

Министерство образования Российской Федерации  
Научно-Исследовательский Университет Высшая Школа  
Экономики

*Факультет Компьютерных Наук  
Факультет Экономических Наук*

# **Применение алгоритмов машинного обучения в сфере здравоохранения**

*Мирзоева Алина БЭАД223*

Москва, 2023

# 1 Введение

Машинное обучение (МО) представляет собой фундаментальную область искусственного интеллекта, фокусирующуюся на создании алгоритмов и моделей, способных обучаться и принимать решения без явного вмешательства человека. Основная идея заключается в том, чтобы компьютерные системы могли адаптироваться к новой информации и совершенствовать свою производительность в процессе взаимодействия с данными. Машинное обучение находит применение в различных сферах, от распознавания изображений и естественного языка до прогнозирования трендов и оптимизации бизнес-процессов. Эта область технологий становится все более важной в нашем цифровом мире, играя ключевую роль в автоматизации и оптимизации разнообразных задач.

В данном обзоре будет рассмотрена литература и полученные в этих исследованиях результаты на тему применения алгоритмов машинного обучения в различных задачах современного здравоохранения.

## 2 Основные термины:

### **Алгоритмы машинного обучения:**

*Алгоритм k ближайших соседей (k-nearest neighbors, или k-NN)* - это метод обучения без учителя, используемый для задач классификации и регрессии. Основная идея заключается в том, чтобы классифицировать объект так: его класс - это класс, наиболее часто встречающийся среди k ближайших соседей точки, обозначающий наш исходный объект.

*Метод опорных векторов (Support Vector Machine, SVM)* - это алгоритм машинного обучения, который используется для задач классификации и регрессии. Основной целью алгоритма SVM является поиск оптимальной гиперплоскости в N-мерном пространстве, которая может разделить точки, принадлежащие разным классам в пространстве признаков. Иными словами, цель - максимизировать

расстояние между ближайшими точками различных классов.

**Свёрточная нейронная сеть** (*Convolutional Neural Network, CNN*) — это тип искусственных нейронных сетей, специально разработанный для обработки и анализа структурированных данных, например, изображений. Основная особенность CNN заключается в использовании свёрточных слоёв для обнаружения и выделения важных закономерностей и признаков из входных данных. Каждый свёрточный слой состоит из фильтров, представляющих собой матрицы, далее каждая часть изображения умножается на матрицу (ядро) свёртки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения. Таким образом выделяются различные признаки изображения и их пространственные зависимости.

**Байесовская классификация** — метод машинного обучения, основанный на теореме Байеса. Суть метода в определении вероятности принадлежности объекта к разным классам, учитывая его признаки и априорные вероятности классов (вероятности принадлежности объекта к каждому классу без учёта признаков). Алгоритм обновляет вероятности после поступления новой информации и принимает решение о классификации на основе постериорных вероятностей.

### **Основные показатели для оценки результатов**

**Матрица ошибок** - матрица, такая что столбцы означают то, к какому классу объект относится в реальности, а строки - к какому классу относится объект по прогнозу модели.

**TP (True Positive)** - Это количество объектов, которые на самом деле принадлежат к положительному классу (истинно положительные).

**TN (True Negative)** - Это количество объектов, которые фактически принадлежат отрицательному классу, но ошибочно классифицированы как положительные. Аналогично определяются False Positive и False Negative

С помощью матрицы ошибок мы можем ввести несколько следующих метрик:

**Accuracy:**  $\frac{TP+TN}{TP+TN+FP+FN}$  — доля исходов, которую наша модель определила правильно.

**Precision:**  $\frac{TP}{TP+FP}$  — какая доля от предсказанных объектов класса 1 и в реальности является объектом класса 1.

**Recall:**  $\frac{TP}{TP+FN}$  — какая доля от тех значений, которые в реальности должны относиться к классу 1, были правильно определены моделью.

*Кривая ROC (Receiver Operating Characteristic)* представляет собой графическое отображение эффективности классификационной модели. На этой кривой сравниваются истинно положительная частота (TPR) и ложно положительная частота (FPR).

*AUC (Area Under the Curve)* представляет собой площадь под ROC-кривой. Чем больше AUC, тем лучше производительность модели. AUC является мерой качества модели, учитывающей её способность разделять классы.

### 3 Обзор литературы

Как было упомянуто выше, машинное обучение активно применяется во многих сферах человеческой деятельности. Одной из наиболее важных и перспективных областей применения этих технологий является здравоохранение. Многие ученые в своих работах предлагают различные подходы к применению машинного обучения в этой отрасли. Так как данные алгоритмы могут превращать огромные объемы различных данных в ценную информацию, то благодаря исследованию их применения ученым предоставляется возможность открыть новые горизонты для диагностики, лечения и управления заболеваниями.

Задача классификации является одной из самых важных задач машинного обучения, именно поэтому достаточно большое количество статей посвящено построению эффективных моделей для прогнозирования развития заболевания на основе различных данных об организме пациента.

Машинное обучение может применяться на различных этапах диагностики болезней: от анализа генов до постановки диагноза на основе уже проявившегося на коже образования.

