

ВЕРОЯТНОСТНАЯ ДИАГНОСТИКА НАРУШЕНИЙ ГЕМОПОЭЗА

Салтанович И.М.

Научный руководитель - канд. физ.-мат. наук Онищук С.А.

Кубанский государственный университет

Анализ медико-биологических данных, создание численных алгоритмов оценки параметров состояния биосистем представляют собой активно развивающиеся направление применения информационных технологий в биомедицинских исследованиях. Эффективность использования многомерных данных (в частности показателей крови) существенным образом определяется возможностями их согласования, т.е. разработки концепции их совместного использования для получения более надежных оценок состояния биосистем. В условиях недостатка априорной информации о свойствах биосистем наиболее адекватными являются оценки, полученные методами статистического моделирования.

Конечно, при использовании вероятностной диагностики диагноз может быть установлен и не со стопроцентной вероятностью, но и при постановке диагноза врачом не исключена возможность ошибки. Поэтому, во-первых, диагноз практически всегда является вероятностным, а во-вторых, с увеличением числа показателей диагноз все с большей вероятностью будет склоняться в сторону одного определенного заболевания.

Целью данной работы является разработка и исследование метода автоматизированной диагностики патологических состояний кроветворной системы человека на примере выбранных заболеваний относительно контрольной группы.

Для оценки эффективности и информативности исследуемых гематологических показателей был проведен *ROC*-анализ, который дает представление об интегральной диагностической эффективности, позволяя отобрать наиболее значимые данные и отсеять малоинформативные. Исследовалась диагностическая эффективность показателей крови по выбранным заболеваниям относительно контрольной группы. Было установлено, что эритроцитарные показатели различаются статистически значимо у лиц с абсолютным и перераспределительным вариантами дефицита железа. У лиц с опухолевыми заболеваниями крови наиболее значимо различаются показатели лейкоцитарной группы.

Своеобразным методом сравнения *ROC*-кривых является оценка площади под кривыми. Теоретически она изменяется от 0 до 1.0, но, поскольку метод всегда характеризуется кривой, расположенной выше положительной диагонали, то обычно говорят об изменениях от 0.5 ("бесполезный" классификатор) до 1.0 ("идеальный" метод). Численный показатель площади под кривой *AUC* (*Area Under Curve*) составил в среднем 0.81, что соответствует по экспертной шкале качества алгоритма показателю "Очень хорошее". На рисунке 1 представлены *ROC*-кривые для таких показателей как *WBC* (абсолютное содержание лейкоцитов), *RBC* (абсолютное содержание эритроцитов), *MCV* (средний объем эритроцитов), *MCH* (средняя концентрация гемоглобина в эритроците), *HGB* (концентрация гемоглобина), *HCT* (гематокрит). Метод с высокой чувствительностью часто дает истинный результат при наличии положительного исхода (обнаруживает положительные примеры). Наоборот, модель с высокой специфичностью чаще дает истинный результат при наличии отрицательного исхода (обнаруживает отрицательные примеры). В данном случае гематологические показатели различаются статистически значимо, т.е. высока диагностическая

чувствительность, что проявляется в гипердиагностике – максимальном предотвращении пропуска больных. Так же были использованы такие тромбоцитарные показатели как *PLT* (абсолютное содержание тромбоцитов) и *MPV* (средний объем тромбоцитов). Однако для них значение *AUC* в среднем равнялось 0.6.

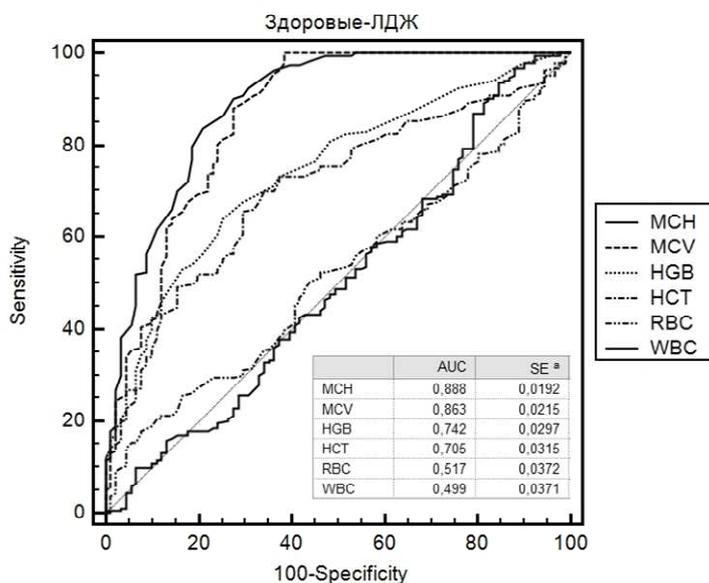


Рис. 1. Чувствительность и специфичность показателей крови для диагностики ЛДЖ относительно контрольной группы.

