

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.855

Хачатрян
Арман Гагикович

Применение методов машинного обучения для диагностики
злокачественных новообразований

АВТОРЕФЕРАТ
на соискание академической степени
магистра
по специальности 1-40 80 05 – Программная инженерия

Научный руководитель

М. С. Абрамович,
к.ф.-м.н., доцент

Минск 2021

ВВЕДЕНИЕ

Меланома кожи является одной из самых опасных злокачественных новообразований человека. Показатели заболеваемости увеличиваются в 2 раза среди населения Европы каждые 10-20 лет (ежегодный прирост составляет 3% - 7%). В Европе меланома занимает 17 место среди наиболее часто диагностируемых видов рака у мужчин и 8 место у женщин. Со временем о меланоме сложилось мнение как о заболевании с переменным и неблагоприятным клиническим течением. В отличие от других злокачественных новообразований, меланома чаще поражает людей относительно молодого возраста, обладает высоким метастатическим потенциалом и резистентностью к лучевой и химиотерапии. Хотя только 4% от всех злокачественных новообразований кожи является меланома, на неё приходится 79% смертей этой группы. Несмотря на это, локальная меланома не является болезнью с однозначно плохим прогнозом. В случае ранней диагностики, поддающимся лечению оказываются до 95% опухолей.

Лечение больных меланомой кожи является одной из самых актуальных проблем онкологии. Среди нерешенных вопросов в лечении больных меланомой кожи остается и вопрос о необходимости проведения профилактической регионарной хирургической операции по удалению лимфоузлов.

На протяжении долгого времени учёные пытаются определить факторы, которые влияют на протекание и точность прогнозов заболевания у больных меланомой кожи. В 2002 г утвердили новую версию классификации меланомы кожи. Изменения были произведены на основании анализа прогностических факторов, изученных у 17 600 больных меланомой, и коснулись следующих параметров:

- толщина опухоли, наличие или отсутствие изъязвления;
- количество метастатически поражённых лимфоузлов и наличие микро- или макрометастазов в лимфоузлах;
- наличие сателлитов вокруг первичной опухоли и транзиторных метастазов (было объединено в единую III стадию заболевания);
- дополнительную информацию о стадии заболевания получают при лимфатическом картировании и биопсии сторожевого лимфоузла.

Это вызвало интерес к исследованию эффективности применения методов машинного обучения для диагностики злокачественных новообразований, так как выявление новообразований на ранних этапах без видимых внешних признаков даст пациенту больше шансов на выздоровление, тем самым снизит смертность среди заболевших.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является исследование эффективности применения методов машинного обучения для диагностики злокачественных новообразований различных локализаций (больных меланомой кожи).

Объектом исследования являются методы машинного обучения.

Предметом исследования является классификация больных меланомой кожи с наличием метастаз в лимфоузлах и их отсутствием.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность прогнозирования отсутствия метастаз в лимфоузлах с целью уменьшения проведения профилактической регионарной хирургической операции по удалению лимфоузлов.

Апробация результатов диссертации

Материалы, положенные в основу работы, докладывались и обсуждались на Международной конференции «ИТС-2020 БГУИР» (Минск, Беларусь, 2020), а также в Республиканской научно-практической конференции «Вычислительные методы, модели и образовательные технологии» (Брест, Беларусь, 2020).

Опубликованность результатов диссертации

По теме диссертации опубликовано 2 печатных работы в сборниках материалов международных научных конференций. Из них 1 работа в сборнике трудов и материалов международной конференции ИТС-2020 БГУИР и 1 работа в сборнике материалов IX Республиканской научно-практической конференции «Вычислительные методы, модели и образовательные технологии».

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, списка использованных источников, списка публикаций автора и приложения. В первой главе представлен аналитический обзор методов машинного обучения, выявлены основные преимущества каждого из методов, показаны основные параметры методов. Вторая глава посвящена описанию и анализу метрик для оценки эффективности применения методов машинного обучения. В третьей главе описаны методы отбора информативных признаков для эффективного применения методов машинного обучения при классификации пациентов. В четвертой главе описано получение результатов применения методов машинного обучения для диагностики злокачественных новообразований.

КРАТКОЕ СОДЕРЖАНИЕ

Во введении рассмотрена актуальность выбранной темы, а также приведена проблема, которая приводит к необходимости исследования.

В первой главе представлен аналитический обзор методов машинного обучения, выявлены основные преимущества каждого из методов, показаны основные параметры методов.

Вторая глава посвящена описанию и анализу метрик для оценки эффективности применения методов машинного обучения.

В третьей главе описаны методы отбора информативных признаков для эффективного применения методов машинного обучения при классификации пациентов.

В четвертой главе описаны исходные данные, представлены полученные результаты применения методов машинного обучения для диагностики злокачественных новообразований.

Библиотека БГУИР

ЗАКЛЮЧЕНИЕ

В ходе исследования были рассмотрены 5 методов машинного обучения:

- Деревья решений;
- Опорных векторов;
- Логистическая регрессия;
- Случайный лес;
- k-ближайших соседей.

Так же были рассмотрены метрики машинного обучения, которые использовались для оценки эффективности применения методов и отбор информативных признаков для повышения эффективности применения методов машинного обучения.

Были проанализированы исходные данные, и получены результаты применения выбранных методов машинного обучения для диагностики злокачественных новообразований. Полученные результаты приведены в таблице 1.

Таблица 1 – Результаты классификации методов

Название	Чувствительность, %	Специфичность, %	Эффективность, %
Деревья решений	79,8	76,0	76,8
Метод опорных векторов	72,0	74,5	74,0
Логистическая регрессия	75,5	75,3	75,4
Случайный лес	89,9	81,5	83,1
k ближайших соседей	84,8	77,8	79,1

Из таблицы 1 можно понять что самым эффективным методом машинного обучения для диагностики злокачественных новообразований является метод "Случайный лес т.к. все показатели эффективности у него выше, чем у остальных трех методов, а именно:

- диагностическая чувствительность = 89,9%;
- диагностическая специфичность = 81,5%;
- диагностическая эффективность = 83,1%.

В дальнейшем данные методы машинного обучения можно использовать и в других отраслях медицины для диагностирования больных.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

[1] Хачатрян, А. Г. Деревья решений как метод диагностики злокачественных новообразований / Хачатрян А. Г. // Компьютерные системы и сети: материалы международной научной конференции аспирантов, магистрантов и студентов, Минск, 18 ноября 2020 г. / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2020. – С. 205 – 206.

[2] Хачатрян, А. Г. Деревья классификаций как метод диагностики злокачественных новообразований / Хачатрян А.Г. // Материалы 9 Республиканской научно-практической конференции «Вычислительные методы, модели и образовательные технологии», Минск, 22 октября 2020 г. / Брестский Государственный Университет Имени А.С. Пушкина. – Брест, 2020.

Библиотека БГУИР