

УДК 004.853

Модель бинарной логистической регрессии для выявления подозрения на злокачественную опухоль щитовидной железы по данным ультразвукового исследования

© О.Б. Цыдыпова

*Иркутский национальный исследовательский технический университет,
г. Иркутск, Российская Федерация*

Аннотация. Информационные системы и технологии находят активное применение в медицине и позволяют повысить качество медицинского обслуживания. Внедрение информационных систем для хранения данных медицинских исследований даёт возможность получить базу данных, на основании которой можно строить модели машинного обучения и использовать их для решения задач медицинской диагностики. В данной работе описана выявленная на основе имеющихся данных модель бинарной логистической регрессии для обнаружения подозрения на злокачественное новообразование щитовидной железы по данным, полученным в ходе ультразвукового исследования. Также рассмотрена созданная автором информационная система, реализующая эту модель и позволяющая автоматизировать работу врача при выполнении УЗИ щитовидной железы. Внедрение системы позволит не только уменьшить время, затрачиваемое врачом на заполнение протокола исследования, но и создать базу данных с возможностью отслеживания динамики заболеваний и накопления статистики, благодаря чему станет доступным применение других методов машинного обучения в задачах диагностики заболеваний щитовидной железы. Таким образом, создан инструмент поддержки принятия решений для врача при выполнении ультразвукового исследования щитовидной железы.

Ключевые слова: искусственный интеллект, анализ данных, система поддержки принятия решений, логистическая регрессия

Binary Logistic Regression Model for Detecting Suspected Thyroid Cancer Based on Ultrasound Data

© Oyuna B. Tsydyпова

*Irkutsk National Research Technical University,
Irkutsk, Russian Federation*

Abstract. Information systems and technologies are actively used in medicine and improve the quality of medical care. The introduction of information systems for storing medical research data makes it possible to obtain a database on the basis of which it is possible to build machine learning models and use them to solve problems of medical diagnostics. This paper describes a model of binary logistical regression identified from the available data to detect suspicion of thyroid malignancy, according to ultrasound data. The article also discusses an information system created by the author that implements this model and allows you to automate a doctor's work when performing a thyroid ultrasound. The implementation of the system will allow you not only to reduce the time spent by a doctor on filling out the study protocol, but also to create a database with the ability to track the dynamics of diseases and accumulate statistics, which will make it possible to use other machine learning methods in diagnosing thyroid diseases. Thus, a decision support tool has been created for a doctor when performing an ultrasound examination of the thyroid gland.

Keywords: artificial intelligence, data analysis, decision support system, logistic regression

Щитовидная железа (ЩЖ) играет важную роль в организме человека: она отвечает за умственное, психическое, физическое развитие, а также за состояние здоровья в целом. В связи с тем, что Иркутская область относится к так называемым йододефицитным регионам, она является одним из неблагоприятных регионов по заболеваниям щитовидной железы [1].

Любой сбой в работе этого органа может привести к возникновению тяжелых

заболеваний, вплоть до злокачественного новообразования [2]. Этот диагноз может быть установлен по результатам целого комплекса исследований, однако выявить подозрение на злокачественное новообразование ЩЖ и необходимость дообследования позволяет ультразвуковое исследование (УЗИ). УЗИ занимает ведущее место в диагностике узловых образований ЩЖ в связи с его доступностью, высокой разрешающей способностью, информативностью.

Совокупность ультразвуковых признаков при исследовании узловых образований ЩЖ позволяет выявить высокий риск злокачественного процесса. УЗИ дает возможность обнаружить рак в 85,5 % случаев [3, 4].

В настоящее время информационные системы применяются во многих сферах жизни нашего общества, однако их использование в задачах медицинской диагностики, в том числе при регистрации данных УЗИ, по крайней мере в нашем регионе встречается довольно редко. В г. Иркутске многие медицинские учреждения (МУ) для регистрации данных УЗИ ЩЖ используют бумажные носители или шаблоны в MS Word и MS Excel. Очевидны недостатки такого способа регистрации данных:

1. сложно хранить и находить информацию о приемах;
2. отсутствует централизованное хранилище результатов исследований;
3. невозможно анализировать данные, следовательно, сложно отслеживать динамику показателей ЩЖ (например, для отслеживания зарождающихся злокачественных новообразований);
4. трудно составлять отчеты по исследованиям.

