



УДК 303.732

© 2024 г. З.Д. Кузьмин¹,

А.А. Яковлев^{1,2}, канд. физ.-мат. наук

¹(Дальневосточный федеральный университет, Владивосток)

²(Институт автоматизации и процессов управления ДВО РАН, Владивосток)

ПРИМЕНЕНИЕ НЕЙРОННОЙ СЕТИ ДЛЯ РАННЕЙ ДИАГНОСТИКИ РАКА ШЕЙКИ МАТКИ НА ОСНОВЕ ЦИТОКИНОВОГО ПРОФИЛЯ ПАЦИЕНТА*

По данным концентрациям цитокинов проведено исследование связей между предикторами. Установлено, что данные не подчиняются нормальному закону распределения и отсутствуют явные корреляционные связи между представленными параметрами. Получена нейронная сеть, позволяющая с высокой точностью прогнозировать рак шейки матки на основе цитокинового профиля пациента.

Ключевые слова: вирус папилломы человека, рак шейки матки, цитокиновый профиль, машинное обучение, нейронная сеть, моделирование.

DOI: 10.22250/18142400_2024_79_1_35

Введение

Вирус папилломы человека (ВПЧ) является распространенной инфекцией, передающейся половым путем, которая может поражать кожу, область гениталий и горло. Опасна она тем, что длительное время может «жить» в организме, никак не проявляя себя [1]. Особая значимость ВПЧ состоит в широкой распространенности, высокой контагиозности и в онкогенном потенциале. Вирус может негативно влиять на репродуктивную систему человека: потенциально вызывать преждевременные роды, раннюю потерю беременности [2]. В большинстве случаев иммунная система очищает организм от ВПЧ. Персистирующая ВПЧ-инфекция шейки матки (нижняя часть матки, которая открывается во влагалище, называемая также родовым кана-

* Работа выполнена в рамках государственного задания FFW-2022-0002.

лом) в отсутствие лечения может стимулировать развитие аномальных клеток, вызывающих в 95% случаев рак шейки матки (РШМ), который занимает первое место в структуре онкологических заболеваний женских половых органов и четвертое место среди всей онкопатологии (после рака молочной железы, колоректального рака и рака легкого) [1].

Ранняя диагностика рака является критически важной для успешного лечения и увеличения выживаемости пациентов.

Лечение РШМ до настоящего времени может быть радикальным лишь при раннем выявлении, тогда как частота запущенных случаев может превышать 40%. Поэтому совершенствование консервативного лечения РШМ остается актуальной задачей. Во многих современных исследованиях рассматривается роль цитокинов в патогенезе и при лечении злокачественных новообразований, поскольку их количественные характеристики и соотношения между собой, как правило, отражают динамику патологического процесса, коррелируют с активностью заболевания и позволяют судить об эффективности проводимого лечения [3, 4]. Уровень цитокинов в биологических образцах может содержать информацию о наличии или отсутствии онкологического заболевания, что делает его потенциальным биомаркером для диагностики. При этом, несмотря на наличие хорошо разработанных методов исследования, стандартизацию подходов к определению тех или иных цитокинов, в публикациях содержатся разноречивые данные о направленности и степени их изменений у больных со злокачественными новообразованиями как в ткани опухоли, так и в сыворотке крови [5].

Традиционные статистические методы могут оказаться недостаточно эффективными для полного выявления сложных взаимосвязей между уровнем цитокинов и наличием рака. Эффективность ИИ в диагностике РШМ и ВПЧ заключается в его способности обрабатывать большие объемы данных, выявлять предикторы, незаметные человеческому глазу, и предоставлять врачам ценную информацию для более точной диагностики и разработки индивидуальных методов лечения.

Системы искусственного интеллекта (ИИ) играют значительную роль в диагностике и прогнозировании онкологических заболеваний, включая ВПЧ и РШМ. Алгоритмы машинного обучения, нейронные сети и другие методы ИИ позволяют анализировать медицинские данные, включая различные биомаркеры, для выявления признаков заболеваний на ранних стадиях и оказания поддержки врачам в принятии решений.

В настоящей работе нами была предпринята попытка применить нейронную сеть для прогнозирования онкопатологии у пациентов на основе

их цитокинового профиля.

