

Решение задачи классификации COVID-19 по МРТ снимкам с использованием нейронных сетей

А.С. Мирошниченко, email: 963565@bsu.edu.ru¹

В.М. Михелев, email: mikhelev@bsu.edu.ru²

Белгородский государственный национальный исследовательский университет

***Аннотация.** В данной статье рассматривается задача классификации COVID-19 по снимкам МРТ с применением технологий сверточных нейронных сетей. Решение данной задачи основывается на использовании faster R-CNN как основного метода поиска патологий на снимках МРТ, далее применение классификатора Хсертпюна.*

***Ключевые слова:** COVID-19, классификация изображений, машинное обучение, нейронные сети.*

Введение

COVID-19 это новое заболевание, которое уже было объявлено всемирной пандемией. Уже более двухсот стран пострадали от этого серьезного вируса, так как данный вирус распространяется в результате взаимодействия с людьми. Более того, важно отметить что симптомы данного вируса очень похожи на симптомы обычного сезонного гриппа. Отбор инфицированных пациентов считается важнейшим шагом в борьбе с COVID-19. Поскольку отсутствуют отличительные инструменты для выявления случаев заболевания COVID-19, потребность в дополнительных диагностических инструментах возросла. Поэтому очень важно, как можно раньше распознавать положительные случаи, чтобы избежать дальнейшего распространения этой эпидемии. Однако, существует ряд методов выявления пациентов с положительным диагнозом COVID-19, которые обычно выполняются на основе респираторных проб, и среди них важнейшим подходом к лечению является радиологическая визуализация или рентгеновская визуализация. Недавние результаты, полученные при помощи методов рентгеновской визуализации, предполагают, что такие изображения содержат важную информацию о вирусе SARS-CoV-2. Применение технологий глубоких сверточных нейронных сетей в сочетании с рентгеновской визуализацией может помочь в точной идентификации этого заболевания, а также может помочь в идентификации этого заболевания и в преодолении проблемы нехватки обученных врачей в

отдаленных областях. В данной статье представлен метод решения классификации на основе структуры сети Xception [1] с применением faster R-CNN [2] для обнаружения пациентов с COVID-19 по снимкам МРТ грудной клетки.

1. Архитектура модели

Сверточные нейронные сети (CNN) – это архитектура обучения на основе глубоких нейронных сетей для обработки большого количества данных. В настоящее время широко применяется в области анализа медицинских изображений. CNN широко используется по сравнению с другими методами машинного обучения, потому что не требует ручного извлечения функций, а также не требует специальной сегментации. Ниже на рис. 1 показана архитектура CNN, которая состоит из нескольких блоков входного уровня, скрытого слоя (слой свертки, слой объединения), полностью связанного слоя и выходного слоя.

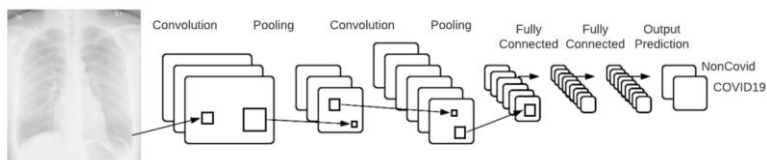


Рис. 1. Диаграмма архитектуры сверточной нейронной сети классификации МРТ изображений

В ходе решения задачи классификации МРТ снимков, вместо создания модели с нуля, было принято решение взять за основу Xception [1] с добавлением Faster R-CNN [2]. Разработанная структура состоит из 24 сверточных слоев, после которых следует два полностью связанных слоя и шесть слоев объединения. Данные слои являются типичными слоями CNN с разным количеством фильтров, размерами и значением шага. Ниже на рис. **Ошибка! Источник ссылки не найден.** показана архитектура Faster R-CNN.

Для использования классификатора в режиме реального времени, мы объединяем модель Xception [1] с Faster R-CNN. Данное объединение моделей обусловлено тем, что такая модель способна улавливать и извлекать небольшие отличия в медицинских изображениях.

