

Кудрейко А.А.^{1,2}, Гильмиярова М.Н.²

РАЗРАБОТКА АЛГОРИТМА АНАЛИЗА МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ ORANGE DATA MINING

¹Уфимский университет науки и технологий, г. Уфа²Башкирский государственный медицинский университет, г. Уфа

В данной работе разработан алгоритм интеллектуального анализа медицинских изображений с помощью программы Orange Data Mining. Выполнено сравнение работы алгоритмов классификации изображений.

Ключевые слова: Orange Data Mining, интеллектуальный анализ медицинских изображений, пневмония, туберкулез.

Разработка интеллектуальных методов анализа и технологий для медицинских изображений стала областью, представляющей значительный интерес. Об этом ясно свидетельствует динамика роста публикаций по анализу медицинских изображений (см. Рис. 1. Динамика роста публикаций с ключевыми словами "medical images" и "artificial intelligence", полученная с помощью базы данных Scopus (дата обращения: 22.01.2024).

). Примечательно, что резкий рост интереса исследователей к данной теме совпал с началом пандемии COVID-19, поскольку во всех странах наблюдался дефицит квалифицированного медицинского персонала. За это время разработаны модели с хорошими показателями точности для применения в клинической практике врачей-рентгенологов.

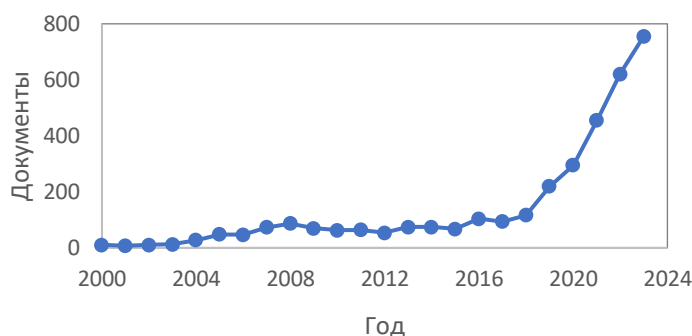


Рис. 1. Динамика роста публикаций с ключевыми словами "medical images" и "artificial intelligence", полученная с помощью базы данных Scopus (дата обращения: 22.01.2024).

Среди наиболее применяемых методик ранней диагностики заболеваний легких наиболее точной является лучевая диагностика. Соответствующие изображения и их интерпретации являются преобладающими инструментами в постановке диагноза. Однако фундаментальный недостаток лучевой диагностики связан с интегральным характером получаемого изображения, что ведет к потере информации по ее глубине. Это вызывает необходимость в получении и анализе множества проекций [1]. Отсюда возникает следующая ситуация: улучшение методов, средств и способов диагностики не всегда приводит к улучшению качества диагностики. Следовательно, решение ряда задач распознавания

медицинских изображений, их классификация по категориям и внедрение в клинические методы позволит разрешить накопленные противоречия.

Цель работы

Целью данной работы является поиск новых подходов, методов и алгоритмов классификации КТ и рентген изображений с диагнозом пневмония и туберкулез в условиях априорной неопределенности координат патологии и относительных уровней яркости объектов на фоне пространственно-неоднородного изображения.

Для достижения поставленной цели были решены следующие **задачи**:

- поиск и подготовка медицинских изображений, необходимых для обучения алгоритмов;
- анализ и подбор подходящих алгоритмов предобработки и классификации изображений;
- разработка, обучение и определение на тестовой выборке ее диагностической информативности;
- анализ результатов исследования.

Решение поставленных задач выполнено в программе Orange Data Mining. Orange Data Mining (ODM) – интерактивная среда визуального программирования и интеллектуального анализа данных [2, 3].

Материал и методы

Объектом исследования являются открытые базы данных КТ снимков легких Chest X-Ray Images (Pneumonia) и Tuberculosis (TB) Chest X-ray Database, загруженные с платформы Kaggle [4, 5]. В работе мы проанализировали набор из 300 снимков, который содержит три типа изображений: норма, пневмония и туберкулез легких. Перед загрузкой совокупности всех изображений в наш алгоритм, они были обработаны на подавление шума.

