

ПРИНЦИПЫ РАЗРАБОТКИ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ

В.В. Битюкова¹, Е.А. Сидоренко², Б.Л. Гуртовой³

Воронежская государственная медицинская академия (1);

Воронежская государственная технологическая академия (2);

Научный центр акушерства, гинекологии и перинатологии РАМН, Москва (3)

Представлена членом редколлегии профессором В.И. Коноваловым

Ключевые слова и фразы: автоматизированная диагностика; вероятность правильной диагностики; интегральный и дифференциальный подход; кластерный анализ; пороговые значения.

Аннотация: Рассмотрены существующие методы автоматизированной медицинской диагностики. Разработан подход, позволяющий выделить симптомы, наиболее значимые для диагностики заболевания. Предложены интегральный и дифференциальный способы диагностики по качественным и количественным признакам на основании малой статистики.

Диагностика гнойно-воспалительных заболеваний в акушерстве в максимально короткие сроки чрезвычайно важна, так как позволяет начать раннее лечение, уменьшить количество осложнений и сроки пребывания больной в стационаре. Она весьма затруднительна при стертом течении заболевания. Для постановки диагноза в таких случаях нередко необходимо проведение дополнительных методов исследования, имеющих достаточно высокую стоимость.

Введение в медицинскую практику нового оборудования, способного определить большое число параметров функционирования организма, резко увеличило количество поступающей к врачу информации, но не облегчило процесс постановки диагноза. Трудность анализа поступающей информации по традиционным методикам вынуждает искать новые способы ее обработки.

Привлечение информационных технологий в практику работы врача привело к появлению широкого спектра автоматизированных методов медицинской диагностики, позволяющих принимать решение о наличии или отсутствии заболевания в диалоговом режиме с ЭВМ в реальном масштабе времени. Очевидно, что это одно из магистральных направлений развития медицинской диагностики.

Однако большинство методов диагностики [4] было создано для технических объектов и оказалось непригодно в медицине, так как, во-первых, у медиков и математиков различные подходы к описанию состояния диагностируемого объекта (технического комплекса или организма человека), во-вторых, параметры нормального состояния человека обладают большим полиморфизмом, в отличие от технических объектов, и, в-третьих, связи между элементами технического объекта хорошо известны, их относительно немного в отличие от взаимосвязей между органами, системами органов и т. п. в организме человека.

Проблема разработки методов автоматизированной медицинской диагностики, не требующих привлечения дорогостоящих вычислительных систем и понят-

ных врачу, а потому принимаемым им в практику работы, и затрачивающих на обработку поступающей информации малое время, актуальна. Отметим, что системы, созданные на базе таких методов, призваны оказать поддержку при постановке диагноза, выдавая несколько вариантов решения в максимально короткие сроки, анализируя взаимосвязь множества параметров функционирования организма. Окончательный диагноз по-прежнему устанавливается врачом.

Целью данной работы является анализ существующих методов автоматизированной диагностики и разработка подхода, позволяющего выделить наиболее значимые для диагностики заболевания симптомы и необходимые для их оценки методы исследования, а также создать систему автоматизированной медицинской диагностики на основании лечебного опыта, выдающую рекомендации для врача, обладающего меньшей квалификацией.

Поскольку организм человека состоит из множества органов, подсистем и систем, то его состояние в медицине описывается с помощью внешних проявлений (симптомов болезни), имеющих большой разброс значений даже в пределах одного заболевания и часто коррелированных. Поэтому для целей автоматизированной медицинской диагностики пригодны математические методы, основанные на анализе признаков (симптомов), выбирающие одно из состояний объекта – диагноз из списка заранее определенных.

Первые системы автоматизированной медицинской диагностики были созданы на основе параметрических методов, наиболее известным из которых является байесовский [2]. Система диагностики приобретенных пороков сердца [5] с помощью решающего правила, содержащего 5 последовательных алгоритмов на основании выборки из 900 верифицированных во время операции диагнозов, позволила получить в специализированной клинике средний для всех пороков результат 95,8 %.

При построении таких систем предполагалось, что все диагностируемые заболевания характеризуются различными признаками, которые, в свою очередь, не коррелируют между собой. Однако, такое предположение весьма редко выполняется, так как существуют общие признаки, такие как изменение температуры, сопровождающие целый ряд заболеваний. Даже если исключить общие признаки из рассмотрения при постановке диагноза, оставшиеся специфические могут проявляться при нескольких заболеваниях, варьируя свои значения. Степень корреляции между признаками весьма тяжело оценить, но отрицать, что она существует, невозможно.

