»,
Москва, Россия, e-mail: kaf46@mephi.ru

Рассмотрены информационно-измерительные системы компьютерной микроскопии для морфологической диагностики, включающие экспертные системы, в основу которых положен метод интерактивного распознавания с применением качественных и количественных информативных признаков опухолей. Формализованные списки признаков сформированы на основе экспертных оценок врачей-цитологов и гистологов и составляют, наряду с цветными изображениями опухолей, ядро базы знаний экспертных систем по различным локализациям. Представлен принцип организации экспертной системы, приведена структура базы знаний, описаны модель интерактивного распознавания, а также логическая структура базы данных, реализованная в СУБД Oracle Database 11g.

Ключевые слова: экспертная система, экспертная оценка, база знаний, логическая модель, модель интерактивного распознавания, база данных, цитологическая диагностика, морфологическое описание, поддержка принятия решений.

The data-measuring systems of computer microscopy for morphological diagnostics include the expert systems based on the method of interactive recognition with application of qualitative and quantitative informative signs of tumors. The formalized lists of signs are formed on the basis of expert judgements of doctors — cytologists and histologists and compose along with color images of tumors the nucleus of the knowledge base of expert systems for different localizations. The principle of the organization of expert system is represented, the structure of the knowledge base is given, the model of interactive recognition as well as the logical structure of the data base realized in Oracle Database 11g is described.

Key words: expert system, expert judgement, knowledge base, data model, logical model, interactive identification model, database, cytological diagnostics, morphological description, decision support.

В современной клинической практике в комплексе диагностических мероприятий цитологическое и гистологическое исследование с морфологической верификацией поражения — необходимое условие своевременного начала и адекватности проводимого лечения онкологических заболеваний [1]. На сегодняшний день существуют следующие проблемы

в обозначенной области: дефицит высококвалифицированных кадров; проблема доступа к знаниям в удаленных от центральных клиник регионах; высокая степень субъективизма при принятии диагностических решений; наличие пограничных поражений; редко встречающиеся нозологические формы [2]. Для решения этих проблем специалистами кафедры компьютерных медицинских систем Национального исследовательского ядерного университета «МИФИ» и врачами-экспертами кафедры клинической лабораторной диагностики Российской медицинской академии последипломного образования и Российского онкологического научного центра им. Н. Н. Блохина РАМН разработаны информационно-измерительные системы (ИИС) компьютерной микроскопии для гисто- и цитологической диагностики по различным локализациям, которые, как показала клиническая практика, являются эффективным средством поддержки принятия решений врачом [2].

Основным интеллектуальным компонентом таких систем являются экспертные системы (ЭС), в основе которых лежит метод интерактивного распознавания с применением качественных и количественных информативных признаков опухолей. Структура ЭС представлена на рис. 1 на примере цитологической диагностики опухолей желудка. Ядром ЭС является база знаний (БЗ), проблема построения которой имеет важнейшее значение при создании ИИС данного класса.

Цель работы — построение информативной БЗ для интерактивного распознавания в ИИС компьютерной микроскопии.



Рис. 1. Структура экспертной системы

Модель базы знаний. База знаний ИИС имеет следующую структуру: для всех локализаций (рис. 2) существует информация по классам патологий (неопухоловое, предопухоловое, опухолевое заболевание (доброкачественное, злокачественное)) и нозологическим формам. Каждая нозологическая форма представлена выборкой цифровых изображений (5—20 шт.), характеризующих различные случаи заболеваний. Для всех изображений помимо класса патологии и цитологического описания указаны набор информативных морфологических признаков и общие характеристики препарата — способ окрашивания, увеличение микроскопа и др.

Общие перечни доступных значений для описания изображений микропрепаратов — морфологические признаки, нозологические формы и другие описания — располагаются в соответствующих отдельных модулях, называемых словарями данных. Для выделения формализованных списков информативных признаков опухолевых и неопухоловых заболеваний построены модели систематизации предметной области на основе обобщения экспертных оценок, сделанных ведущими цитологами страны [3, 4]. Также БЗ содержит раздел со справочными материалами, включающими терминологический словарь, список литературы и обзор интернет-ресурсов с наиболее полезной и актуальной информацией по цитологической диагностике. Помимо этого в БЗ содержится не только фактическая информация, но и правила вывода, позволяющие проводить вероятностную оценку принадлежности исследуемого объекта к одному из заданных классов.