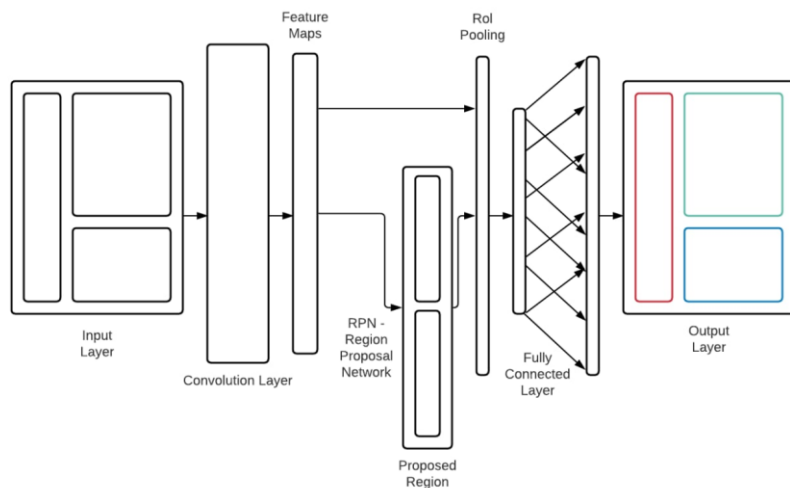


Рис. 2. Схематичная структура Faster R-CNN сверточной нейронной сети

2. Датасет

При подготовке набора данных использовалась двухэтапная процедура подготовки данных. В первую очередь использовались снимки МРТ пациентов с COVID-19, которые доступны как данные из открытых источников. Кроме того, был сформирован набор данных предназначенный для обучения и тестирования, который включал 5450 МРТ снимков грудной клетки для 2500 случаев пациентов. Для эксперимента, мы объединили и изменили два разных общедоступных набора данных: набор данных рентгеновского снимка грудной клетки COVID, подготовленный доктором Джозефом Коэном, научным сотрудником Монреальского университета [3], и RSNA набор данных задачи обнаружения от Kaggle [4]. Далее мы перешли к недавно доступному набору данных, специально разработанному для выявления случаев заболевания COVID-19 с использованием рентгеновских снимков грудной клетки, названного COVIDx [5]. Количество образцов изображений, которые мы использовали для обоих (пользовательских и COVIDx), набор данных показан в таблице (используемые данные в этой статье). Для пользовательского набора данных мы использовали 5450 образцов изображений, тогда как в наборе данных COVIDx был сформирован набор данных с 13800 изображениями. Набор данных

COVIDx постоянно обновляется изображениями, предоставленными исследователями из разных регионов. По состоянию на 7 декабря 2020 года имеется 183 рентгеновских снимка с диагнозом COVID-19, 8066 пациентов в норме и 5551 случай идентифицирован как не связанная с COVID пневмония. Объединив "нормальную" и "не-COVID-пневмонию" в один класс "не-COVID", мы разработали его как набор данных двоичного класса.

Таблица 1

Используемые данные в рамках решения задачи классификации снимков МРТ

Датасет	Тип	Кол-во	Non-COVID	COVID19	Всего
Сформированный	Train	1100	3000	80	4180
Сформированный	Test	300	950	20	1270
COVIDx	Train	7966	5451	152	13569
COVIDx	Test	100	100	31	231

3. Результаты и выводы

Показатели предложенной модели и результаты перекрестной проверки различных метрик классификатора и их средние результаты приведены в таблице (Результаты перекрестной проверки и метрики классификатора). Стоит отметить, что разработанная модель обеспечила среднюю точность 97,36% и получила среднее значение чувствительности, специфичности и F1-мера 97,75%, 95,18% и 98,46% соответственно. Он также обеспечивает значение точности около 97,46%.

Применение технологий глубоких сверточных нейронных сетей в сочетании с рентгеновской визуализацией может помочь в точной идентификации этого заболевания, а также может помочь в идентификации этого заболевания и в преодолении проблемы нехватки обученных врачей в отдаленных областях.

Таблица 2

Результаты перекрестной проверки и метрики классификатора

Номер теста	Чувствительность %	Специфичность %	F1-мера %	Точность %
1	98.50	96.77	98.99	98.27
2	99.50	99.00	99.75	99.57
3	97.00	93.55	97.98	96.54

4	99.50	93.55	99.25	98.70
5	98.50	99.00	99.24	98.70
6	97.50	93.55	98.24	96.97
7	97.50	96.77	98.23	96.97
8	96.50	90.32	97.47	96.67
9	96.50	99.00	98.22	96.97
10	96.50	90.32	97.22	95.24
Среднее	97.75	95.18	98.46	97.46

Работа выполнена при поддержке гранта РФФИ 19-07-00133_A.

Список литературы

1. A.S. Miroshnichenko, V.M. Mikhelev - Classification of COVID-19 patients from medical images using convolutional neural networks // «Information technologies in science, education and production» (ITSEP-2020). Collection of works of the VIII International Conference on Science and Technology (ISBN 978-5-9571-2993-6) Belgorod, 24-25 September, 2020.

2. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks arxiv:1506.01497

3. Cohen JP, Morrison P, Dao L. COVID-19 image data collection. 2020. p. 11597. arXiv:2003.

4. Radiological Society of North America. RSNA pneumonia detection challenge [электронный ресурс]. – Режим доступа: <https://www.kaggle.com/c/rsnapneumonia-detection-challenge/data>; 2019.

5. Wang L, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. 2020. arXiv preprint arXiv:2003.09871.