Байесовский метод позволяет получить вероятности наличия заболеваний из ряда заранее выбранных, на основе оценки частоты проявления каждого признака при выбранном заболевании, так называемой априорной вероятности. Сделать такую оценку можно, только собрав большую (порядка 700-1000 историй болезни) статистику. Сбор большого числа историй болезней требует длительного времени, поскольку многие формы заболеваний являются крайне редкими, и в условиях одного стационара могут встречаться раз в несколько лет.

Системы, основанные на байесовском подходе, зачастую невозможно использовать в медицинской диагностике, так как статистика заболеваний невелика, симптомы их перекрываются и коррелируют между собой, поэтому добиться высоких результатов можно лишь в специализированных клиниках.

Другим подходом, разработанным позднее, является диагностика на основе моделирования процесса принятия решения опытным врачом, входящим наряду с данными о заболеваниях в экспертную систему. Экспертные системы показывают весьма высокие результаты диагностики. Например, система распознавания ранних стадий артериальной гипертонии с привлечением теории нечетких множеств на основании 35 иммунобиохимических параметров по 57 случаям заболеваний [3], верные решения были получены в 73...94 % случаев.

Однако создание такой системы требует привлечения специалистов из разных областей – программистов-разработчиков, инженеров, формализующих знания врачей, и нескольких врачей-экспертов, что сопряжено с большими временными и материальными затратами на разработку.

В большинстве практических задач медицинской диагностики мы имеем дело с ограниченной статистикой заболевания, неодинаковой для различных форм болезни, так как обычно фиксируются тяжелые случаи, в то время как средние и особенно легкие – часто пропускаются, пока не переходят в тяжелую форму (тяжесть стертых форм часто недооценивается в связи с «размытостью клинической картины»). В медицинской практике это приводит к инвалидизации пациентов, а следовательно, к увеличению стоимости лечения и длительности восстановительного периода.

Разрабатываемая система автоматизированной медицинской диагностики, должна давать высокие результаты (около 90 % случаев правильной диагностики) с использованием небольшой статистики и быть относительно недорогой. Кроме того, она должна оценивать необходимость выполнения дополнительных методов исследования и отбирать только те из них, что дают максимальное количество информации о состоянии пациента.

Для построения такой системы необходимо разработать эвристические методы, которые позволяют проанализировать совокупность симптомов заболевания и выработать такие математические правила, которые бы позволили выделить данное заболевание среди других, то есть провести границы (пороги) между заболеваниями.

Симптомы заболевания могут быть как количественными (измеряются с помощью аппаратуры и выражаются в числовом значении, например, СОЭ, формула крови), так и качественными (характеризуют некоторое состояние и не могут быть измерены: запах, патологические включения и т.п.). Математический анализ их совокупности возможен лишь в случае, когда они имеют одинаковую размерность, либо все симптомы безразмерны. Поскольку привести все значения симптомов к одинаковой размерности не представляется возможным, то наиболее простым вариантом является переход к их безразмерным значениям.

Для диагностики заболевания каждый симптом должен иметь значение из конкретного диапазона, например, температура тела для послеродового эндометрита изменяется от 37,5 до 39 °С. Поэтому не столько важно конкретное значение симптома, сколько его попадание в указанный диапазон. Разбив все признаки заболевания на диапазоны значений, соответствующих различным нозологиям, мы получим для всех симптомов одинаковую характеристику – число диапазонов (число градаций симптома), на основании которой можно вести математический анализ клинической картины.

Поскольку заболевание характеризуется как общими, так и специфическими симптомами, то каждый из них несет меньшую (в первом случае) или большую (во втором) информацию. Совокупные данные, получаемые при анализе всех симптомов, для того или иного заболевания будут различны. Следовательно, можно выбрать граничные значения информации о каждой нозологической форме и ставить диагноз на основании сравнения полученного количества информации о состоянии больного с граничным значением информации о заболевании.

Оценить необходимость включения в рассмотрение нового симптома можно, рассчитав общее количество информации для нового симптомокомплекса. Если оно увеличилось по сравнению с предыдущим, то симптом включается, если уменьшилось – исключается из рассмотрения.

