

**Региональная конференция по научно-техническому творчеству
школьников «Лабиринты науки»**

государственное общеобразовательное учреждение

Ярославской области “Лицей № 86”

Творческий проект

**Разработка искусственного интеллекта для
выставления предварительного диагноза на основе
общего анализа крови**

Выполнил ученик 9Н класса

ГОУ ЯО “Лицей №86”

Фадеев Макар Сергеевич

Научный руководитель:

Скрыгитель Алексей Игоревич

Ярославль,

2024

Содержание

Введение	3
1. Анализ существующих решений проблемы	3
2. Сбор и подготовка данных для обучения модели.....	4
3. Разработка и обучение модели	5
4. Валидация и тестирование.....	6
Вывод	6
Список использованной литературы	7

Введение

Развитие современной медицины напрямую зависит от развития информационных технологий. Продвинутое оборудование позволяет врачам оказывать не только скорую, но и качественную помощь пациенту. Традиционные методы оценки результатов могут быть подвержены субъективности и ошибкам, что подчеркивает необходимость разработки автоматизированных систем. Данный проект направлен на создание модели искусственного интеллекта, которая будет использовать алгоритмы машинного обучения для выставления предварительного диагноза на основе результатов общего анализа крови.

Проблема: Высокая роль человеческого фактора в интерпретации результатов общего анализа крови.

Цель: Разработка модели ИИ, способной автоматически анализировать результаты общего анализа крови и предоставлять предварительный диагноз.

Задачи:

1. Анализ существующих решений проблемы
2. Сбор и подготовка данных для обучения модели
3. Разработка модели
4. Валидация и тестирование

1. Анализ существующих решений проблемы

Перед выполнением поставленной задачи важно оценить уже существующие решения проблемы, найти их недостатки и исправить их в своей работе. Мной было проведено исследование, в результате которого я выявил три основных решения проблемы.

- 1. Поиск объяснения анализа онлайн.** Простой и доступный способ узнать значение своего анализа с использованием одного из множества сайтов. Главной проблемой такого решения является крайне неточная оценка состояния здоровья пациента. Это связано с непрактичным методом оценки анализа в целом, ведь за раз можно проанализировать только один параметр общего анализа

крови. Значит, наша разработка должна иметь возможность оценивать сразу все элементы крови в совокупности.

2. Консультация квалифицированным персоналом. Один из самых распространенных и практичных способов поставления диагноза на данный момент. Отличается трудоемкостью, ведь требует анализа и выставления оценки от обученного сотрудника. Главной проблемой является возможность ошибки со стороны человека, ставящего диагноз, в связи с человеческим фактором. Наш продукт будет оценивать состояние здоровья пациента полагаясь исключительно на машинный код, что делает шанс ошибки практически нулевым.

3. Использование готовых систем поддержки принятия решений. Продвинутый способ решения проблемы, эффективен, быстр, но не сильно популярен. К недостаткам можно добавить неполноценность тех СППР, что находятся в открытом доступе (открытые решения способны принимать себя лишь до 8 параметров, когда на самом деле их 15). Возьмем на заметку то, что стоит анализировать как можно больше элементов крови сразу.

Подытожим исследование, собрав техническое задание. Разработка должна ставить предварительный диагноз пациенту на основе общего анализа крови, причем оценивать состояние здоровья необходимо именно по совокупности абсолютно всех показателей анализа, а также дополнительно проверять свои диагнозы, чтобы исключить возможность ошибки со стороны алгоритма.

2. Сбор и подготовка данных для обучения модели

Теоретический анализ проблемы показал, что большая часть существующих вариантов решения проблемы упирается именно в необъективность анализа, связанную с недостатком входных данных. В таком случае, мы должны найти базы данных с обезличенными общими анализами крови и диагнозами, соответствующими этим анализам.