Анализ диагностической эффективности показателей крови латентного дефицита железа показал, что достоверно выявить заболевание способны только *MCH*, *MCV*, *HGB* и *HCT* (рис 9.), что вполне закономерно ввиду того, что дефицит железа в организме проявляется в изменение параметров эритроцитов (средний объем, концентрация гемоглобина и др.). Абсолютное содержание лейкоцитов и эритроцитов не является диагностически эффективными для данного заболевания. Однако при диагностике хронического миело-лейкоза лейкоциты имеют одну из наибольших значимостей (рис 2.).

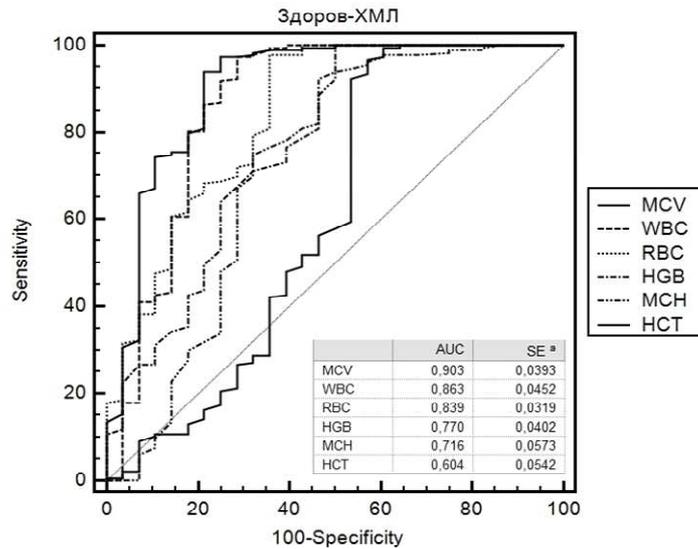


Рис. 2 Чувствительность и специфичность показателей крови для диагностики ХМЛ относительно контрольной группы.

Для исследуемых групп по всем используемым показателям крови были построены гистограммы плотности вероятности и аппроксимированы гладкой математической функцией. На рисунке 3 показан общий график для *HGB* для 7 групп.

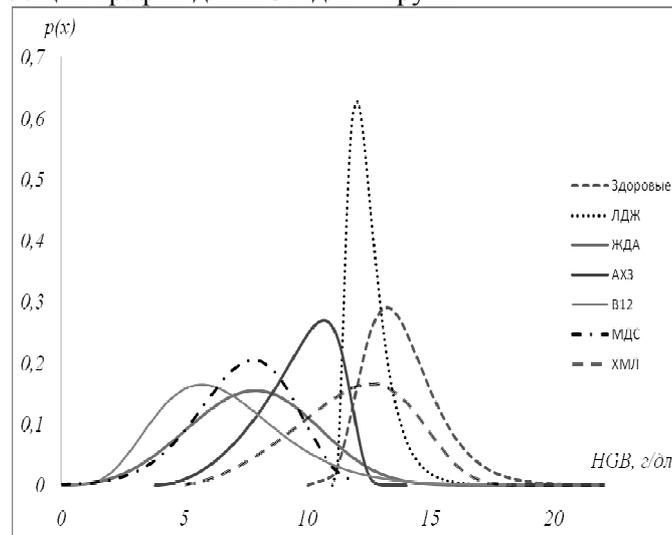


Рис. 3. Плотности вероятности показателя *HGB* для исследуемых групп.

Далее строится вероятностный график, который рассчитывается следующим образом. По диаграмме плотности вероятности для каждого значения данного показателя суммируются плотности вероятностей всех заболеваний и нормируются на 100%. Если известны вероятностные диагнозы, например, по восьми показателям, то вероятности каждого заболевания перемножаются и нормируются таким образом, чтобы сумма вероятностей всех возможных диагнозов была равна единице (или 100%).

Ни по одному показателю в отдельности нет возможности установить диагноз с большой вероятностью (хотя бы 70-80%). Но при сопоставлении всё большего числа показателей с высокой вероятностью (90% и выше) можно утверждать отсутствие или наличие патологии.