Предмодельный анализ данных

Объектом исследования послужила группа из 111 женщин в возрасте от 18 до 70 лет, направленных врачами МЦ «Асклепий» и женских консультаций Приморского края, а также самостоятельно обратившихся для диагностических исследований ВПЧ в ООО ДЦ «Лабораторная диагностика – Асклепий» г. Владивостока.

Комплексное клинико-лабораторное исследование проводили в амбулаторных условиях по единой программе, включающей молекулярно-генетическое, цитологическое и иммунологическое обследование с целью детализации патофизиологических механизмов папилломавирусной инфекции в урогенитальном тракте. Биоматериалом для исследования послужили соскобы, взятые с эпителия цервикального канала и шейки матки. Оценивался локальный уровень концентрации цитокинов IL-1, TNF- α , IL-6, IL-10, IL-18, IL-8, VEGF, TGF- β 1.

По результатам лабораторных исследований врачами-экспертами пациентки были разнесены на 5 групп: контрольная группа – женщины без папилломовирусной инфекции (ПВИ) и отсутствием интраэпителиальных поражений; группа 1 – с ПВИ и отсутствием интраэпителиальных поражений; группа 2 – с ПВИ и интраэпителиальными поражениями низкой степени риска; группа 3 – с ПВИ и интраэпителиальными поражениями высокой степени риска (предраковое состояние); группа 4 – с ПВИ и раком шейки матки (РШМ). Распределение обследованных пациенток по группам представлено на рис. 1.

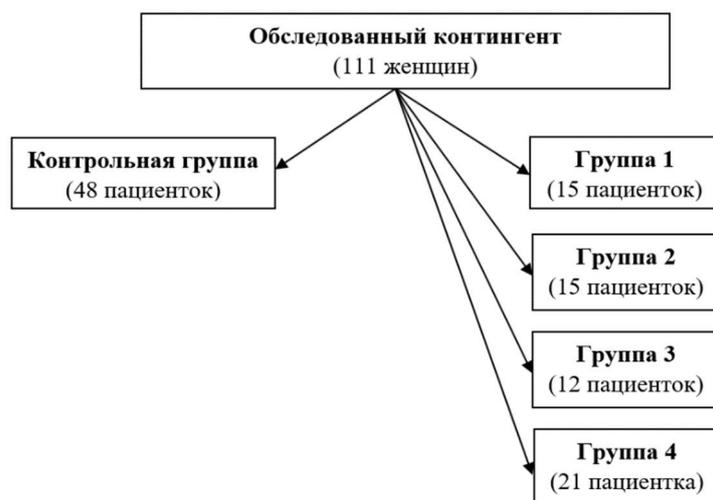


Рис. 1. Распределение пациенток по группам в зависимости от результатов лабораторных исследований.

Все цитокины из каждой группы пациенток были проверены на нормальное распределение с помощью критерия Шапиро – Уилка [6]. За нулевую гипотезу H_0 принимали предположение, что выборка принадлежит нормальному распределению. Проверка нулевой гипотезы осуществлялась на уровень значимости $p = 0.05$. Если полученный уровень значимости принимал значение $p > 0.05$, то нулевая гипотеза принималась.

По результатам проверки было выявлено, что все исследуемые цитокины в каждой из групп женщин не подчиняются нормальному закону распределения. На рис. 2 показан пример проверки для цитокина IL-18 из второй группы. Здесь по вертикальной оси отложена величина плотности распределения, по горизонтальной – концентрация цитокина. Синей линией обозначена кривая плотности нормального распределения.