Главная проблема заключается в практически полном отсутствии данных в открытом доступе, что становится проблемой. Мной было найдено несколько статей, в которых рассказывалось о процессе борьбы с лейкемией и анемией, в которых были приведены в пример результаты общего анализа

крови у больных. Кроме того, попросив помощи у знакомых своих родственников, мне удалось получить несколько анализов людей, страдающих от гепатита и диабета, а также просто здоровых людей. С помощью распределения Гаусса я расширил свою базу данных для тренировки нейросети, в итоге научившись различать пять состояний здоровья у человека. Важно отметить, что в перспективе расширять список определяемых болезней можно до бесконечности, просто скармливая ИИ новые данные, но в этом проекте мне пришлось ограничиться такой выборкой из-за труднодоступности данных.

3. Разработка и обучение модели

Было решено использовать язык программирования Python в связи с его популярностью в сфере машинного обучения, наличия множество упрощающих библиотек и простотой в освоении. Алгоритм работает на основе tensorflow - библиотеки, содержащей в себе базовые инструменты для разработки нейросетей, а также numpy – библиотеки, позволяющей проводить продвинутые математические вычисления и оптимизирующей некоторые аспекты стандартного Python. Нейросеть состоит из входного слоя с методом активации ReLU, трех скрытых слоев с методами активации ReLU, ELU и tanh, и выходного слоя с методом softmax. Такое количество слоев было выбрано не случайно – результат работы выходного слоя softmax напрямую зависит от точности подаваемых в него данных, а в связи с мелкими различиями между анализами крови разных диагнозов большое число скрытых слоев обеспечивает точность выводимых данных даже в том случае, если мы имеем выборку из более чем сотни различных заболеваний.

Так как данные для тренировки мы подготовили заранее, нам остается только обучить нейросеть с их помощью, отрегулировав количество циклов обучения и их глубину. В результате, мы получили файл с весами для нейронов, при использовании которого нейросеть способна определить тип болезни на основе анализа крови

4. Валидация и тестирование

Для получения наилучших результатов важно обучить нейросеть несколько раз с разными параметрами обучения, чтобы в будущем получать наиточнейшие результаты. Если циклов будет слишком мало, то нейросети просто не хватит времени на выявление паттернов и закономерностей, но и слишком много циклов ставить нельзя, ведь в таком случае нейросеть может оказаться просто неспособной работать с данными, отличающимися от тренировочных. К сожалению, точной стратегии для нахождения лучшей глубины и количества циклов не существует, поэтому приходится вручную перебирать комбинации для нахождения оптимальной обучающей связки.

Таким образом, за день подобрав лучшую комбинацию, я все-таки натренировал нейросеть и опробовал её на нескольких реальных примерах. Результат был отличный: все десять поданных случаев нейросеть правильно распределила по классам болезней, что означает её превосходную функциональность.

Вывод

В результате работы был создан алгоритм, способный по общему анализу крови определить, болен ли человек лейкемией, анемией, диабетом или гепатитом. Такой проект не имеет известных аналогов, а также способен расширить число распознаваемых болезней в перспективе с учетом, что в него будут добавляться новые данные пациентов с различными заболеваниями. Системы поддержки принятия решений делают повседневную жизнь человека проще, автоматизируя механическую работу и исключая шанс человеческой ошибки.

Список использованной литературы

1. Душкин Р. В. Искусственный интеллект. – М.: ДМК Пресс, 2019. – 280 с., ISBN 978-5-97060-787-9. – Текст: непосредственный Рассел, Стюарт, Норвинг, Питер. Искусственный интеллект:
2. Современный подход, 2-е изд.: Пер. с англ. — М.: Издательский дом «Вильямс», 2006. — 1408 с.: ил. — Парал. тит. англ. ISBN 5-8459-0887-6 (рус.) – Текст: непосредственный
3. Datasets for machine learning. – Текст: электронный // Kaggle: официальный сайт – 2024 – URL: <https://www.kaggle.com/datasets?tags=13302-Classification> (дата обращения: 25.10.2024)
4. Выбор слоя активации в нейронных сетях: как правильно выбрать для вашей задачи. – Текст: электронный // Habr: официальный сайт. – 2024 – URL: <https://habr.com/ru/articles/727506/> (дата обращения: 10.09.2024)
5. Общий анализ крови (ОАК). – Текст: электронный // Гемотест: официальный сайт. – 2024 – URL: <https://gemotest.ru/info/spravochnik/analizy/obshchiy-analiz-krovi/> (дата обращения: 25.10.2024)