Алгоритм распознавания медицинских изображений

Для сегментации легочной ткани применили алгоритм, использующий метод каскадных окон. После выделения всех объектов, соответствующих диапазону плотностей, результатом сегментации было множество выделенных объектов – кандидатов в патологию. Поскольку сегментация основана на показателях шкалы Хаунсфилда, в поле зрения могли попадать объекты, совпадающие по рентгеновской плотности с признаком «пневмония», но являющиеся при этом анатомическими объектами (например, участками сосудов). Для выделения сосудов и других плотных объектов в массиве сегментированной легочной ткани использовался так называемый алгоритм заливки (flood fill): многоуровневая процедура порогового определения на каждом слайсе [6]. Совокупность вычисленных метрических

расстояний позволяет представить результаты анализа в том или ином виде. В нашем алгоритме это реализовано в виджетах «Image grid», «Distance map» и «Hierarchical clustering».

Результаты и обсуждение

Результаты анализа рентгеновских изображений

Для визуализации диапазона расстояний, представленных в матрице, обычно используется цветовая палитра. Ниже мы визуализировали наиболее коррелирующие атрибуты (расстояния по столбцам) в наборе данных о легочных заболеваниях (Рис. 2). Можно заметить, что в нашей выборке изображений существует три кластера данных: пневмония, туберкулез и норма, каждый из которых разбивается на дополнительные кластеры.

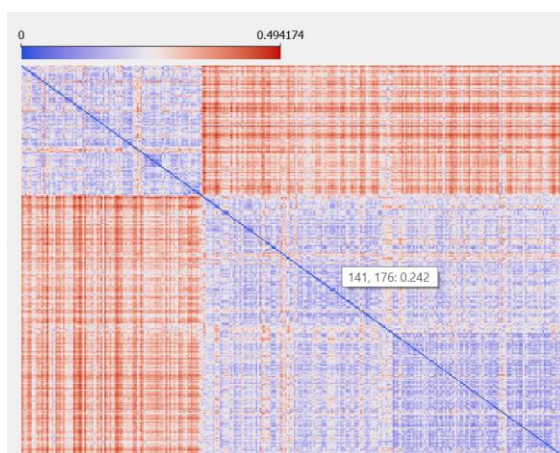


Рис. 2. Визуализация кластерного анализа изображений с двумя разными диагнозами и нормой.

На сетке изображений - Image Grid (Рисунок 3) четко и отдельно выделяется кластер изображений с туберкулезом.

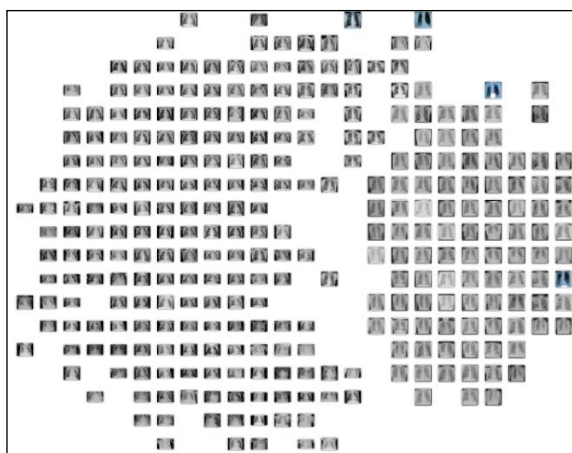


Рис. 3. Фрагмент результата работы алгоритма сравнения изображений при поиске сходств и различий.

		Predicted			Σ
		1. Туберкулез	2. Пневмония	NORM	
Actual	1. Туберкулез	99	1	0	100
	2. Пневмония	1	88	11	100
	NORM	2	11	87	100
Σ		102	100	98	300

Рис. 4. Матрица ошибок.

