

Моргаева В.А.

*Научный руководитель: к. т. н., доц. каф. ИС Подгорнова Ю.А.
Муромский институт (филиал) федерального государственного образовательного
учреждения высшего образования «Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирская обл., ул. Орловская, 23
E-mail: vika_ozhegina@mail.ru*

Распознавание кисты молочной железы с помощью алгоритма случайного леса

Киста является достаточно распространенной патологией молочной железы. Она имеет пузыревидную форму и наполнена жидким содержимым. Данное образование относится к доброкачественным, но оно опасно возможностью развития воспалительного процесса и зарождением внутри капсулы онкологического заболевания.

Своевременная постановка диагноза пациенту, поможет минимизировать риск развития злокачественных образований на фоне доброкачественных изменений.

Маммография является одним из самых распространенных способов диагностики заболеваний молочной железы [1]. Однако не всегда на маммографических снимках можно точно определить диагноз и назначить правильное лечение пациенту.

Для повышения точности постановки корректного диагноза используются компьютерные системы для обработки и анализа заболеваний молочной железы на маммографических снимках.

Каждый алгоритм распознавания новообразований содержит несколько основных этапов:

1) Предварительная обработка изображения: этот шаг важен для некоторых методов, с целью улучшения изображения и уменьшения шума с минимальным искажением функций изображения. В некоторых САПР отсутствует этап предварительной обработки[2].

2) Сегментация изображений. Основная цель сегментации - выделение области интереса (ROI), соизмеримой с желаемыми свойствами.

3) Расчет признаков: на этом этапе извлекаются различные признаки в соответствии с характеристиками поражений из изображения. Так как существует большое количество признаков, используемых для классификации образований, необходимо провести анализ и выбрать из них наиболее эффективные для поставленной цели. В последующем эти признаки используются для различения доброкачественных и злокачественных образований. От эффективности отобранных признаков во многом зависит результат корректной классификации областей.

4) Классификация: в соответствии с выбранными характеристиками подозрительные области классифицируются на доброкачественные и злокачественные на основе различных методов классификации.

Предварительная обработка изображения выполняется с помощью адаптивного выравнивания гистограммы, медианной фильтрации для улучшения контрастности и подавления шума на изображении. Затем для полученного изображения выполняется сегментация методом Оцу для выделения области интереса.

После выполнения сегментации изображения выполняется расчет признаков построение дерева решений и на его основе выполняется классификация новообразования с помощью алгоритма случайного леса.

Для классификации изображения используются признаки, рассчитанные на основе гистограммы направленных градиентов (HOG), а также матрицы смежности уровней серого (GLCM).

На основе признаков, вычисленных для обучающих изображений, выполняется построение дерева решений.

Основная идея любого алгоритма дерева решений заключается в следующем:

- выбор лучшего атрибута, используя Мету выбора атрибутов (ASM), чтобы разделить записи;
- установка выбранного атрибута узлом решения и разбейте набор данных на меньшие подмножества.

Построение дерева, повторяя этот процесс рекурсивно для каждого потомка, пока не будет выполнено одно из условий:

- все кортежи принадлежат одному и тому же значению атрибута;
- больше нет оставшихся атрибутов;
- больше нет экземпляров;

На основе построенных деревьев решений выполняется классификация алгоритмом случайного леса.

Случайный лес — это множество решающих деревьев. В задаче регрессии их ответы усредняются, в задаче классификации принимается решение голосованием по большинству. Все деревья строятся независимо по следующей схеме:

Выбирается подвыборка обучающей выборки размера (м.б. с возвращением) – по ней строится дерево (для каждого дерева — своя подвыборка).

Данная модель использует две ключевые концепции, которые и делают этот лес случайным:

- случайная выборка образцов из набора данных при построении деревьев.
- при разделении узлов выбираются случайные наборы параметров [3].

Для построения каждого расщепления в дереве просматриваем некоторое количество (определяется случайным образом) случайных признаков (для каждого нового расщепления — свои случайные признаки) [4].

Выбирается наилучший признак и расщепление по нему (по заранее заданному критерию). Дерево строится, как правило, до исчерпания выборки (пока в листьях не останутся представители только одного класса), но в современных реализациях есть параметры, которые ограничивают высоту дерева, число объектов в листьях и число объектов в подвыборке, при котором проводится расщепление.

Для оценки эффективности результата работы алгоритма распознавания были взяты маммографические снимки из международной базы данных MIAS. В результате тестирования алгоритма было выявлено, что 3 маммограммы из 35 были классифицированы не правильно, остальные были классифицированы корректно. Чувствительность алгоритма составила 88,89%, специфичность 94,12%, а точность алгоритма 91,43%.

Литература

1. Заридзе Д.Г. Динамика заболеваемости злокачественными новообразованиями и смертности от них в России. / Д.Г. Заридзе, А.Д. Каприн, И.С. Стилиди// Вопросы онкологии. 2018. Т.64. №5. С. 578-591.
2. Afsaneh Jalalian, Syamsiah Mashohor , Rozi Mahmud , Babak Karasfi , M. Iqbal B. Saripan , Abdul Rahman B. Ramli. Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection. // EXCLI Journal. 2017. №16. С. 113-137.
3. Дьяконов А. Случайный лес (Random Forest) – URL: <https://dyakonov.org/2016/11/14/%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D1%8B%D0%B9-%D0%BB%D0%B5%D1%81-random-forest/> (Дата обращения: 15.03.2022).
4. Штукатуров С. Реализация и разбор алгоритма «случайный лес» на Python – URL: <https://tproger.ru/translations/python-random-forest-implementation/> (Дата обращения: 18.03.2022).