Модель интерактивного распознавания изображений микропрепаратов на основе экспертных оценок при микроскопическом анализе состоит из следующих этапов.

1. Определение списка классов (списка нозологических форм), между которыми будет осуществляться вероятностная оценка принадлежности объекта X , $\{K_j\}$.

2. Формирование выборки эталонных объектов (цифровых изображений микропрепаратов) из хранимых в БЗ экспертных оценок, представляющих заданные классы $\{M_{эkj}\}$.

3. Построение словаря с информативными морфологическими признаками. Ему присуща двухступенчатая иерархическая структура (морфологический признак/возможные значения), которая представляет собой вектор свойств \mathbf{S} и вектор признаков \mathbf{P} , причем каждое свойство описывается своим вектором признаков \mathbf{P}_s . Таким образом, каждому входному объекту X ставится в соответствие набор $(\mathbf{S}_x, \{\mathbf{P}_s\}_x)$

из указанного словаря, т. е. $X \rightarrow (\mathbf{S}_x, \{\mathbf{P}_s\}_x)$.

4. Создание морфологического описания объектов из эталонной выборки классов с помощью словаря морфоло-

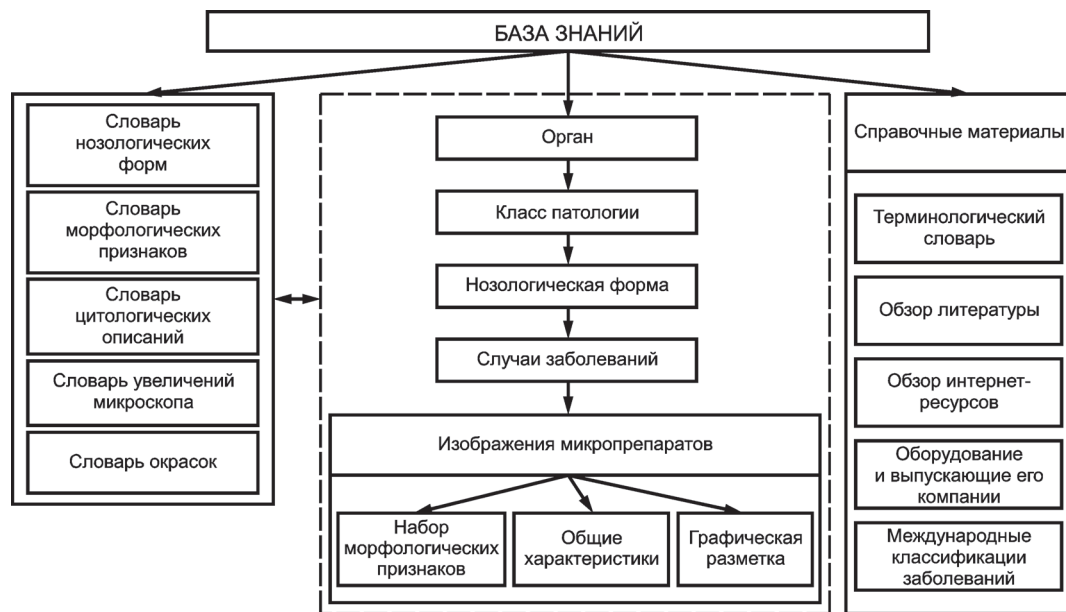


Рис. 2. Модель базы знаний

гических признаков $\{M_{эkj}\}$, т. е. каждому объекту из класса ставится в соответствие набор свойств и признаков из словаря $M_s \rightarrow (\mathbf{S}_s, \{\mathbf{P}_s\}_s)$.

5. При анализе исследуемого объекта лицо, принимающее решение, определяет из указанного словаря набор свойств и признаков, характерных для исследуемого объекта X .

