

НЕКОТОРЫЕ МЕТОДИЧЕСКИЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ НЕЙРОННЫХ СЕТЕЙ В МОРФОЛОГИЧЕСКОЙ ДИАГНОСТИКЕ НОВООБРАЗОВАНИЙ МОЛОЧНЫХ ЖЕЛЁЗ И МЕТАСТАЗОВ В ЛИМФАТИЧЕСКИХ УЗЛАХ.

Яценко Инна Владимировна, Борбат Артём Михайлович
МБУ ИНО ФГБУ ГНЦ ФМБЦ имени А.И. Бурназяна ФМБА России
inna.yatsenko@bk.ru

Сверточная нейронная сеть (CNN) является одним из наиболее популярных алгоритмов для анализа материала в глубоком обучении. Модели учатся непосредственно на изображениях, используя шаблоны для классификации и устраняя потребность в ручном извлечении признаков. Одним из лимитирующих факторов для развития технологий нейросетевого анализа в микроскопической диагностике является необходимость учебных наборов данных, для создания которых требуются опытные, квалифицированные морфологи. Мы сообщаем о подходе к расширению существующих наборов гистологических данных с использованием уже обученной нейросети. Цель: сравнить качество обучения нейросетевой модели в зависимости от разных способов формирования набора данных, подготовленных с использованием исходной модели, на примере опухолей молочной железы и метастазов в лимфатические узлы.