Разработка и внедрение информационной системы для регистрации и хранения результатов УЗИ ЩЖ позволят устранить указанные недостатки:

1. обеспечит удобное хранение и быстрый доступ к информации;
2. позволит ускорить формирование протокола исследования и, как следствие, уменьшить время, затрачиваемое на «бумажную работу» врача;
3. создаст централизованное хранилище результатов исследований, в котором будет накапливаться статистика и динамика показателей при различных заболеваниях ЩЖ, что даст возможность расширить применение методов искусственного интеллекта (ИИ) и анализа данных в задачах медицинской диагностики.

Последний пункт особенно важен в связи с тем, что направление медицины считается одним из стратегических и перспективных с точки зрения эффективного внедрения ИИ, используемого в качестве поддержки принятия решения. ИИ обрабатывает огромные массивы данных с целью подготовки прогнозов и таким образом помогает повысить эффективность МУ, а также облегчить работу врачей [5, 6].

Систем, позволяющих регистрировать данные УЗИ щитовидной железы, крайне мало. При этом ни одна из тех, которые были рассмотрены, не обеспечивает поддержку принятия решений, которая позволила бы осуществить поддержку принятия решений с помощью ИИ [7–9].

В связи с этим актуальной является задача создания информационной системы (ИС) для работы с данными УЗИ ЩЖ и с поддержкой принятия решений о необходимости дообследования на предмет наличия злокачественного новообразования ЩЖ.

Для решения задачи разработана система регистрации и учета данных УЗИ ЩЖ с применением метода логистической регрессии для оценки необходимости дообследования на предмет наличия злокачественного новообразования. Данная система формирует рекомендацию для врача на основе 27 признаков ЩЖ о необходимости дообследования.

Для проектирования и реализации системы использовались следующие инструментальные средства:

- инструмент моделирования бизнес-процессов Ramus,
- инструмент для построения UML-диаграмм StarUML,
- инструмент, позволяющий создавать блок-схемы ClickCharts,
- графический редактор MS Visio для создания макетов,
- инструмент Draw.io для проектирования модели данных,
- СУБД Microsoft SQL Server Developer,
- среда разработки Visual Studio.

Система реализована с использованием языков программирования C# и Python.

Для хранения данных обследований сформирована база данных. Модель данных на логическом уровне представлена далее (рис. 1).

Для проведения УЗИ ЩЖ необходимо обеспечить хранение информации о медицинских учреждениях (МУ), аппаратах, работниках, пациентах, исследованиях. Сущность «Персонал МУ» связывает работника с конкретным МУ. Сущность «Должность» определяет должность работника. Сущность «Врач – медсестра» позволяет разграничить права доступа. Сущность «Значения показателей» хранит значения показателей исследования.

Для обеспечения поддержки принятия решений в системе построена модель логистической регрессии [10, 11].