Матрица ошибок - Confusion Matrix (Рисунок 3) показывает следующее: 99 снимков в отношении пациентов с туберкулезом определены точно. Это означает, что из 100 пациентов, которые были классифицированы как имеющие заболевание туберкулез, 99 действительно имеют это заболевание. Из 100 снимков пациентов с пневмонией, которые были классифицированы как имеющие это заболевание, 88 действительно имеют это заболевание. Из 100 снимков с нормой 87 определены точно. В следствии того, что человеком могут быть допущены ошибки в интерпретации медицинских изображений, то основными причинами этих ошибок чаще всего бывают небольшие размеры патологий. В случае применения искусственного интеллекта причинами ошибок могут быть другие причины, например, прилегание к средостению сосуду или плевре, слияние с ателектазированной легочной тканью, расположение узелков на фоне диффузного уплотнения легочной ткани. Таким образом, обучение алгоритма на большем количестве данных, включающем атипичные случаи, сможет сделать его диагностическую способность сопоставимой с группой врачей. Однако отсутствие статистически значимого различия не позволяет говорить об использовании алгоритма без врача [6].

Модель ошибочно классифицировала одного пациента с туберкулёзом как имеющего другое заболевание, хотя он на самом деле не имел его; из больных с пневмонией одного пациента определила как больного с туберкулезом и 11 без заболевания. И среди снимков с нормой - 11 как пневмония и 1 как туберкулез.

Таким образом, развитие алгоритмов искусственного интеллекта в области диагностики легочных заболеваний достигло высоких результатов. Некоторые алгоритмы не уступают показателям врачей и могут использоваться для первичного прочтения, однако для этого требуется соответствующая нормативно-правовая основа [7]. Такие показатели достигнуты благодаря тому, что мы намеренно выделили 3 класса изображений. Если бы мы выделили различные виды туберкулеза, пневмонии и добавили другие патологии, то искусственный интеллект проявил бы себя иначе. Например, при дифференциальной

диагностике различных видов пневмоний искусственный интеллект сталкивается с рядом проблем и ограничений, связанных с рентгенологическими симптомами пневмонии, необходимостью анализа клинической картины, боковых проекций и широким диапазоном заболеваний.

Заключение и выводы

Разработанная модель искусственного интеллекта для поддержки принятия решений по классификации КТ и рентгеновских снимков, экспериментальная апробация которой проведена на контрольных выборках по классификации медицинских изображений по классам «пневмония», «туберкулез» и «норма» показали высокую диагностическую эффективность: около 90%.

Проведена экспериментальная апробация программного обеспечения интеллектуальной системы по классификации заболеваний по классам «пневмония» и «туберкулез». Матрица ошибок показала, что 99 изображений из 100 с меткой «туберкулез» определены точно. Из 100 снимков пациентов с пневмонией, правильно классифицированы 88.

Применяя этот метод к медицинским изображениям, медицинские работники могут повысить точность диагностики, оптимизировать планы лечения и в конечном итоге улучшить результаты пациентов. Кроме того, понимание, полученное в результате этого анализа, может способствовать разработке новых технологий медицинской визуализации и достижений в области радиологии.

ЛИТЕРАТУРА

1. Т. Г. Авачева, М.К. Алмазова, Ю.В. Зубцова. Применение искусственного интеллекта в радиологии. in Всероссийская научно-практическая конференция с международным участием «Информационный обмен в междисциплинарных исследованиях», 2022 года. Рязанский государственный радиотехнический университет им. В.Ф. Уткина.
2. Demsar J, C.T., Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan B. 2013; Available from: <https://orangedatamining.com/>.
3. Юсупов, Н., А. Савельева, and О. Леонова, Исследование методов классификации в программе Orange. Молодежная школа-семинар по проблемам управления в технических системах имени АА Вавилова, 2020. 1: p. 27-30.
4. MOONEY, P. Chest X-Ray Images (Pneumonia). Available from: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>.
5. RAHMAN, T. Tuberculosis (TB) Chest X-ray Database. Available from: <https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset>.
6. Мелдо, А.А. Разработка и внедрение системы искусственного интеллекта в лучевой диагностике очаговых образований в легких. 2021; Available from: <https://search.rsl.ru/ru/search#q=%D0%BC%D0%B5%D0%BB%D0%B4%D0%BE>.

7. С. П. Морозов, Д.Ю.К., Н. А. Павлов, Ю. С. Кирпичев, В. А. Гомболевский, А. Е. Андрейченко, Клинические аспекты применения искусственного интеллекта для интерпретации рентгенограмм органов грудной клетки. Туберкулёз и болезни лёгких, 2021.