

Разработанный алгоритм автоматизации логистического учета, помимо информационной системы 1С: предприятие, может быть реализован в любой среде программирования. Таким образом, возможна реализация отдельного программного обеспечения под любую платформу. При этом сложность разработки будет определяться только возможностями среды программирования. В качестве перспективного направления реализации логистической системы, предполагается создание интеллектуального блока анализа логистической информации. [5, 6]

#### Список литературы

1. Академик. [Электронный ресурс] Режим доступа: [http://dic.academic.ru/dic.nsf/econ\\_dict/6534](http://dic.academic.ru/dic.nsf/econ_dict/6534), свободный. – Загл. с экрана – Яз. Рус. Дата доступа – 12.04.2014.
2. Ассоциация автоматической идентификации "ЮНИСКАН / ГС1 РУС." [Электронный ресурс] Режим доступа: <http://www.gs1.ru.org/technologies/codes/>, свободный. – Загл. с экрана – Яз. Рус. Дата доступа – 22.04.2014.
3. ID-Expert [Электронный ресурс] – Режим доступа: <http://www.idexpert.ru/technology/121/>, свободный. – Загл. с экрана – Яз. Рус. Дата доступа – 22.04.2014.
4. Rmk-shop [Электронный ресурс] – Режим доступа: <http://www.rmk-shop.ru/rfid-technology.html>, свободный. – Загл. с экрана – Яз. Рус. Дата доступа – 10.05.2014.
5. Шайдуrow А.А. Финансовое моделирование при помощи многокритериальной оптимизации // Вестник Алтайской академии экономики и права. 2008. № 12. С. 110-111.
6. Шайдуrow А.А. // Экономический анализ предприятий на основе применения слоистых нейронных сетей. Предпринимательство. 2009. № 3. С. 74-79.

УДК 004.94

### КЛАССИФИКАЦИЯ ОНКОЛОГИЧЕСКИХ ЗАБОЛЕВАНИЙ НА ОСНОВЕ АНАЛИЗА РЕЗУЛЬТАТОВ ТЕХНОЛОГИИ IMMUNOSIGNATURE

**А.С. Бубликов**, Алтайский государственный университет, Барнаул, Россия,  
E-mail: madsilver@rambler.ru

### CLASSIFICATION CANCER BY ANALYSIS OF THE RESULTS TECHNOLOGIES IMMUNOSIGNATURE

**A.S. Bublikov**, Altai State University, Barnaul, Russia, E-mail: madsilver@rambler.ru

В статье рассматривается анализ возможности построения модели классификации онкологических заболеваний с помощью слоистых искусственных нейронных сетей. Исследовались данные, полученные при помощи технологии IMMUNOSIGNATURE. Сложность анализа данных заключается стохастическом влиянии внешних факторов на получаемые результаты 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 в следствии влияния таких внешних факторов, как температура, влажность, концентрация вещества и т.д. В качестве классификатора используется искусственная нейронная сеть обратного распространения ошибки.

**Ключевые слова:** искусственная нейронная сеть, метод доверительных интервалов, корреляционный анализ, технология IMMUNOSIGNATURE, онкология, рак молочной железы.

На основании современных исследований о биочипах, появилась технология «immunosignature», простой и недорогой способ, используемый для ранней диагностики онкологических заболеваний.

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 капля 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, и 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 пост-симптоматической медицины (когда опухоль уже развилась в организме) к пред-симптоматической медицине.

Суть технологии — в том, чтобы выявить иммуносигнатуры, характерные для какого-то одного 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 речи не идёт о постановке страшного диагноза. [1]

В ходе экспериментов были получены результаты исследования иммуносигнатур для двух групп людей. Одна группа людей – контрольная группа людей с не выявленной онкологией (25 человек), а другая группа – пациенты с диагнозом С50 «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]

Логарифм отношения нужен нам для того, чтобы увидеть, есть ли существенное различие в показателях средней и медианной светимостей. И если таковое имеется, можно сказать о том, что результаты содержат в себе «артефакты», не характерные для данной выборки.

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 микрочастицы пыли, находящиеся в окружающем воздухе. Пылинки, попадая на сектора с аминокислотными последовательностями вызывают сильное свечение соответствующих ячеек. Этот эффект может сильно влиять на конечный результат. На рисунке 1 приведена гистограмма распределения средней светимости пептидов ЕМРТУ по трем экспериментам для десяти пациентов с раком молочной железы. Теоретически, пептиды ЕМРТУ должны давать единый минимальный уровень светимости (светимость ЕМРТУ должна быть на уровне светимости фона). Однако, как видно на гистограмме светимости ЕМРТУ для одного пациента в трех экспериментах могут отличаться в несколько раз.