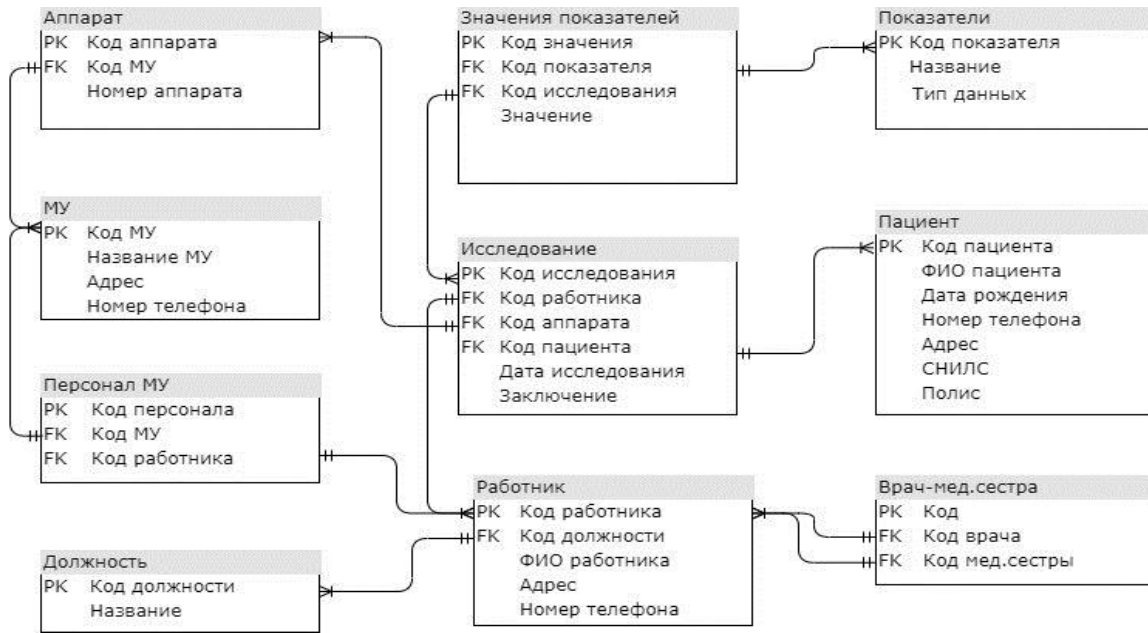


Рис. 1. Модель данных на логическом уровне

Обучение модели предполагает нахождение по имеющейся выборке коэффициентов a_0, a_1, \dots, a_m в выражении

$$Z = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m. \quad (1).$$

Значение логистической функции $F(z) = \frac{1}{1+e^{-z}}$ определяет вероятность принадлежности наблюдений к первому или второму подмножеству наблюдений.

Для обучения модели логистической регрессии использовалась выборка из 995 наблюдений результатов УЗИ ЩЖ. В каждом наблюдении оценивались 27 показателей: размеры, экзогенность, контур, наличие очаговых изменений и прочее. Все наблюдения разделены на 2 подмножества: на потенциально неопасные (значение введенного признака равно 0) и опасные, которые требуют дообследования (значение введенного признака равно 1).

Для принятия решения о необходимости дообследования используется библиотека машинного обучения scikit-learn с языком программирования Python. Обучение модели логистической регрессии выполнено с помощью метода `model.fit(X_train, y)`. Значения коэффициентов в модели логистической регрессии, полученные в результате обучения по имеющимся данным, приведены ниже (табл.).

Для каждого нового наблюдения, для случая бинарной логистической регрессии система возвращает одномерный массив из двух элементов, где первый элемент – вероятность принадлежности наблюдения к подмножеству наблюдений, для которых потенциальная опасность отсутствует, а второй элемент – вероятность принадлежности наблюдения к подмножеству потенциально опасных и требующих дообследования. Для каждого нового наблюдения, для случая бинарной логистической регрессии система возвращает одномерный массив из двух элементов, где первый элемент – вероятность принадлежности наблюдения к подмножеству наблюдений, для которых потенциальная опасность отсутствует, а второй элемент – вероятность принадлежности наблюдения к подмножеству потенциально опасных и требующих дообследования.

Когда значение второго элемента массива больше чем 0,1, значение атрибута «Оценка необходимости дообследования» равно 1, в противном случае оно равно 0. Если использовать полученную модель, то можно классифицировать новые наблюдения к одному из двух подмножеств, за счет чего осуществляется поддержка принятия решения о необходимости дообследования на предмет злокачественного поражения ЩЖ.

Главное окно системы представлено ниже (рис. 2). На нем располагаются меню, фильтры и таблица с исследованиями.

Значение коэффициента для каждого признака

| № п/п | Признак | Значение коэффициента |
|-------|-------------------------|-----------------------|
| 1 | Возраст | 0.38 |
| 2 | Длина правой доли | 0.09 |
| 3 | Ширина правой доли | -0.79 |
| 4 | Высота правой доли | -0.11 |
| 5 | Объем правой доли | 0.15 |
| 6 | Длина левой доли | 0.002 |
| 7 | Ширина левой доли | -0.19 |
| 8 | Высота левой доли | 0.23 |
| 9 | Объем левой доли | 0.20 |
| 10 | Толщина перешейка | 0.05 |
| 11 | Общий объем железы | 0.19 |
| 12 | Расположение | 0.22 |
| 13 | Эхогенность | 1.85 |
| 14 | Структура | 0.01 |
| 15 | Контур | -1.74 |
| 16 | Очаговые изменения (ОИ) | 0.84 |
| 17 | Расположение ОИ | 0.11 |
| 18 | Длина ОИ | -0.19 |
| 19 | Ширина ОИ | -0.38 |
| 20 | Высота ОИ | 0.44 |
| 21 | Структура ОИ | -0.09 |
| 22 | Контур ОИ | 0.15 |
| 23 | Кальцинат | 0.68 |
| 24 | Фокальные изменения | -0.03 |
| 25 | Кровоснабжение в узле | 0.44 |
| 26 | Кровоток в паренхиме | 0.68 |
| 27 | Заключение | -0.01 |