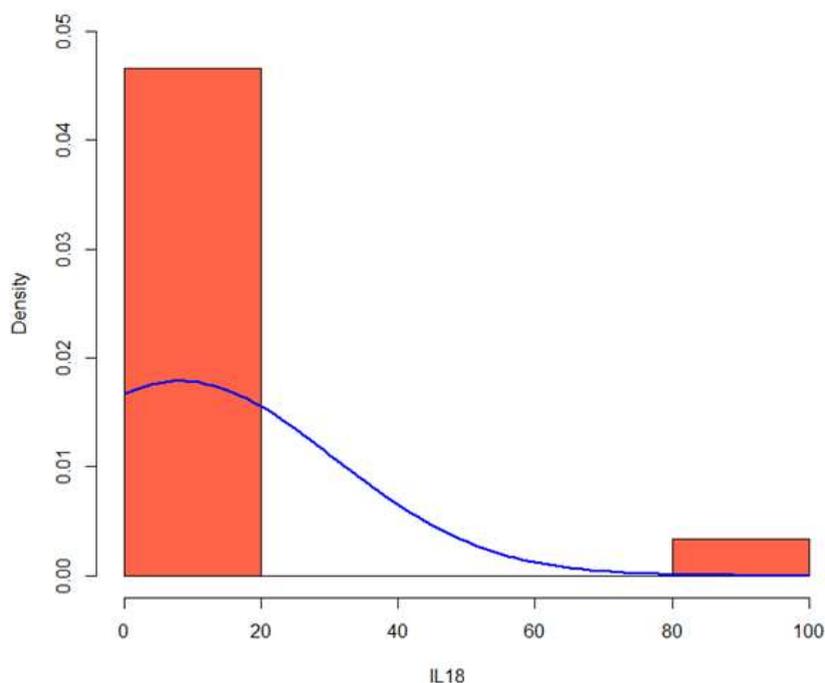


Рис. 2. Гистограмма распределения цитокина IL-18 для второй группы.

Так как исследуемые данные не подчиняются закону нормального распределения, то для оценки различий в концентрации цитокина между независимыми группами пациенток было принято решение использовать критерий Манна – Уитни [7].

Данный критерий считается для уровня значимости $p\text{-value} = 0.05$. Нулевой гипотезой H_0 считается утверждение, что концентрации цитокинов схожи. Если значение $p\text{-value} > 0.05$, то принимается нулевая гипотеза.

После проведения исследования был получен следующий результат: различия в концентрациях цитокинов существенны только в единичных случаях, в основном это проявилось между четвертой, раковой группой и остальными группами.

Какой-то системы при анализе результатов критерия Манна – Уитни обнаружено не было. Результаты критерия Манна – Уитни для цитокинов IL-6 и IL-8 представлены в табл. 1, 2. Здесь W – значение статистики, p – уровень значимости.

Таблица 1

		I группа	II группа	III группа	IV группа
Контрольная группа	W	343	479	367	435
	p	0.783	0.056	0.149	0.372
I группа	W		160	116	140
	p		0.054	0.213	0.585
II группа	W			86	104
	p			0.864	0.089
III группа	W				88
	p				0.160

Таблица 2

		I группа	II группа	III группа	IV группа
Контрольная группа	W	371	286	152	273
	p	0.865	0.139	0.012	0.003
I группа	W		78	46	93
	p		0.158	0.033	0.039
II группа	W			61	115
	p			0.167	0.1762
III группа	W				92
	p				0.200

Как видно из табл. 1, различий в концентрациях для цитокина IL-6 не зафиксировано. А вот для цитокина IL-1 (табл. 2) обнаружены различия между контрольной и первой группой, с одной стороны, и третьей и четвертой, – с другой.

Для визуальной оценки степени взаимосвязи между переменными и для анализа взаимосвязей между цитокинами была составлена матрица корреляций между каждой парой цитокинов (рис. 3).

В результате анализа полученной матрицы очевидных корреляций между цитокинами обнаружено не было.

Проведенный предмодельный анализ данных показал, что с помощью стандартных статистических методов невозможно выявить четкие закономерности между цитокинами и в дальнейшем построить модель для прогнозирования предварительного диагноза пациента. Поэтому было принято решение в дальнейшей работе использовать нейронные сети. Переход к использованию нейронных сетей обоснован необходимостью выявить более сложные зависимости в данных и открывает новые перспективы для даль-

нейшего исследования.

IL-1, пг/мл	1	0.18	0.37	0.27	0.37	0.56	0.62	-0.0038
TNF- α , пг/мл	0.18	1	0.51	0.31	0.5	0.5	0.21	0.19
IL-6, пг/мл	0.37	0.51	1	0.54	0.53	0.41	0.44	0.15
IL-10, пг/мл	0.27	0.31	0.54	1	0.19	0.21	0.34	0.063
IL-18, пг/мл	0.37	0.5	0.53	0.19	1	0.46	0.47	0.29
IL-8, пг/мл	0.56	0.5	0.41	0.21	0.46	1	0.57	-0.051
VEGF, пг/мл	0.62	0.21	0.44	0.34	0.47	0.57	1	0.0069
TGF- β , нг/мл	-0.0038	0.19	0.15	0.063	0.29	-0.051	0.0069	1
	IL-1, пг/мл	TNF- α , пг/мл	IL-6, пг/мл	IL-10, пг/мл	IL-18, пг/мл	IL-8, пг/мл	VEGF, пг/мл	TGF- β , нг/мл

Рис. 3. Матрица корреляций между цитокинами.