В рамках нахождения совокупного значения критериев заболевания существует два подхода: интегральный и дифференциальный [1]. Интегральный – основан на суммировании информации, получаемой от каждого симптома. Суммарная

информация сравнивается затем с пороговым значением, на основании чего делается вывод о наличии заболевания. В рамках интегрального подхода можно оценить вероятность правильной диагностики, что повышает достоверность этого метода. Однако существенным недостатком этого подхода является большое число неправильных диагнозов при близких формах заболевания. Это обусловлено тем, что в суммарном значении информации растворяются значения отдельных симптомов, важных для диагностики близких форм. Поэтому интегральный подход лучше всего использовать при диагностике двух хорошо различающихся случаев, например, норма и заболевание или легкая и тяжелая степень.

Указанный недостаток ликвидирует дифференциальный подход, основанный на оценке степени «похожести» диагностируемого случая с рядом уже имеющихся, образующих кластер. Построив на первом этапе кластеры для диагностируемых заболеваний по обучающей выборке, на втором проводится анализ состояния конкретной больной. Ставится диагноз из того кластера, к которому ближе всего оказался рассматриваемый случай. Кластерный анализ позволяет выполнить «тонкую» оценку заболевания, продиагностировать степень тяжести. Однако кластерный анализ дает хорошие результаты только в случае одинаковых кластеров по всем рассматриваемым диагнозам. Это значит, что число историй болезней, заложенных в обучающую выборку для каждого кластера, должно быть одинаковым, что в реальных условиях невозможно.

Поэтому для надежной диагностики заболевания необходимо сочетание интегрального и дифференциального способов диагностики. Вначале проводится «грубая» диагностика с использованием интегрального подхода, а затем более «тонкая» дифференциальная диагностика для предварительно отобранных случаев.

Таким образом, построение системы автоматизированной диагностики проводится в два этапа. Первый – этап обучения системы – проводится с участием врача-эксперта, на котором:

- 1) определяются заболевания или формы одного заболевания, для диагностики которых создается система;
- 2) собирается статистика заболевания для заранее поставленных врачом-экспертом диагнозов;
- 3) отбираются наиболее значимые с точки зрения эксперта симптомы заболевания, закладываемые в обучающую выборку, и определяются их весовые коэффициенты, отражающие степень значимости;
- 4) определяются математические правила, по которым проводится постановка диагноза;
- 5) проводится автоматизированная диагностика случаев, заложенных в обучающую выборку для некоторой комбинации результатов анализов (обычно вначале используют данные тех методов исследования, которые быстрее всего выполняются или наиболее дешевы);
- 6) выбираются пороговые значения полученных с помощью математических правил величин, разделяющие результаты автоматизированной диагностики на заранее определенные экспертом группы диагнозов таким образом, чтобы число расхождений с диагнозами эксперта было минимальным;
- 7) составляется новая комбинация данных анализов, и вновь проводится автоматизированная диагностика обучающей выборки, подбираются новые пороговые значения до тех пор, пока не будет составлена комбинация результатов, и выбраны пороговые значения величин, дающие минимальное число расхождений с экспертом.

После этого созданная система готова к функционированию на втором этапе, который называется консультационным, поскольку позволяет в процессе эксплуатации системы врачом получить альтернативы диагноза. Он включает в себя:

- 1) сбор и ввод в систему только тех данных, которые были отобраны в конечном итоге на этапе обучения;

2) автоматический расчет по математическим правилам и сравнение полученных результатов с максимально допустимыми пороговыми значениями величин;

3) выдачу примерного диагноза.

Руководствуясь полученными результатами, врач ставит диагноз. Такая автоматизированная система диагностики, благодаря двухуровневой постановке диагноза вначале интегральным, а затем дифференциальным способами, очевидно должна давать результаты, близкие к экспертным системам, но, в отличие от последних, быть простой и дешевой в построении.

На основе вышесказанного была синтезирована автоматизированная система диагностики послеродового эндометрита. На первом этапе были определены границы диагностики трех форм заболевания (легкой, средней и тяжелой), и сформирован симптомокомплекс, отражающий максимальную информацию о состоянии больной, включающий температуру тела, качественную характеристику лохий, размеры матки (по данным эхографии) и их изменения, особенности кесарева сечения (плановое или экстренное), кровопотерю, число лейкоцитов, бактериологическое исследование, наличие патологических включений, качественная характеристика полости (форма), состояние послеоперационного шва, гиперемии, толщину послеоперационного шва.