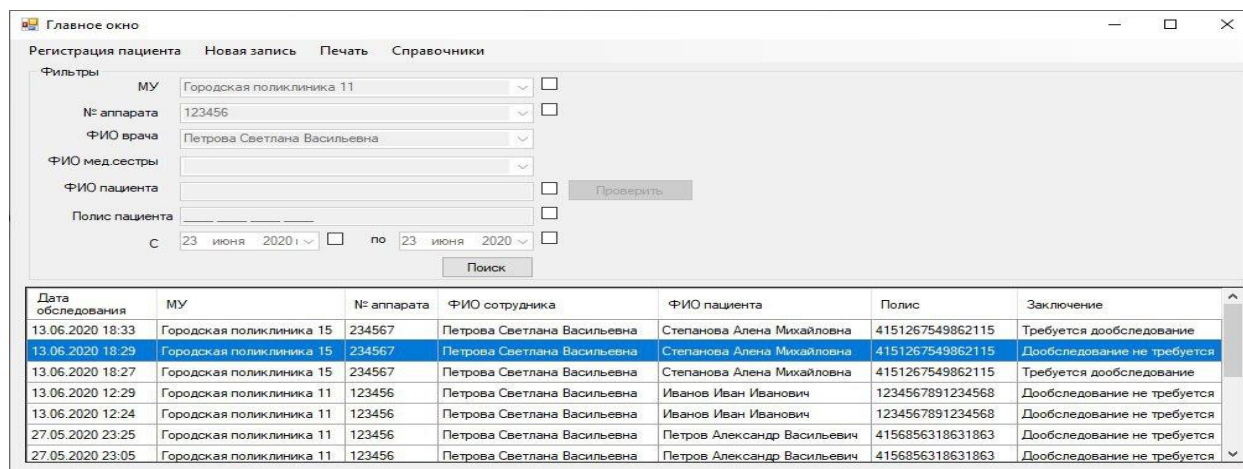


Рис. 2. Главное окно системы

Окно добавления новой записи исследования изображено далее (рис. 3).

После заполнения показателей ЦЖ нужно рассчитать оценку необходимости дообследования в нижней части окна. На основании значения полученных коэффициентов формируется рекомендация. Так как система не отвечает за итоговое заключение, а только рекомендует его, врач формирует заключение сам.

Также система может выполнять следующие функции:

- регистрацию новых пациентов,
- изменение данных исследования,
- поиск исследований по медицинским учреждениям, № аппарата, ФИО пациента, полису пациента, дате исследования,
- просмотр проведенных исследований,
- формирование протоколов,
- экспорт протокола в MS Word,
- формирование отчета,
- экспорт отчета в MS Excel.

Новая запись

МУ:

№ аппарата:

ФИО врача:

ФИО мед.сестры:

ФИО пациента:

Дата обследования:

Показатели ЩЖ

| | Левая доля | Правая доля |
|----------|---|--|
| Длина | <input type="text" value="1.4"/> см | <input type="text" value="1.4"/> см |
| Ширина | <input type="text" value="1.2"/> см | <input type="text" value="0.9"/> см |
| Высота | <input type="text" value="4.2"/> см | <input type="text" value="4.3"/> см |
| Объем | <input type="text" value="3.38"/> см ³ | <input type="text" value="2.6"/> см ³ |
| Операция | <input type="text" value="Нет"/> | <input type="text" value="Нет"/> |

Очаговые изменения

Толщина перешейки: мм

Операция на перешейке:

Объем ЩЖ: см³

Расположение:

Эхогенность:

Структура:

Контур:

Кровоток:

Лимфоузлы:

Рассчитать

Рекомендация:
Требуется дообследование.
Коэффициенты:
p0 = 0,68277,
p1 = 0,31723

Печать

Заключение:

Добавить

Отменить

Рис. 3. Окно добавления новой записи

Результатом работы является система регистрации и учета данных УЗИ ЩЖ, в которой применен метод логистической регрессии для оценки необходимости дообследования на предмет злокачественного новообразования ЩЖ.