Разработка нейронной сети

Искусственные нейронные сети представляют собой технологию, уходящую корнями во множество дисциплин: нейрофизиологию, математику, статистику, физику, компьютерные науки и технику. Они находят применение в таких разнородных областях как моделирование, анализ временных рядов, распознавание образов, обработка сигналов благодаря одному важному свойству – способности обучаться на основе данных при участии учителя или без его вмешательства [8].

Для начала работы с нейронными сетями был выполнен важный этап – нормирование данных. В данном случае выбран метод, при котором каждый признак был поделен на модуль максимального значения в соответствующем столбце данных. Этот подход уменьшает разброс значений, обеспечивает стабильность обучения некоторых нейронных сетей и предотвращает проблемы, связанные с различными масштабами признаков.

Дополнительным шагом в подготовке данных для обучения нейронных сетей стало упрощение классификации групп. На основании проведенного

анализа данных критерием Манна – Уитни (табл. 1 и 2) было решено укрупнить исследуемые группы. Вместо изначального деления на 5 групп (контрольную и 4 степени заболевания) по результатам исследования было решено использовать только две категории:

1. Условно здоровая группа. Эта категория объединяет контрольную группу и группы без рака (первую и вторую степени заболевания).

2. Онкологическая группа. Эта категория включает третью и четвертую группы, т. е. с раком и предраковым состоянием.

Это решение было принято с целью упростить задачу классификации и сфокусироваться на более общих паттернах между контрольной группой и группой с отклонениями, в данном случае – с раком.

Разработка нейронной сети осуществлялась в языке объектно-ориентированного программирования Python.

Для решения задачи классификации было проведено сравнение различных нейронных сетей с использованием следующих классификаторов из библиотеки `scikit-learn`:

1. `RandomForestClassifier`. Это ансамблевый метод, основанный на построении множества деревьев решений. Он обучается на подмножествах данных и использует механизм голосования для принятия решений.

2. `SVC (Support Vector Classifier)`. SVM (метод опорных векторов) используется для задач классификации и регрессии. Он строит гиперплоскость в пространстве признаков, максимизируя расстояние между разными классами.

3. `MLPClassifier`. Многослойный перцептрон (или нейронная сеть прямого распространения) с несколькими скрытыми слоями. Он обучается с использованием метода обратного распространения ошибки.

4. `DecisionTreeClassifier`. Классификатор, основанный на построении деревьев решений. Он разбивает данные на подгруппы, принимая решения на основе значений признаков.

5. `KNeighborsClassifier`. Метод k-ближайших соседей. Классификация основывается на большинстве классов среди k ближайших соседей в пространстве признаков.

6. `GradientBoostingClassifier`. Ансамблевый метод, строящий деревья решений последовательно, при этом каждое следующее дерево направлено на исправление ошибок предыдущего.

7. `LogisticRegression`. Логистическая регрессия используется для бинарной классификации. Она моделирует логарифм отношения шансов как линейную комбинацию признаков.

8. `AdaBoostClassifier`. Еще один ансамблевый метод, который комби-

нирует слабые модели (чаще всего деревья решений) с целью улучшить производительность.

Эксперимент был проведен на 100 случайных разбиениях исходного набора данных с соотношением обучающей к тестовой выборке 75% и 25% соответственно. Результаты показали, что из перечисленных нейронных сетей наилучшей производительностью обладает MLPClassifier, что может быть обусловлено его способностью выявлять сложные взаимосвязи в данных.

Архитектура разработанной нейронной сети представляет собой входной слой, два скрытых слоя и выходной слой (рис. 4).