Рабочим этапом, на котором выполняется диагностика заболевания, является второй. Система предлагает врачу ввести значения по каждому симптому из комплекса, приведенного выше, а затем производит расчет с помощью интегрального, дифференциального методов и двухуровневое распознавание. Полученные три результата осмысливаются врачом, и выносятся окончательный диагноз. Эффективность диагностики такой системы достигает 91 % (вероятность правильной диагностики до 0,965) в случае, если известны все значения, составляющие симптомокомплекс. Если использовались не все методы исследования, то вероятность правильной классификации снижается. Однако, если имеются значения специфических признаков эндометрита, диагностика возможна с вероятностью до 0,89.

Пример 1. В истории болезни отсутствовали результаты некоторых лабораторных анализов и УЗИ. Врачом была поставлена тяжелая форма эндометрита. Диагностика проводилась по следующим признакам: температуре тела, качественной характеристике лохий, особенностям кесарева сечения (плановое или экстренное), кровопотере, числу лейкоцитов при бактериоскопии, бактериологическому исследованию, состоянию полости (при гистероскопии), наличию патологических включений, состоянию послеоперационного шва (налет фибрина, провисание лигатур), гиперемии, изменению длины полости при зондировании.

Вероятность правильной диагностики с помощью интегрального метода составила 0,890. Диагноз – тяжелая степень эндометрита, поставленная уже после проведения первого этапа – интегральной диагностики.

Пример 2. Врачом была определена легкая степень эндометрита. Автоматизированная диагностика проводилась по небольшому количеству симптомов (особенности кесарева сечения, качественные характеристики лохий, состояние послеоперационного шва, наличие патологических включений, изменение длины полости при зондировании, кровопотеря, число лейкоцитов при бактериоскопии).

В результате диагностики интегральным методом была получена вероятность правильной диагностики 0,848. Так как вероятность правильной диагностики не превысила 0,85, был проведен расчет по второй ступени с помощью дифференциального метода. Окончательный диагноз, выданный системой – легкая степень эндометрита, совпал с врачебным.

Таким образом, синтезированная система решающих правил позволяет получать высокие результаты автоматизированной диагностики состояния системы, даже если имеются не все значения наиболее информативных признаков состояния системы.

Список литературы

1. Базарский О.В. Автоматизированная диагностика гнойно-воспалительных заболеваний / О.В. Базарский, В.К. Битюков, В.В. Битюкова, Б.Л. Гуртовой, Е.А. Сидоренко. Воронеж: Изд-во ВГТА, 2001.
2. Беллман Р. Математические методы в медицине. – М.: Мир, 1987. – 200 с.
3. Богомолов А.В. Использование лингвистических переменных и методов обработки экспертной информации для автоматизированного распознавания ранних стадий нарушения функционального состояния человека // Информационные технологии, 2000. – № 8. – С. 50-54.
4. Подвальный Е.С. Модели индивидуального прогнозирования и классификации состояний в системах компьютерного мониторинга. – Воронеж: Изд-во ВГТУ, 1998.
5. Биологическая и медицинская кибернетика. Справочник. – Киев: Наукова думка, 1986. – 376 с.

Principles of Development of Automated System of Disease Diagnostics

V.V. Bityukova¹, E.A. Sidorenko², B.L. Gurtovoy³

Voronezh State Medical Academy (1);

Voronezh State Technological Academy (2);

Scientific Center of Obstetrics, Gynecology and Perinatology, RAMS, Moscow (3)

Key words and phrases: automated diagnostics; probability of correct diagnostics; integral and differential approach; cluster analysis; threshold values.

Abstract: Existing methods of automated medical diagnostics are considered. The approach allowing to determine symptoms the most relevant for disease diagnostics is developed. Integral and differential ways of diagnostics by quality and quantity characteristics on the basis of minor statistics are suggested.

Prinzipien der Erarbeitung des automatisierten Systems der Diagnostik der Erkrankungen

Zusammenfassung: Es sind die existierenden Methoden der automatisierten medizinischen Diagnostik untersucht. Es ist das Herangehen, das die für die Diagnostik der Erkrankungen am meisten bedeuteten Symptome zu wählen zuläßt, entwickelt. Es sind die Integral- und Differentialmethoden der Diagnostik nach den qualitativen und quantitativen Merkmalen aufgrund der kleinen Statistik angeboten.

Principes de la conception automatisée du système du diagnostic des maladies

Résumé: On a examiné les méthodes existant pour le diagnostic médical automatisé. On a élaboré la méthode qui permet de dégager les symptômes qui sont les plus importants pour le diagnostic des maladies. On a proposé les moyens intégral et différentiel du diagnostic d'après les indices qualitatives et quantitatives à la base de la petite statistique.