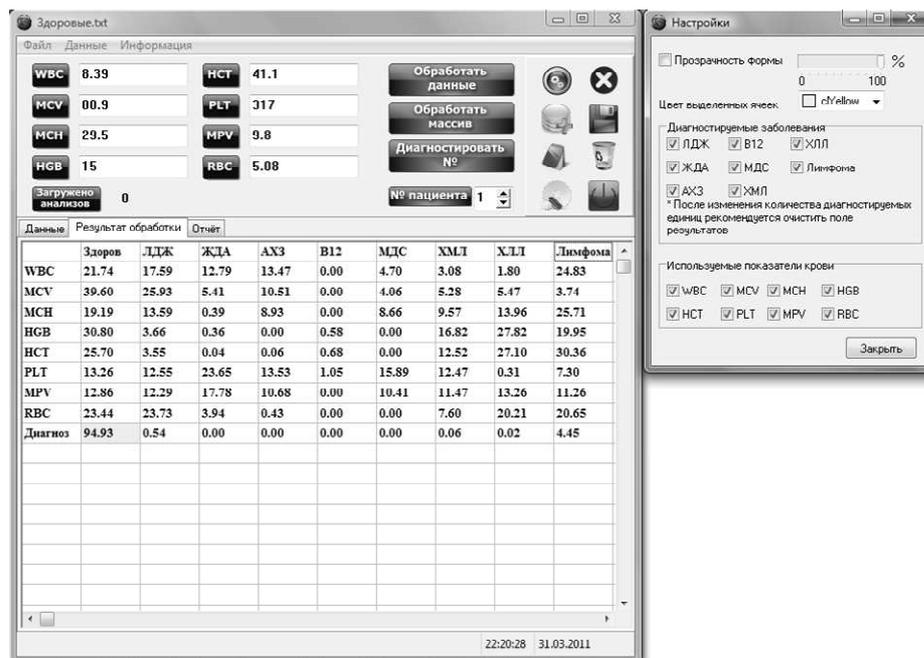


Рис 4. Окно программы вероятностной диагностики

На рисунке 4 представлена программная реализация метода вероятностной диагностики, написанная на языке C++. Возможности программы включают диагностику по числу параметров от 1 до 8 и по количеству групп от 2 до 8. Реализована функция обработки большого массива данных, что является весьма ценным при проведении профилактических обследований населения. В настоящее время выделить из большого потока обследуемых людей лиц, нуждающихся в более полной диагностике, достаточно трудоемко. Применение автоматизированной системы вероятностной диагностики позволит существенно снизить временные затраты медицинского персонала и повысить эффективность скрининга населения.

Был обследован массив данных, состоящий из 3900 гемограмм, полученных при скрининге населения г. Краснодара. По результатам вероятностной диагностики 37% относятся к группе здоровых лиц, 33% имеют дефицит железа, в 26% случаев имеет место наличие патологии кроветворной системы, еще у 4% диагноз не был установлен ввиду того, что ни по одной из групп вероятностный диагноз не превысил 50%.

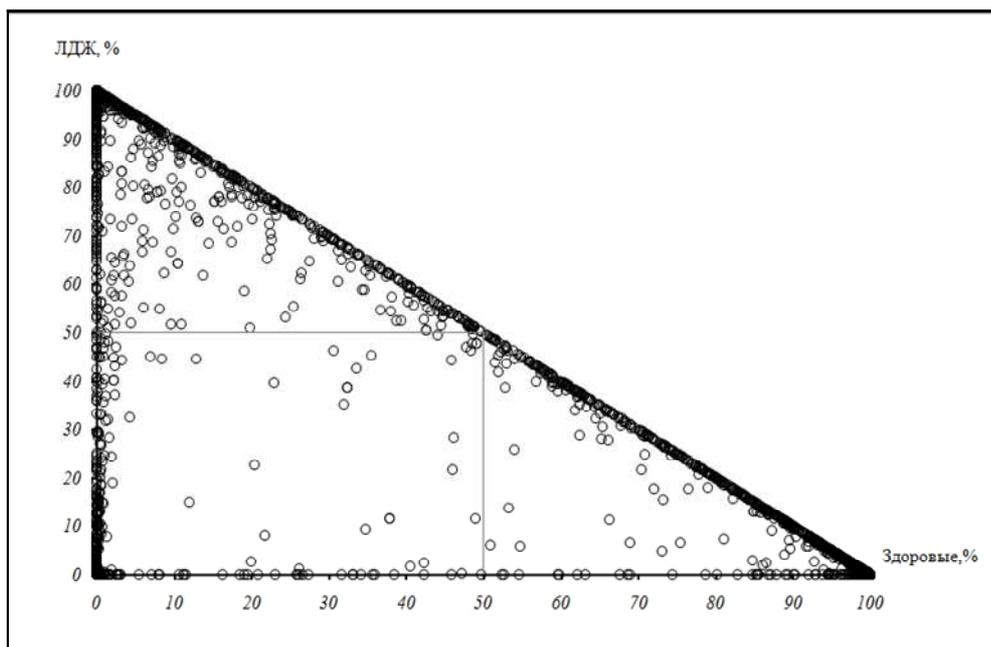


Рис 5. График зависимости вероятностных диагнозов “Здоров” и “ЛДЖ”

На рисунке 5 показан график зависимости вероятностных диагнозов “Здоров” и “ЛДЖ”. Третья “сторона” представленного треугольника состоит из значений диагнозов, которые разделяются только между двумя состояниями: отсутствием патологии и дефицитом железа. Значения, расположенные внутри “треугольник” подразумевают наличие некоего третьего состояния.

По данным Всемирной организации здравоохранения треть населения планеты в той или иной степени имеет недостаток железа в организме. Данный метод позволяет, помимо диагностики, получить общую картину рассматриваемой популяции, что может быть интересно с точки зрения статистики.

В результате работы предложена автоматизированная методика диагностики нарушений кроветворной системы человека. Использование в полной мере информации, поступающей от гематологических анализаторов, позволяет повысить точность, а использование ПК и скорость обработки данных. Другим преимуществом предлагаемой методики является способность диагностической системы к развитию и самосовершенствованию, ведь с каждым верифицированным диагнозом, пополняется база данных, уточняются вероятностные распределения. Выявленная высокая частота заболеваемости диктует необходимость повсеместной диспансеризации населения.