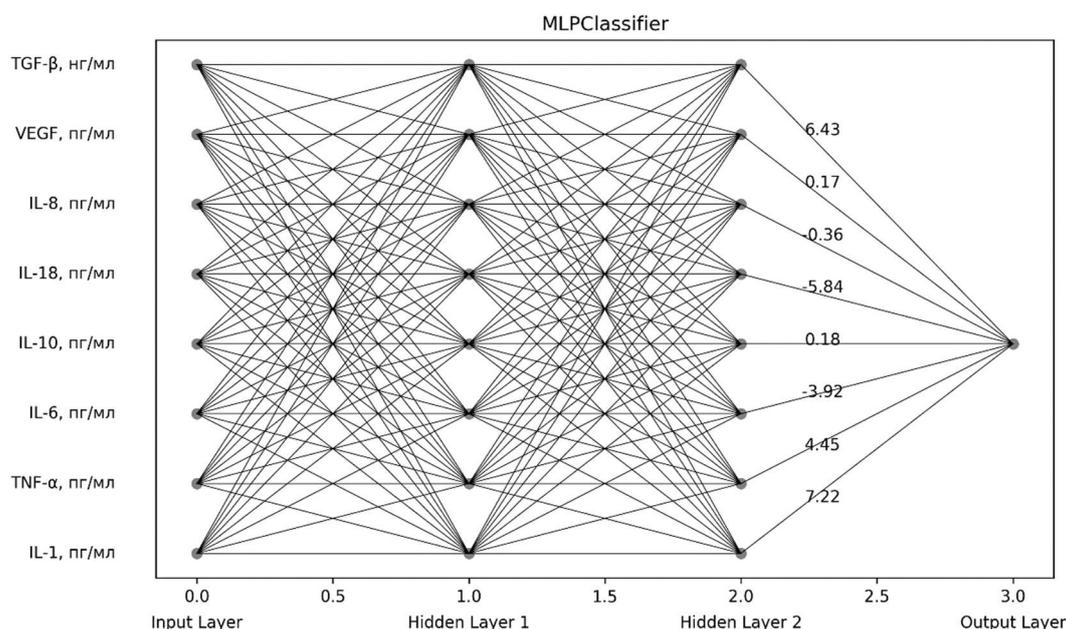


Рис. 4. Архитектура разработанной нейронной сети.

Для достижения оптимальной производительности модели был проведен тщательный процесс подбора семантических весов нейронной сети. Эта процедура является традиционным методом настройки нейронных сетей. Слои с подобранными семантическими весами показаны на рис. 5.

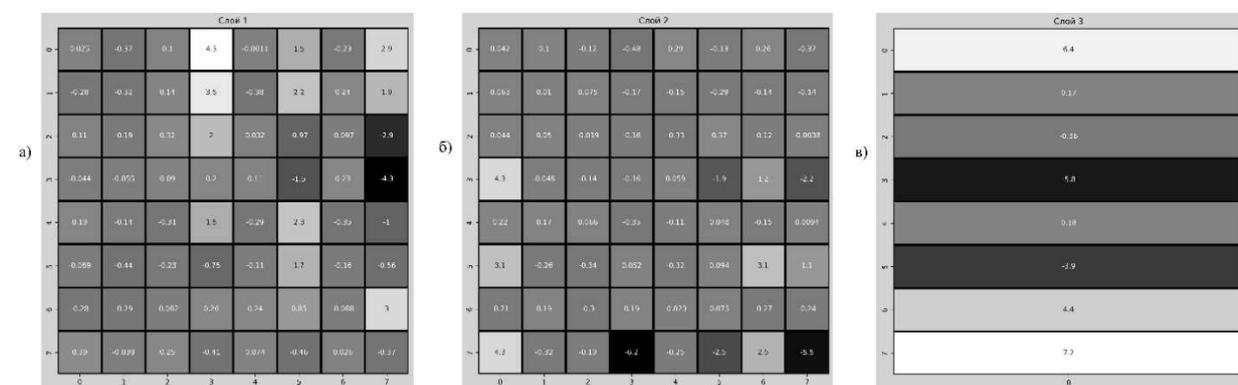


Рис. 5. Слои нейронной сети с семантическими весами: а) первый скрытый слой; б) второй скрытый слой; в) выходной слой.

Полученная комбинация параметров демонстрирует высокую производительность и эффективность в решении задачи классификации на данном наборе данных.

Результаты моделирования

Результаты моделирования были собраны в матрицу неточностей [9], представленную на рис. 6.

