

УДК 004.94

## ПРИМЕНЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ ТЕХНОЛОГИЙ ДЛЯ ДИАГНОСТИКИ ОНКОЛОГИЧЕСКИХ ЗАБОЛЕВАНИЙ

А. А. Шайдунов

Алтайский государственный университет,  
г. Барнаул

В статье предлагается метод построения модели классификации онкологических заболеваний с помощью слоистых искусственных нейронных сетей. Используются данные, полученные при помощи технологии IMMUNOSIGNATURE. В статье рассматриваются вопросы коррекции получаемых данных при помощи корреляционного анализа для избавления от случайных артефактов. Рассмотрен вопрос нормализации данных для минимизации влияния системных искажений, возникающих вследствие влияния внешних факторов. В качестве классификатора используется искусственная нейронная сеть обратного распространения ошибки.

**Ключевые слова:** искусственные нейронные сети, технология IMMUNOSIGNATURE, классификация онкологических заболеваний.

Роль информационных технологий в развитии общества чрезвычайно велика. С ними связано начало революции в области накопления, передачи и обработки информации. Эта революция, следующая за революциями в овладении веществом и энергией, затрагивает и коренным образом преобразует не только сферу материального производства, но и интеллектуальную, духовную сферы жизни.

Рост производства компьютерной техники, развитие информационных сетей, создание новых информационных технологий приводят к значительным изменениям во всех сферах общества: в производстве, науке, образовании, медицине и т.д.

В качестве яркого примера внедрения информационных технологий в систему здравоохранения можно привести технологию «immunosignature» – простой и недорогой способ, используемый для ранней диагностики онкологических заболеваний, появившийся на основании современных исследований о биочипах.

Суть метода иммуносигнатур заключается в отслеживании иммунных реакций. Каждое заболевание вызывает ответ иммунной системы, по которому можно узнать, болен ли человек и чем именно.

Для получения иммуносигнатуры капли крови обследуемого, содержащая в себе клетки иммунной системы, наносится на биочип, разделённый на сектора, в каждом из

которых находится уникальная аминокислотная последовательность. Всего на пластинке от 10 000 до 330 000 секторов. В зависимости от того, с какой интенсивностью в конкретных ячейках проявится иммунная реакция – формируется профиль (сигнатура) для конкретного человека. Естественно, у каждого человека он индивидуален. Сама пластинка имеет небольшой размер, аналогичный предметному стеклу для светового микроскопа. Её удобно перевозить, и для неё не требуется большого объёма исследуемой крови. Иммуносигнатура показывает, какие последовательности аминокислот активируют иммунную систему пациента, к чему у него выработано больше антител, то есть с чем организм борется в данный момент. Зная это, можно понять, чем человек болен.

Революционный прорыв в данной технологии заключается в том, что стало возможно совершить продвижение с постсимптоматической медицины (когда опухоль уже развилась в организме) к предсимптоматической медицине.

Суть технологии – выявление иммуносигнатур, характерных для определенного заболевания. Технология интенсивно изучается, исследования обширны и дают хорошие результаты в плане точности диагностики. Особенно метод перспективен для ранней диагностики рака, возможной ещё на тех стадиях, когда раковая опухоль не может быть выявлена традиционными методами [1].

В ходе проводимых экспериментов были получены результаты исследования иммуно-сигнатур для двух групп людей. Одна группа людей – контрольная группа людей с не выявленной онкологией (25 человек), а другая группа – пациенты с диагнозом C50 «Злокачественное новообразование молочной железы» (10 человек). Чтобы получить достоверные данные, для каждого пациента проводилось три эксперимента, в разные моменты времени на разных чипах.

Полученный результат представляет собой набор различных показателей более чем по десяти тысячам пептидам. К основным показателям, используемым нами можно отнести: среднее и медианное значения светимости пептида, логарифм отношения для каждого пептида, среднее и медианное значения светимости фона в окрестности пептида.

Среднее значение светимости может дать искаженное представление, так как оно является слишком чувствительным параметром к так называемым «артефактам» — нехарактерным для изучаемой выборки. Медианное значение светимости более устойчиво к «артефактам». В случае же нормального распределения медиана совпадает со средним значением. [2]

При анализе результатов исследования иммуносигнатур было обнаружено, что одни и те же показатели в разных экспериментах могут довольно сильно отличаться. Это поставило под вопрос адекватность полученных данных. Для того чтобы разобраться в причинах появления столь значимых различий, было выдвинуто предположение, что на результат эксперимента оказывают влияние как микрочастицы пыли, находящиеся в окружающем воздухе, так и технические погрешности, возникающие при печати пептидных последовательностей. Артефакты, располагающиеся на секторах с аминокислотными последовательностями вызывают сильное свечение соответствующих ячеек. Этот эффект может сильно влиять на конечный результат.