Например, Т.С. Фьюри и соавторы [9] в статье "Support vector machine classification and validation of cancer tissue samples using microarray expression data" еще в 2000 году провели эксперименты по разработке нового метода анализа данных, состоящих из значений уровней экспрессии тысяч генов при использовании ДНК-микрочипов, который будет выполнять как задачу классификации образцов тканей при диагностике раковых заболеваний, так и исследования данных на предмет неправильно маркированных или сомнительных. В исследовании были использованы значения экспрессии на 97,802 клеточных клонированных ДНК для 31 образца ткани. Данные были поделены на три класса ("раковая ткань яичника", "нормальная ткань яичника" или "нормальная ткань, не связанная с яичниками"). Последние два класса были объединены в один класс - "ткань, не содержащая раковых клеток". Изначально проведены эксперименты с

использованием данных экспрессии на всех 97,802 генах. Затем гены ранжируются, и формируются наборы данных, включающие топ-25, топ-50, топ-100, топ-500 из них, такие, что алгоритм предполагает, что их важность в процессе анализа и предсказания выше, чем у других. В данном исследовании лучшие результаты классификации достигнуты при использовании топ-50 признаков с диагональным коэффициентом 2 или 5 ( $TP = 12$ ,  $TN=14$ ,  $FP= 2$ ,  $FN = 3$ ). Несмотря на то, что меньшие наборы данных показывают немного более точные результаты по сравнению со всеми признаками (значения для результатов на всех признаках:  $FP = 7$ ,  $FN = 3$ ,  $TP = 12$ ,  $TN = 8$ ), авторы не считают этот прирост значительным. Доступные во время исследования наборы данных содержат относительно мало примеров и, таким образом, эта проблема не позволила авторам точно исследовать модель на больших объемах данных. Авторы заключают, что SVM проявляет высокую эффективность при использовании простого ядра на небольших наборах данных, и предполагают, что с появлением более крупных наборов данных может потребоваться использование более сложных ядер для поддержания хорошей производительности. Дополнительным преимуществом метода SVM является возможность выявления ошибочно помеченных данных (как например данные, которые в целом не содержали в себе ткани исследуемого органа).

Также достаточно глубоко была исследована возможность применения алгоритмов для предсказания дальнейшего течения заболевания:

В статье Купера и др "Predicting dire outcomes of patients with community acquired pneumonia"(2005) [5] была поднята проблема прогнозирования тяжелых исходов, таких как смертность или тяжелые клинические осложнения, при лечении пациентов с внебольничной пневмонией (далее - ВП). Этот аспект является особенно важным в лечении пациента, так как позволяет врачам определить следует ли лечить пациента в стационаре или амбулаторно. Для этого авторы использовали обучающую выборку из 1601 случаев пациентов с ВП, чтобы построить 11 статистических моделей и моделей машинного обучения, которые предсказывают тяжелые последствия. Спроекти-

рованные 11 моделей включают в себя: разные модификации Байесовской классификации, логистическую регрессию, нейронные сети. Оценка результатов каждой модели производилась на валидационной выборке, состоящей из 686 дополнительных случаев пациентов с ВП. Была создана модель нейронной сети с использованием двух новых методов: первый метод заключается в обучении модели предсказывать тяжелые последствия, посредством предсказания так же сопутствующих ему осложнений, а второй метод фокусируется на максимизации AUC. Именно эта специальная версия модели искусственной нейронной сети показала лучшие результаты в предсказании тяжелых исходов ( $AUC = 0,84$  в среднем). Также был проведен анализ применения данной модели с точки зрения потенциальной финансовых выгоды для пациентов, которых будут отправлять на амбулаторное лечение и для больниц, которые будут снижать нагрузку на свои отделения. Однако, авторы приходят к выводу, что если слишком сильно максимизировать этот показатель, то это может привести к неточности модели и увеличению смертности после рекомендации амбулаторного лечения. Поэтому поиск моделей с максимальным возможным уровнем предсказательной эффективности имеет важное значение по мнению авторов, а их модель требует дальнейшей валидации. Однако авторы подчеркивают, что даже небольшие улучшения предсказательной эффективности для распространенных и дорогостоящих заболеваний, вероятно, приведут к значительным улучшениям эффективности предоставления медицинской помощи.

Исследование "Feature selection in bayesian classifiers for the prognosis of survival of cirrhotic patients treated with tips"(2005) Р.Бланко и др. [4] проведено с целью предсказания выживаемости пациентов в первые 6 месяцев после проведения трансъюгулярного внутрипечёночного портосистемного шунтирования (TIPS). Для этого был использован датасет из 107 случаев и 77 переменными. В работе применены модели байесовской классификации. Так как часто не все медицинские данные необходимы для конечного прогнозирования авторами были использованы методы отбора подмножества признаков. Результаты исследования, подтверждают гипотезу о том, что не

**все признаки из медицинских записей необходимы для точной классификации (приведет к смерти или нет) течения заболевания, и уменьшение их количества приводит к повышению точности модели.** Например, выборочный байесовский классификатор повышает ассигасу до 92%. Работа подчеркивает эффективность байесовских классификаторов в данном контексте и необходимость поиска оптимальных моделей с минимальным числом признаков для улучшения результатов.