Использование разработанной системы позволит:

- повысить качество оказания медицинских услуг благодаря применению поддержки принятия врачебных решений,
- реализовать безбумажную технологию с сохранением возможности получения копии документов протокола и отчета,
- организовать структурированное хранение информации о медучреждении,

персонале, пациентах, обследованиях, а также обеспечить быстрый доступ к ним,

- сократить временные затраты на ведение журнала УЗИ ЩЖ.

Результатом внедрения системы окажется база данных результатов исследований, где будет отражена динамика развития заболеваний. По мере накопления данных возможно привлечение других методов машинного обучения для диагностики заболеваний щитовидной железы, а также уточнение имеющихся моделей для повышения качества диагностики и предоставления врачу инструмента поддержки принятия решений.

Библиографический список

1. Такая маленькая и такая важная – щитовидная железа // Информационный сайт Иркутской области «Ирк.ру» [Электронный ресурс]. URL:

<http://www.irk.ru/news/articles/20160615/thyroid/> (30.06.2020).

2. Основные симптомы нарушения функции щитовидной железы // Медицинский портал Schitovidka03.ru [Электронный ресурс]. URL:

<https://schitovidka03.ru/bolezni/narushenie-funkcii> (30.06.2020).

3. Ракитина Д.А., Беляев А.М., Раджабова З.А., Костромина Е.В., Котов М.А. Оптимизация диагностических мероприятий и лечебной тактики при распространенном раке щитовидной железы // Современные проблемы науки и образования. 2016. № 6. С. 3–4. [Электронный ресурс]. URL: <https://science-education.ru/ru/article/view?id=25875> (01.07.2020).

4. Котляров П.М., Канорская Г.А., Соловьева С.В. Рак щитовидной железы – значимость вероятностной диагностики по данным ультразвукового исследования // Вестник РНЦРР МЗ РФ. 2004. № 3. С. 1. [Электронный ресурс]. URL: http://vestnik.rncrr.ru/vestnik/v3/papers/kotl_v3.htm (01.07.2020).

5. Гусев А.В., Добридюк С.Л. Искусственный интеллект в медицине и здравоохранении // Информационное общество. 2017. № 4-5. С. 78–93. [Электронный ресурс]. URL: <http://emag.iis.ru/arc/infosoc/emag.nsf/BPA/0d74c2999f2435374425829c0028d93b> (02.07.2020).

6. Поряева Е.П., Евстафьева В.А. Искусственный интеллект в медицине // Вестник науки и образования. 2019. № 6-2 (60). С. 15–19. [Электронный ресурс]. URL:

<https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-medicine-1/viewer> (02.07.2020).

7. Смирнов В.В., Беляев М.В. Программа «Редактор протоколов УЗИ» Версия 1.0 [Электронный ресурс]. URL: <http://pantherman.narod.ru/versions.html> (06.07.2020).

8. Митюшин А.А. Компьютерная программа «Протоколирование и анализ результатов клинических ультразвуковых исследований» [Электронный ресурс]. URL: <http://программа-узи.рф/> (06.07.2020).

9. Программное обеспечение «Протокол.УЗИ» // ООО «БиоМедТех» [Электронный ресурс]. URL: <https://protocol-uzi.ru/index.php/protokol-uzi> (06.07.2020).

10. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989. 608 с. [Электронный ресурс]. URL: <https://obuchalka.org/2015030282984/prikladnaya-statistika-klassifikacii-i-snijenie-razmernosti-aivazyan-s-a-buhstaber-v-m-enukov-i-s-meshalkin-l-d-1989.html> (06.07.2020).

11. Hosmer D.W., Lemeshow S. Applied Logistic Regression. New York, 2002. 392 p.

Сведения об авторе / Information about the Author

Цыдыпова Оюна Батоевна, студентка группы ИСТБ-16-1, Институт информационных технологий и анализа данных, Иркутский национальный исследовательский технический университет, 664074, г. Иркутск, ул. Лермонтова, 83, Российская Федерация, e-mail: batoevnaa@mail.ru

Oyuna B. Tsydyпова, Student, Institute of Information Technology and Data Analysis, Irkutsk National Research Technical University, 83 Lermontov Str., Irkutsk, 664074, Russian Federation, e-mail: batoevnaa@mail.ru