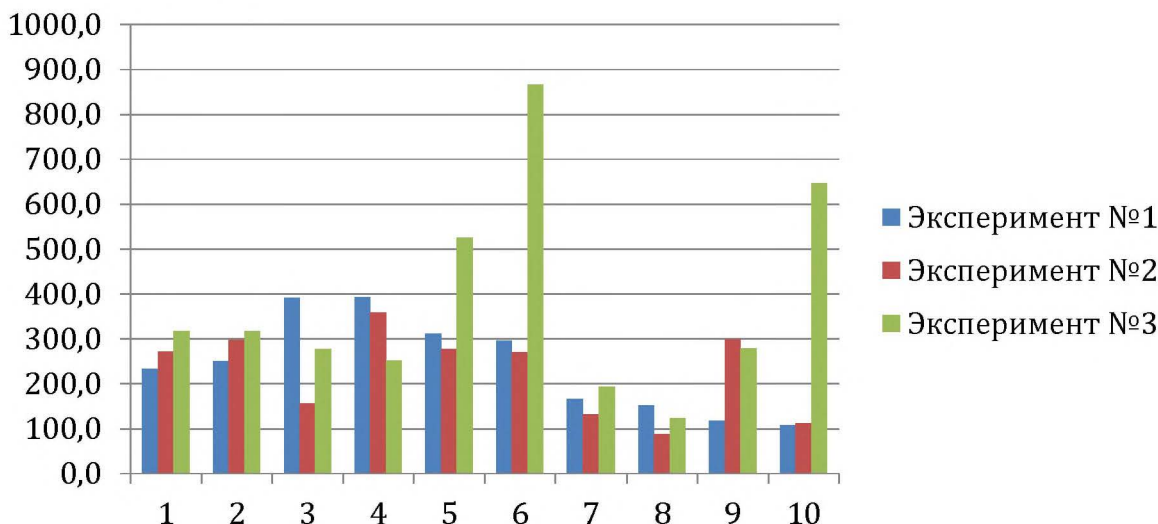


Рис.4 Светимости пептидов ЕМРТУ по трем экспериментам для группы пациентов

Таким образом, на начальном этапе исследований, необходимо было выявить, как сильно внешнее воздействие пыли может влиять на конечный результат исследований. Влияние такого внешнего фактора было решено оценить методом корреляционного анализа. Чтобы проверить, верно ли предположение о попадании пыли на подложку и влиянии ее на конечный результат, 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:  $\log(\text{Mean}/\text{Median}) \leq 0,1$ . В результате, корреляция несколько улучшилась. Но чтобы окончательно отсеять всю пыль, ограничение пришлось увеличить:  $\log(\text{Mean}/\text{Median}) \leq 0,01$

В результате были исключены все артефактные пептиды. Однако на результаты пациентов, у которых корреляция изначально соответствовала норме, это практически не повлияло. Но для пациентов группы с показателями, отклоненными от нормы, она значительно улучшилась. Это подтвердило предположение, что пыль, попавшая на подложку, оказывает значительное воздействие на результат анализа. Результаты эксперимента приведены в таблице 1.

Таблица 1 Значения коэффициента корреляции при различных порогах учета пыли

| Номер пациента | пыль не учитывается | Log (Mean/Median) <0,1 | Log (Mean/Median) <0,001 |
|----------------|---------------------|------------------------|--------------------------|
| 1              | 0,92                | 0,92                   | 0,92                     |
| 2              | 0,97                | 0,97                   | 0,97                     |
| 3              | 0,93                | 0,93                   | 0,93                     |
| 4              | 0,96                | 0,96                   | 0,96                     |
| 5              | 0,91                | 0,92                   | 0,92                     |
| 6              | 0,83                | 0,84                   | 0,86                     |
| 7              | 0,82                | 0,84                   | 0,91                     |
| 8              | 0,71                | 0,72                   | 0,89                     |
| 9              | 0,62                | 0,65                   | 0,69                     |
| 10             | 0,72                | 0,73                   | 0,81                     |

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. На рисунке 2 изображено типичное распределение средней светимости пептидов EMPTY и остальных пептидов по трем экспериментам для одному пациента.