Материалы и методы

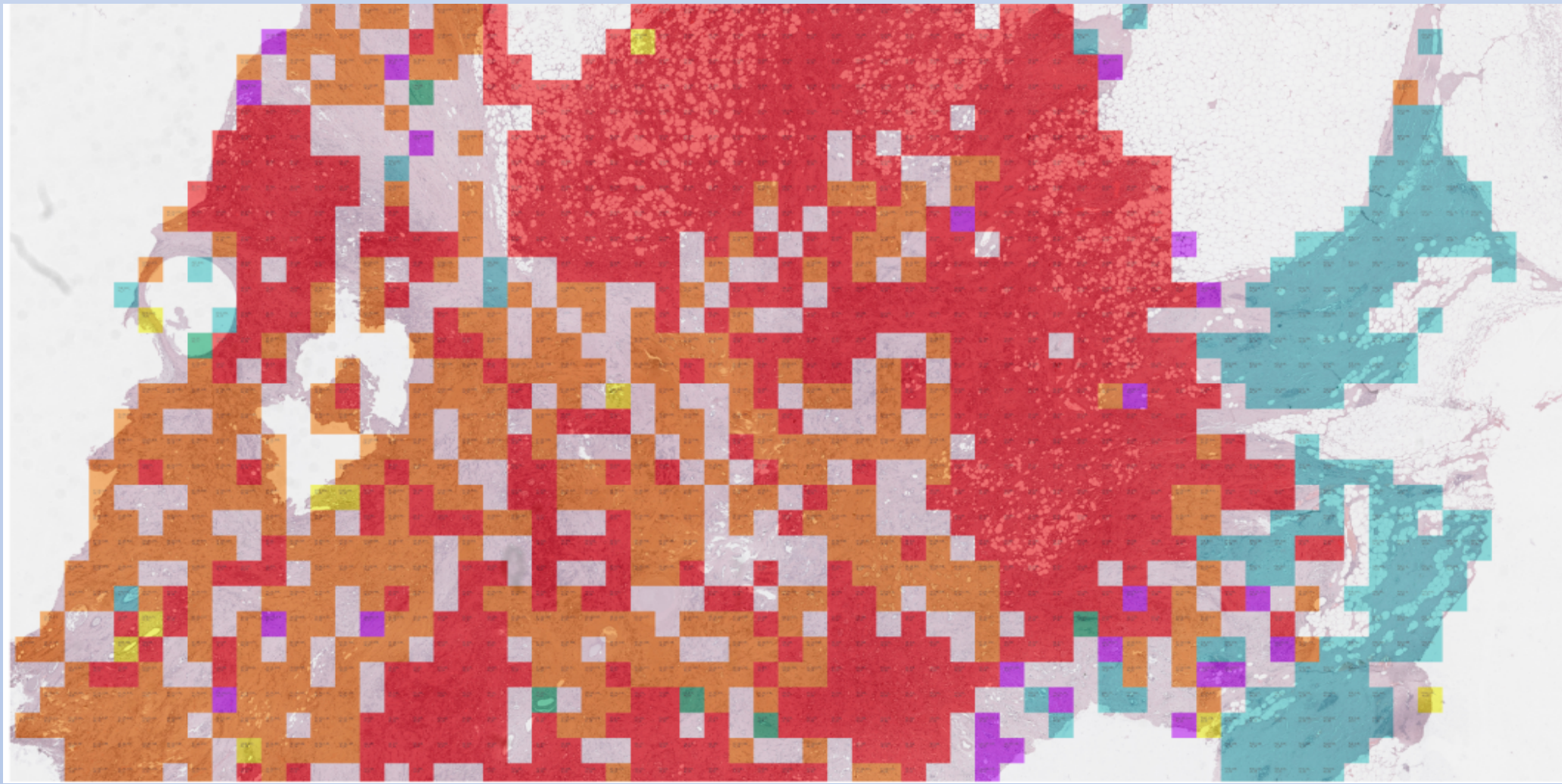
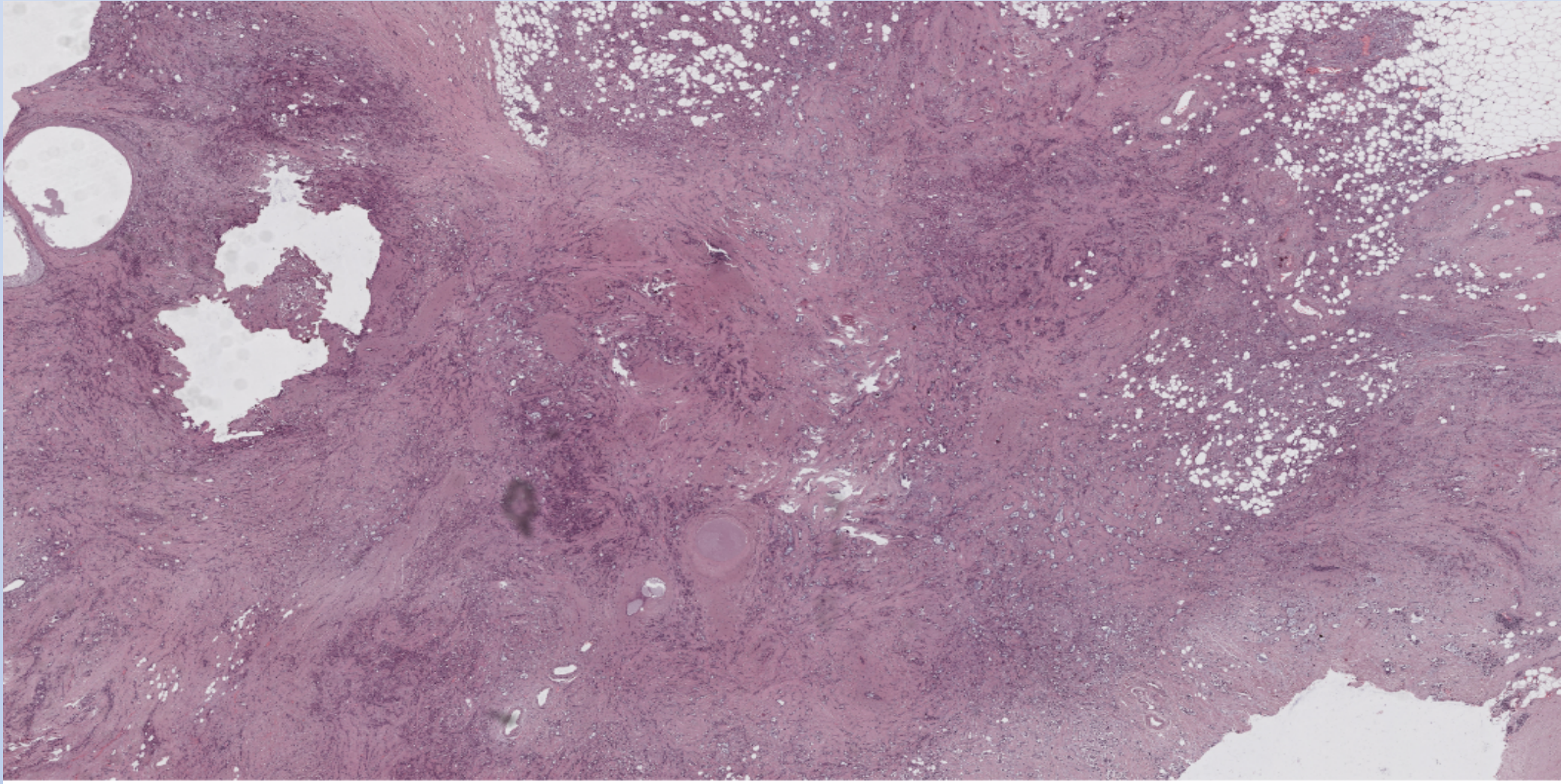
Ранее опубликованный набор данных с различными патологическими процессами молочных желез (104 случая, 23 тыс. изображений), 57 новых случаев для расширения набора данных и 50 новых тестовых случаев. Классификационные группы: неспецифический инвазивный дифференцированный рак, неспецифический инвазивный низкодифференцированный рак, дольковый инвазивный рак, неинвазивная протоковая карцинома, фиброаденома, фиброзно-кистозные изменения, папиллома и фон.