6. В системе происходит анализ введенной комбинации признаков входного объекта X и оценивается наличие выбранного набора признаков в совокупности признаков для каждого класса эталонной выборки. Таким образом, определяется перечень наиболее вероятных классов для входного объекта X . Достоверность работы системы зависит от объема и качества хранимых в БЗ данных (экспертных оценок). При малом количестве случаев заболеваний по определенным нозологическим формам данные являются недостоверными и недостаточными для проведения анализа. Для набора достаточного количества данных необходимы годы работы врачей-цитологов с системой. Из-за крайней субъективности самого цитологического метода экспертные оценки должны проводиться несколькими экспертами для обеспечения точности морфологического описания микроизображений и клинически подтвержденного диагноза.

Логическая модель базы данных. Ядром БЗ является база данных, которая содержит 16 сущностей с установленными отношениями между ними (рис. 3). Каждая сущность включает от 2 до 12 атрибутов, проведена нормализация, обеспечена достоверность, непротиворечивость и целостность данных. Центральными сущностями являются сущность «пациент» и сущность «изображения», имеющие отношение «один-ко-многим». В отдельные сущности выделены данные, которые содержат вполне определенный фиксированный набор значений: «цитологические диагнозы», «гистологические диагнозы», «цитологические описания», «гистологические описания», «морфологические признаки», «окраски», «увеличения». База данных реализована в СУБД Oracle Database 11g.