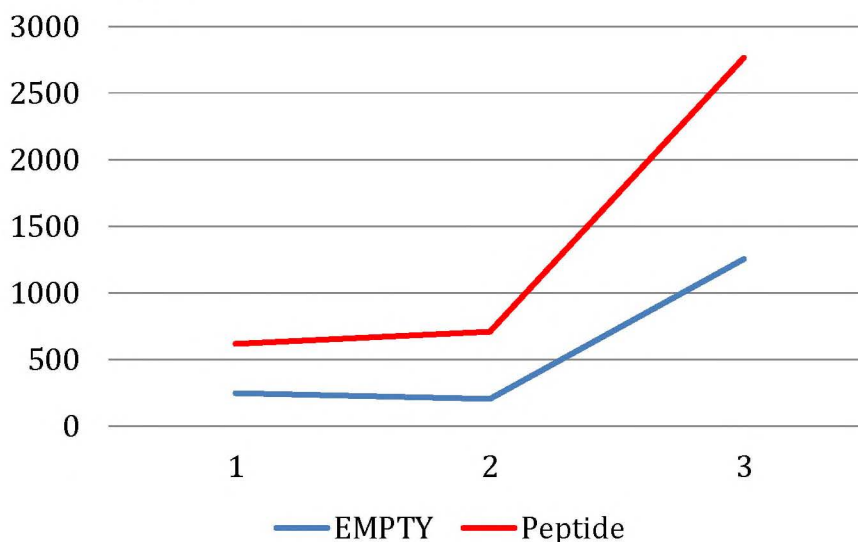


Рис. 5. Средние значения светимостей по трем экспериментам для одного пациента

Как видно, на эксперимент, влияют внешние параметры. При этом суммарная светимость для одного пациента может изменяться в несколько раз при проведении нескольких экспериментов.

На основе эмпирического анализа, был предложен метод нормализации данных. Цель нормализации заключается в том, чтобы в результате нормализации все светимости EMPTY имели одинаковые и минимальные значения:

1. Рассчитать нормировочный коэффициент по формуле  $1/(\text{средняя светимость EMPTY})$ .
  2. Все светимости пептидов преобразовывать следующим образом:
- если светимость больше удвоенной средней светимости для EMPTY, то она умножается на

- нормировочный коэффициент;
- если светимость меньше или равна удвоенной средней светимости для EMPTU, то она приравнивается к нулю;
- Если пептид является артефактным, то он исключается из дальнейших расчетов.

На следующем этапе 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 относящих пациента к данной нозологической группе.

Дальнейший анализ данных был осуществлен при помощи слоистых нейронных сетей на основе нейропарадигмы «Backpropagation» [3, 4]. В ходе проектирования искусственной нейронной сети (ИНС) «Backpropagation», была выбрана оптимальная архитектура, позволяющая распознать рак молочной железы. На рисунке 3 представлен сравнительный анализ архитектур ИНС.

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 был получен график зависимости ошибки обобщения от числа нейронов в скрытом слое. Эксперимент показал, что сети с 4 нейронами в скрытом слое оптимально классифицирует пациентов из выборки.

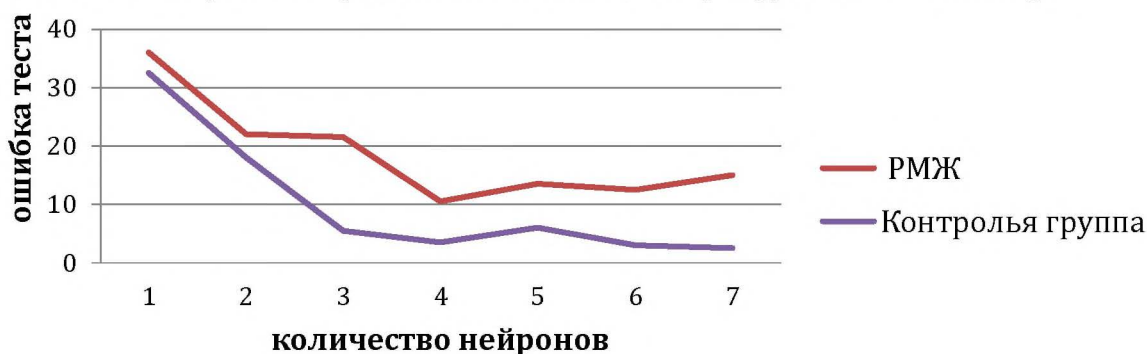


Рис. 6. Динамика ошибки обобщения искусственной нейронной сети при изменении количества скрытых нейронов

Таким образом, в ходе эксперимента были реализованы методы использования ИНС при решении задачи классификации данных [5]. 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.
- 4 Нейроинформационные технологии. учебное пособие / А.А. Шайдулов. Барнаул, 2014.
- 5 Шайдулов А.А. Построение многокритериальной нейросетевой модели оптимизации в задачах диагностики заболеваний со слабо выраженной симптоматикой. // Известия Алтайского государственного университета. 2013. № 1-2 (77). С. 135-138.