

ОБНАРУЖЕНИЕ РАКА ЛЁГКОГО ПО СНИМКАМ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

Сухов Н.Ю.

Белорусский государственный университет информатики и радиоэлектроники

г. Минск, Республика Беларусь

Калугина М.А. – к.ф.-м.н., доцент

В докладе рассматривается построение автоматизированной системы определения рака легкого человека, а также описывается инструмент, показывающий злокачественные образования на снимке.

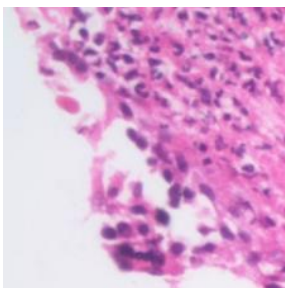
Рак легких является основной причиной смерти от рака в Соединенных Штатах и во всем мире. Прогрессирование заболевания и ответ на лечение при раке легких широко варьируется среди пациентов. Следовательно, точный диагноз имеет решающее значение при выборе и планировании лечения для каждого пациента с раком легких. Микроскопическое исследование тканей легкого остается важным этапом диагностики рака. Это требует, чтобы патолог распознал тонкие гистопатологические паттерны на очень сложных изображениях тканей. Этот процесс отнимает много времени и является субъективным. Также, чтобы поставить верный диагноз, требуются высококвалифицированные специалисты. Поэтому построение автоматизированной системы является крайне важной задачей. На сегодняшний день с применением алгоритмов машинного и глубинного обучения такие задачи решаются весьма успешно.

В имеющемся датасете имеется 15000 гистологических снимков лёгкого. 5000 из них представляют из себя доброкачественную ткань, а 10000 имеют злокачественные образования. Целью является построить бинарный классификатор, где класс 0 свидетельствует о здоровой ткани, а класс 1 - о раке. Классификатор был построен на основе сверточной нейронной сети, так как данный тип нейронных сетей за последних несколько лет себя очень хорошо зарекомендовал в задачах компьютерного зрения. Работа сверточной нейронной сети обычно интерпретируется как переход от конкретных особенностей изображения к более абстрактным деталям, и далее к еще более абстрактным деталям вплоть до выделения понятий высокого уровня. При этом сеть самонастраивается и вырабатывает сама необходимую иерархию абстрактных признаков (последовательности карт признаков), фильтруя маловажные детали и выделяя существенное. В данной работе использовалась собственно написанная нейронная сеть, которая представляет из себя чередующиеся сверточные слои с функцией активации RELU и слои пулинга (для уменьшения пространственной размерности).

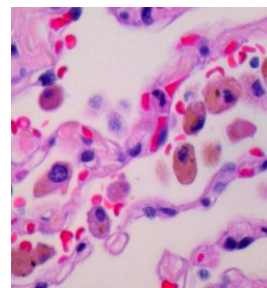
Сеть, описанная выше, показала очень хорошие результаты, однако приблизиться к результату врачей не удалось. Далее я стал использовать предобученную нейронную сеть. Обычно используют сети предобученные на датасете ImageNet, которая содержит более 15 миллионов изображений. Однако, в рамках данной задачи, использование этого датасета только ухудшает предыдущий алгоритм, так как он состоит из изображений котиков, собак, домов и т. д., а в нашем наборе очень специфические данные. Среди долгих поисков данных выбор пал на гистологические снимки толстой кишки. В этих данных есть общее с нашими данными для лёгких: много похожих клеток и структур.

Ещё в помощь для построения хорошей классификации пришёл алгоритм кластеризации KMeans (исследовались и другие алгоритмы кластеризации, например, DBSCAN, однако, особого прироста в качестве они не давали, а обучались очень долго, что также повлияло на время выдачи предсказаний всей системы на тестовых данных). Оказалось, что “прогнав” через кластеризацию снимки лёгких, у нас неплохо отделились кластера с различными клетками и межклеточное пространство. Обучаясь на кластеризованном датасете, снизилось немного переобучение, так как снимки могут приходиться в немного различной цветовой гамме. Финальное решение было выбрано

такое: предобученная сверточная нейронная сеть на снимках толстой кишки и дообучение ее на кластеризованных изображениях лёгких, “разморозив” три последних сверточных слоя. Также была использована функция потерь, которая сильнее штрафует за ложно-отрицательные результаты, так как поставить диагноз рака, когда его на самом деле нет, не так опасно, если сказать, что его нет, когда он есть. Это решение достигло качества примерно 97%, когда качество врачей составляет примерно 95 процентов. Таким образом, удалось построить систему, определяющую рак лёгких не хуже, чем врач-специалист.



Снимок толстой кишки



Снимок легкого

Также, чтобы врачу-специалисту можно было понять, почему алгоритм принял то или иное решение, была сделана подсветка областей, смотря на которые, классификатор принял решение, что изображение содержит злокачественное образование. Это было сделано с помощью сегментационной сети Deeolabv3+, которая на данный момент является лучшей сетью для сегментации изображений и в данной задаче также показала очень хорошие результаты.

Список использованных источников:

1. Artificial Intelligence in Lung Cancer Pathology Image Analysis [Электронный ресурс]. – Режим доступа: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6895901/>.
2. Классификация рака лёгких [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/post/445054/>.

Рис. 1 — Модель искусственного нейрона