

Форма рецензии статей в сборник трудов DLCP2023.

ФИО первого автора: Ledovskikh M.A.

Название работы: Recognition of skin lesions by images

ФИО рецензента: Mikhail A. Krinitskiy

1. Содержание работы (выбрать вариант из предложенных).

1.1 Соответствует ли статья тематике конференции

да

1.2. Отражает ли название тему работы.

да

1.3. Отражает ли аннотация основное содержание работы.

недостаточно четко

1.4. Сформулирована ли научно-техническая проблема, на решение которой направлена работа:

да

1.5. Представлено ли современное состояние этой проблемы (state of arts):

недостаточно четко

1.6. Сформулирована ли новизна предложенного подхода по сравнению с имеющимися и в чем его ожидаемое преимущество:

нет

1.7. Сформулирован ли результат работы, в частности, достигнут ли ожидаемый результат:

да

2. Оформление работы:

2.1. Английский язык:

удовлетворительно

2.2. Качество изложения материала, включая структуру статьи (название, аннотация, ключевые слова, введение, обзор текущего состояния, предлагаемый метод, проведенные исследования и обсуждение полученного результата, заключение, литература):

хорошо

2.3. Рекомендация по объему статьи:

краткая

3. Общая рекомендация:

отклонить

4. Ваша уверенность в данной рекомендации. средняя

5. Комментарий для программного комитета (по желанию).

В статье приведено решение задачи, которая представляется уже решенной много раз в уже существующих исследованиях. Описанный способ решения не представляется эффективным или хоть сколько-нибудь конкурентным по сравнению с уже существующими. При этом во Введении не приведено достаточно адекватное положение современных исследований по теме статьи.

В целом считаю, что в силу отсутствия новизны статью не следует публиковать.

При этом сами по себе упомянутые обстоятельства не являются жесткими препятствиями к публикации. В случае, если программный комитет все же склоняется к публикации статьи, рекомендация состоит в том, чтобы существенно редуцировать статью до объема краткой (5-9 стр.) в силу отсутствующей или неочевидной новизны. Альтернативно можно рекомендовать сместить фокус статьи на исследование экспертно разрабатываемых признаков или подчеркнуть новизну. Как дополнительная альтернатива, статья может быть опубликована в форме некоторого обучающего материала (т.н. tutorial), однако для этого следует существенно переформатировать ее под такой формат.

6. Комментарий для авторов работы.

В работе представлен подход, основанный на методах машинного обучения (МО), для классификации дерматоскопических фотографий, содержащих повреждения кожи, в отношении типа повреждений; преимущественный фокус исследования направлен на классификацию новообразований. В работе представлен подход, основанный на классических моделях МО, который резонно потребовал реализации этапа дизайна признакового описания фотографий, который расписан достаточно подробно для воспроизведения исследования. Постановка задачи упомянута лишь опосредованно в Разделе 6, где упоминается, что выбираются модели МО для решения задачи классификации дефектов кожи. Дизайн исследования описан достаточно ясно; четко описаны меры качества, которыми оцениваются решения. Результаты представлены достаточно четко и ясно. Выводы подкреплены результатами.

К исследованию есть целый ряд существенных замечаний.

В эпоху искусственных нейронных сетей, демонстрирующих решающее преимущество в большинстве задач обработки изображений перед классическими моделями МО, и при условии постановки задачи как классификации изображений, выбор способа решения в пользу экспертного дизайна признаков и применения классических моделей МО выглядит необоснованным. Подход выглядит тем более неоправданным при условии, что начиная с 2017 года уже опубликован целый ряд статей (например, [1,2]), демонстрирующих весьма высокое качество в задачах (бинарной) классификации меланомы и в целом поражений кожи с использованием свёрточных нейросетей. Тем более неясно, почему эти исследования, довольно широко цитируемые, не упомянуты в обзорной части во Введении. В связи с этим современное состояние проблемы, решаемой в исследовании, нельзя считать представленным даже в начальном объеме. Скорее, следует считать представленный обзор современных методов для решения похожих конечных задач сильно устаревшим. В связи с этим нельзя

считать, что в исследовании сформулирована новизна по сравнению с существующими исследованиями. В случае, если выбор методов решения и подходов формирования признакового пространства был продиктован какими-то специальными соображениями, это следует отдельно подробно описать во Введении или в отдельном разделе после Введения, посвященном формулировке задачи и выбору подхода ее решения.