Исходная модель лимфатических узлов включала 57 микропрепаратов (6 случаев, 81 тыс. изображений), 72 дополнительных микроперпарата для расширения набора данных и 82 дополнительных микропрепарата со 151 лимфатическим узлом для тестирования моделей.



Первоначальная модель сверточной нейронной сети была обучена с использованием опубликованного набора данных и применена к новым случаям на основе полученных квадратов 300x300 пикселей. Для дальнейших шагов были использованы два вида квадратов: квадраты с однозначными результатами, с вероятностью классификации 80% и более (всего 46 тыс. квадратов) и квадраты с неоднозначными результатами, которые включали две конкурирующие категории, каждая с вероятностью не менее 40% (всего 28 тыс. квадратов). Третья группа состояла из двух групп 40+80 % (всего 50 тыс. квадратов). Группы по молочной железе были добавлены в опубликованный набор данных.

На основе расширенных наборов данных были обучены и протестированы три новые модели с тестовыми примерами. Результаты анализа оценил патологоанатом.



РЕЗУЛЬТАТЫ

Диагностика патологических состояний в молочной железе	Диагностика метастазов в лимфатических узлах
Наилучший результат показала модель, обученная на материале 40+80% - совпадение по диагностическим категориям составляло 54 % (27 из 50), по категории злокачественный/доброкачественный процесс 96% (48 из 50). По отсечению 40% и 80% получились следующие значения: по диагностическим категориям - 54% и 50%, по категории злокачественный/доброкачественный процесс - 92% и 86%.	Наилучший результат показала модель, обученная на материале 80% - чувствительность 100%, специфичность 67%. Модели 40% и 40+80% - чувствительность 100% и 100%, специфичность 51% и 33%, соответственно.

ЗАКЛЮЧЕНИЕ

Изначально была сформулирована и по предварительным данным принята гипотеза о том, что наполнение набора данных неоднозначными случаями должно улучшить обучение модели. По результатам данного исследования гипотеза подтвердилась для первичных поражений молочной железы, но не нашла подтверждения при диагностике метастазов в лимфатические узлы, где ключевым фактором оказался объем материала для обучения модели.