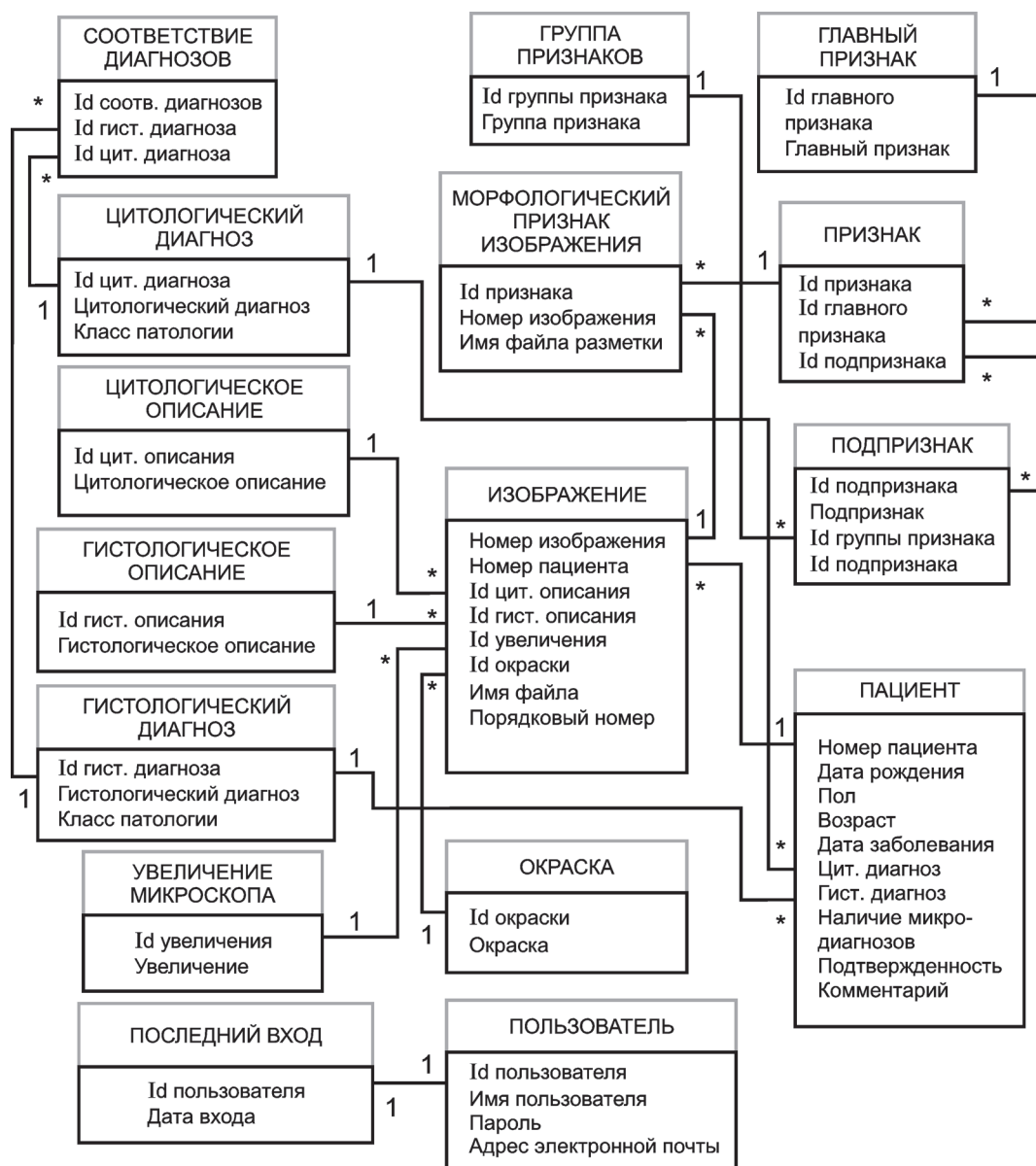


Рис. 3. Логическая модель базы данных

Заключение. Информационно-измерительные системы компьютерной микроскопии для морфологической диагностики являются эффективным инструментом в руках врача для повышения достоверности принимаемых ответственных врачебных решений в сложных случаях заболеваний. Интеллектуальным ядром таких комплексов являются ЭС, в основу которых положен метод интерактивного распознавания с применением качественных и количественных информативных признаков опухолей, которые наряду с изображениями и правилами принятия решений составляют БЗ по различным локализациям. Формализованные списки признаков сформированы на основе экспертных оценок врачей-цитологов и гистологов.

Интерактивное распознавание изображений микропрепаратов позволяет проводить вероятностную оценку принадлежности исследуемого объекта к той или иной нозологической форме при вводе из словаря перечня морфологических признаков, характерных для исследуемого объекта. Во время ввода изображения и его морфологического опи-

сания происходит выдвижение гипотезы, которая помогает врачу в постановке диагноза. При этом, когда создаются системы, составляется комплекс наиболее информативных морфологических признаков, присущих заболеваниям различных локализаций.

Представлен принцип организации ЭС, приведена структура БЗ, описаны модель интерактивного распознавания, а также логическая структура базы данных. В статье все структуры приведены на примере цитологической диагностики желудка.

Для реализации БЗ спроектирована и физически реализована база данных, которая включает 16 сущностей с установленными отношениями между ними. При этом в систему по опухолям желудка внесена информация о более чем 170 пациентах и около 2000 изображений цитологических микропрепаратов. Среда разработки: СУБД Oracle Database 11g+ Microsoft Visual Studio 2010 (с применением языка C#) + Microsoft Expression Studio 4 (с применением технологии WPF).

Работа по созданию ИИС по цитологической диагностике опухолей желудка проводилась кафедрой компьютерных медицинских систем Национального исследовательского ядерного университета «МИФИ» совместно с врачами-экспертами кафедры клинической лабораторной диагностики Российской медицинской академии последипломного образования. Система успешно внедрена в этом центре.

Литература

1. Шабалова И. П., Полонская Н. Ю. Основы клинической цитологической диагностики. М.: ГЭОТАР-Медиа, 2009.
2. Никитяев В. Г., Бердникович Е. Ю. Системные принципы построения компьютерных комплексов гистологической диагностики // Инженерная физика. 2004. № 4. С. 58—68.
3. Михайлов В. Н. и др. Статистические оценки в системах поддержки принятия решений // Научная сессия МИФИ-2003: Сборник научных трудов. В 14 т. Т. 1. М.: МИФИ, 2003. С. 238.
4. Касоян Т. Н. и др. IMAGE-система в цитологической диагностике заболеваний щитовидной железы // Клиническая лабораторная диагностика. 2003. № 9. С. 30.

Дата принятия 06.04.2012 г.