Выбор моделей МО, примененных в задаче, никак не обоснован. Особенно странно выглядит включение модели решающих деревьев (DT), которая известна склонностью к переобучению и чрезвычайно высокой чувствительностью обобщающей способности к выбору гиперпараметров. В то же время, в исследовании не использованы ансамблевые модели, демонстрирующие наиболее высокое качество на современных задачах с табличными данными, а именно градиентный бустинг над решающими деревьями или полносвязные искусственные нейронные сети. Реализации таких моделей доступны и в пакете `scikit-learn`, и отдельно (например, `CatBoost`). Выбор списка моделей может быть продиктован разными соображениями, и можно предположить, что в свете каких-то резонансов приведенный крайне неудачный набор тестируемых моделей может оказаться обоснованным. Эти соображения должны быть приведены.

Модель KNN - "K nearest neighborhood" (модель классификации или регрессии в зависимости от сценария применения). Приведенная в статье расшифровка "K-means method" - модель кластеризации, не релевантная исследованию.

Выбор мер качества, использованных в исследовании, никак не обоснован. Более того, применяется либо мера Accuracy (доля верных ответов), известная ненадежностью в случае сильного дисбаланса классов (что наблюдается в исследовании), либо Recall (полнота, чувствительность), односторонняя мера, не характеризующая специфичность или точность. В существующих исследованиях такого же рода наряду с полнотой применяются другие меры, такие как специфичность, индекс Жаккара, F1-мера, *etc.* Подбор меры качества в задаче классификации дефектов кожи для неидеального классификатора при условии сильно несбалансированной выборки - сложный многогранный вопрос, в котором затрагиваются этические, социальные, статистические аспекты. При условии несогласованности выбранных мер качества с существующими высоко цитируемыми исследованиями обязательно следует обосновать выбор именно Accuracy и Recall. Альтернативно следует воспользоваться сложившейся практикой и опираться на меры, используемые в существующих исследованиях.

Отдельно следует отметить, что в результате чистки наборов данных существенное количество экземпляров были исключены из рассмотрения. При этом исследование позиционируется как шаг в направлении к автоматизированной системе классификации дефектов кожи, которая должна обрабатывать все снимки, поступающие на анализ. Таким образом, вопрос ограничений применимости разработанных моделей никак не освещен. В тексте заключения должны быть упомянуты такие ограничения.

Далее приводятся замечания по форме изложения исследования.

Описание процесса подбора признаков приведено чересчур подробно. В деталях описывается трудный многофазный путь авторов, который в конечном итоге приводит к одному единому способу формирования признакового описания. В случае, если была поставлена цель продемонстрировать исследование релевантности признаков, это должно быть описано отдельно; должны быть приведена методика исследования, приведены результаты этого исследования и проведены соответствующие выводы. Однако ни в аннотации, ни в заголовке статьи, ни во Введении не обозначена цель исследования релевантности признаков. Считаю, что описание подробного пути, приводящего к правильным признакам, следует существенно сократить и переформулировать.

Саму задачу, решаемую в исследовании, следует сформулировать во Введении, а не в Разделе 6. Читателю с самого начала статьи должно быть ясно, какая решается задача. Это не до конца ясно даже из аннотации, в которой значится буквально следующее: "In this work, a

program was developed for segmentation of affected and unaffected skin areas. The resulting masks were used to train machine learning models." Для решения какой именно задачи обучаются модели МО - неясно. Для чего разрабатывается программа сегментации чистых и поврежденных участков кожи - тоже неясно. Эти детали должны быть обоснованы и прояснены во Введении или отдельном разделе, посвященном формулировке задачи.

Общая рекомендация по статье состоит в том, чтобы существенно редуцировать статью до объема краткой (5-9 стр.) в силу отсутствующей или не представленной новизны. Альтернативно можно рекомендовать сместить фокус статьи на исследование экспертно разрабатываемых признаков или подчеркнуть новизну. Как дополнительная альтернатива, статья может быть опубликована в форме некоторого обучающего материала (т.н. tutorial), однако для этого следует существенно переформатировать ее под такой формат.

СПИСОК ЛИТЕРАТУРЫ

- [1] Yu, Lequan, et al. "Automated melanoma recognition in dermoscopy images via very deep residual networks." *IEEE transactions on medical imaging* 36.4 (2016): 994-1004
- [2] Lopez, Adria Romero, et al. "Skin lesion classification from dermoscopic images using deep learning techniques." 2017 13th IASTED international conference on biomedical engineering (BioMed). IEEE, 2017.