Таким образом, на начальном этапе исследований, необходимо было выявить, как сильно артефакты могут влиять на конечный результат исследований. Влияние такого внешнего фактора было решено оценить методом корреляционного анализа. Для этого были взяты анализы группы из 10 пациентов людей по трем экспериментам, и рассчитан коэффициент корреляции для трех экспериментов по каждому пациенту.

На первом этапе велся расчет по всем

пептидным последовательностям. В результате у пациентов коэффициент корреляции находился в пределах от 0,623 до 0,97. Далее, мы поставили ограничение на логарифм отношения средней светимости ячейки к медианной светимости:  $\log(\text{Mean/Median}) \leq 0,1$ . В результате, корреляция несколько улучшилась. Но чтобы окончательно отсеять всю пыль, ограничение пришлось увеличить:  $\log(\text{Mean/Median}) \leq 0,01$

В результате были исключены все артефактные пептиды. Для пациентов группы с показателями, отклоненными от нормы, коэффициент корреляции значительно улучшился. Это подтвердило предположение, что пыль, попавшая на подложку, оказывает значительное воздействие на результат анализа.

На следующем этапе необходимо было осуществить нормализацию данных для минимизации влияния внешних факторов (температура и влажность окружающей среды, концентрация компонентов и т.д.) при протекании реакции связывания пептидов в процессе осуществления технологии IMMUNOSIGNATURE.

В ходе проведенных экспериментов было выявлено влияние внешних факторов на светимость пептидных последовательностей. При этом суммарная светимость пептидов для одного пациента может изменяться в несколько раз при проведении нескольких экспериментов.

На основе эмпирического анализа, был предложен метод нормализации данных. Рассчитанные нормализованные данные исследовались методом доверительных интервалов. То есть целью численного эксперимента стало выявление наиболее значимых пептидных последовательностей, которые на протяжении всех проведенных экспериментов имеют стабильное значения светимости с небольшим отклонением. Выявление таких последовательностей должно позволить осуществить классифицирование между разными группами пациентов.

В результате проведенных исследований были выявлены пептиды, которые для всех пациентов в пределах одной группы имеют стабильное значение светимости с малым отклонением. Таким образом, для группы пациентов с диагнозом C50 «Злокачественное новообразование молочной железы» было выявлено 20 характерных пептидов, однозначно относящихся пациента к данной нозологической группе.

Дальнейший анализ данных был осуществлен при помощи слоистых нейронных се-

тей на основе нейропарадигмы «Backpropagation» [3]. В ходе проектирования искусственной нейронной сети (ИНС) «Backpropagation», была выбрана оптимальная архитектура, позволяющая распознать рак молочной железы.

Эксперимент проходил следующим образом. В исходной выборке присутствовало 10 пациентов с диагнозом и 25 человек контрольной группы. Так как по каждому пациенту было осуществлено три эксперимента, то общее число записей выборки составило 105. Исходная выборка была разделена на 2 части: обучающая (35 записей) и тестовая (70 записей). В обучающую выборку были включены записи первых экспериментов, а в тестовую – записи вторых и третьих экспериментов.

В ходе нейросетевого исследования число скрытых нейронов ИНС менялось от 2 до 7. В результате чего были определены зависимости ошибки обобщения от числа нейронов в скрытом слое. Эксперимент показал, что сети с 4 нейронами в скрытом слое оптимально классифицирует пациентов из выборки.

Таким образом, в ходе эксперимента были реализованы методы использования

ИНС при решении задачи классификации данных. Разработанная модель анализа данных технологии IMMUNOSIGNATURE на основе применения традиционных и нейросетевых алгоритмов позволила с достаточной точностью (90%) классифицировать пациентов с онкологическим диагнозом C50 «Злокачественное новообразование молочной железы».

#### СПИСОК ЛИТЕРАТУРЫ

1. Alexa K Hughes, Zbigniew Cichacz, Adrienne Scheck, Stephen W. Coons, Stephen Albert Johnston, Phillip Stafford Immunosignaturing Can Detect Products from Molecular Markers in Brain Cancer, 2012, DOI: 10.1371/journal.pone.0040201
2. Реброва О. Ю. Среднее или всё же медиана? / О. Ю. Реброва // Троицкий Вариант.- 2011.- № 90.- С. 13
3. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. - Новосибирск: Наука, 1996. - 276 с.

**Шайдунов Александр Алексеевич – к.т.н., доцент кафедры вычислительной техники и электроники, тел.: (3852) 367-059, e-mail: shaidurov@phys.asu.ru.**