Факт	1	FN 1	TP 6
	0	TN 16	FP 0
		0	1
		Предсказание	

Рис. 6. Матрица неточностей.

В полученной матрице введены следующие обозначения:

1. True Positive (TP): 6 (верно предсказанных больных);
2. True Negative (TN): 16 (верно предсказанных здоровых);
3. False Positive (FP): 0 (ложноположительных предсказаний);
4. False Negative (FN): 1 (ложноотрицательных предсказаний)

По полученным результатам была рассчитана модельная производительность разработанной нейронной сети.

Точность:

$$Accuracy \approx \frac{TP + TN}{TP + TN + FP + FN} \approx \frac{6 + 16}{6 + 16 + 0 + 1} \approx 0.957.$$

Модель верно классифицирует около 95.7% всех случаев.

Чувствительность:

$$Sensitivity \approx \frac{TP}{TP + FN} \approx \frac{6}{6 + 1} \approx 0.857.$$

Модель обнаруживает примерно 85.7% всех фактических случаев заболевания раком.

Специфичность:

$$\text{Specificity} \approx \frac{TN}{TN + FP} \approx \frac{16}{16 + 0} = 1.$$

Модель различила всех здоровых пациентов.

Ложноотрицательные предсказания: модель совершила 1 ложноотрицательное предсказание, т. е. ошиблась, считая одного больного здоровым.

Заключение

Короткий ряд данных, всего 111 пациентов, конечно, повлиял на результаты моделирования, но полученная матрица неточностей показывает верно выбранное направление в применении нейронных сетей для ранней диагностики онкологических заболеваний. Использование современных методов машинного обучения представляется перспективным подходом для обнаружения скрытых паттернов и выявления непрямых связей в данных.

В перспективе результаты данного исследования позволят выявить наиболее эффективные модели машинного обучения для ранней диагностики рака на основе уровня концентрации цитокинов индивидуального цитокинового профиля пациента. Это может привести к разработке новых инструментов и подходов для более точной и ранней диагностики рака, что в конечном итоге будет способствовать улучшению результатов лечения и выживаемости пациентов.

Авторы благодарят д.м.н., профессора Е.В. Маркелову и аспирантку ТГМУ М.А. Черникову за предоставленные данные.

ЛИТЕРАТУРА

1. *Всемирная организация здравоохранения. Официальный сайт. URL: <https://www.who.int/ru/news-room/fact-sheets/detail/cervical-cancer> (дата обращения 24.01.2024).*
2. *Бахтияров К.Р., Щукина А.С. Вирус папилломы человека – современный взгляд на проблему // Здоровье и образование в XXI веке. – 2017. – Т. 19, № 12. – С. 37–42.*
3. *Цитокины как индикаторы состояния организма при инфекционных заболеваниях. Анализ экспериментальных данных / А.А. Яковлев, А.И. Абакумов, А.В. Костюшко, Е.В. Маркелова // Компьютерные исследования и моделирование. – 2020. – Т. 12, № 6. – С. 1409–1426.*
4. *Цитокиновый статус доноров крови и её компонентов / Г.А. Зайцева, О.А. Вершинина, О.И. Матрохина и др. // Фундаментальные исследования. – 2011. – № 3. – С. 61–65.*
5. *Кенбаева Д.К., Манамбаева З.А. Особенности цитокиновой регуляции у больных злокачественными новообразованиями шейки матки и влияние противоопухолевой терапии // Онкология. Журнал им. П.А. Герцена. – 2013. – № 1. – С. 17–20.*

6. https://ru.wikipedia.org/wiki/Критерии_нормальности (дата обращения 06.02.2024).
7. https://ru.wikipedia.org/wiki/U-критерий_Манна-Уитни (дата обращения 01.02.2024).
8. Хайкин С. Нейронные сети: полный курс. – Изд. 2-е/пер. с англ. – М.: Изд. дом «Вильямс», 2006.
9. https://learnmachinelearning.fandom.com/ru/wiki/Confusion_matrix (дата обращения 09.02.2024).

Статья представлена к публикации членом редколлегии А.И. Абакумовым.

E-mail:

Кузьмин Захар Данилович – kuzmin.zd@dvfu.ru;

Яковлев Анатолий Александрович – yakovlev-aa@iacr.dvo.ru.