За последние годы было изучено применение машинного обучения в диагностике множества различных заболеваний

В статье "Diagnosis of valvular heart disease through neural networks ensembles"(2009) И.Туркоглу и др [6] представили новую методологию, использующую базовое программное обеспечение SAS 9.1.3 для диагностики пороков сердца. В центре нашей предложенной системы лежит инновационный ансамблевый метод нейронных сетей. Эти методы, основанные на ансамблях, создают новые модели, объединяя апостериорные вероятности или прогнозируемые значения из нескольких предыдущих моделей. Такой подход позволяет создавать более эффективные и точные модели. База данных содержит снимки аортальных и митральных клапанов 132 мужчин и 83 женщин в возрасте от 15 до 80 лет (средний возраст 48 лет). Всего было идентифицировано 215 классификаций клапанов, в том числе 56 нормальных и 54 аномальных аортальных клапана, а также 39 нормальных и 66 аномальных митральных клапанов. В ходе экспериментов с предложенным алгоритмом была достигнута впечатляющая точность классификации в 97,4% при анализе набора данных, включающего 215 образцов.

В исследовании "A classification and regression tree algorithm for heart disease modeling and prediction."(2023) М.Озкан и др [12] был использован алгоритм Classification and Regression Tree (CART), метод обучения с учителем, для прогнозирования сердечных заболеваний и выделения правил принятия решений, разъясняющих взаимосвязи между переменными. **Кроме того, результаты исследования ранжируют важность признаков, влияющих на сердечные заболева-**



ния, что является полезным знанием для множества исследований в области как диагностики, так и профилактики сердечно-сосудистых заболеваний. При учете всех параметров производительности точность прогноза на уровне 87% подтверждает надежность модели. Извлеченные из деревьев решений правила, полученные в исследовании, могут помогать не только медицинским работникам в увеличении эффективности их работы, но и пациентам, сокращая процесс диагностики и затраты для его осуществления.

Одним из самых быстроразвивающихся отраслей применения машинного обучения является использование глубинного обучения и других подтипов машинного обучения для распознавания различных индикаторов развития заболевания (опухоли, меланомы и др.) на медицинских изображениях.

С.Хоукинс и соавторы [10] в 2016 году проводили исследование по определению возможности анализа изображений, получаемых при начальном скрининге легких для обнаружения злокачественных опухолей посредством компьютерной томографии. Результаты исследования направлены на проверку гипотезы о том, что анализ компьютерных томографических снимков на начальном этапе может точно предсказать, разовьется ли опухолевидное образование в клинический рак. Авторами были использованы публичные данные Национального института рака, которые были организованы в две группы: одна состояла из 104 пациентов с диагностированным раком легких в результате скрининга, а другая из 92 пациентов с аналогичным диагнозом. Затем эти группы были сопоставлены с датасетами из 208 и 196 пациентов, у которых были выявлены доброкачественные опухоли. Во время подбора самого эффективного алгоритма использовались такие методы как Байесовская классификация, метод опорных векторов (SVM), Random Forest classifier, JRIP Rule classifier. Приоритетными наборами свойств были приняты свойства из базы данных Reference Image Database to Evaluate Therapy Response (RIDER). В результате анализов было выделено 23 признака. Категория наиболее стабильных признаков включала описание размеров опухолей, а признаки текстуры проявили более низкие уровни стабильности из-за их

высокой зависимости от силы рентгеновских лучей сканера. Параметры сканера, такие как поле зрения, влияющие на размер пикселей, также влияли на текстуру изображения. То есть одним из допущений исследования является то, что качество признаков достаточно сильно зависит от качества устройства, которое производит снимок, что может осложнить обучение модели на наборе изображений с разных устройств. По результатам исследования, максимальная достоверность прогнозирования развития рака через 1 год при использовании исходного сканирования достигла 80,1% (площадь под кривой [AUC] = 0,83; частота ложноположительных результатов [FPR] = 9%) при использовании Random Forest classifier со свойствами, определенных по RIDER.

В статье "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer" (2017) Б.Бенджорди и др [3] предметом исследования становится применение алгоритмов глубокого обучения для обнаружения метастазов в лимфатических узлах у женщин с раком молочной железы, а также производится оценка эффективности автоматизированных алгоритмов глубокого обучения и сравнение результатов их работы с диагнозами врачей в диагностических условиях. Для разработки алгоритмов машинного обучения были собраны 399 изображений, сделанных в первой половине 2015 год у пациентов, прошедших хирургическое вмешательство по поводу рака груди в двух больницах Нидерландов. Тренировочные данные состоят из 110 изображений с метастазами и 160 изображений без метастазов. Оценка алгоритмов производилась на выборке из 129 изображений (49 с метастазами и 80 без). Содержание или отсутствие метастазов было подтверждено с использованием специального химического анализа. Тестовые данные также подвергались оценке от группы из 11 врачей. Будучи ограниченными по времени, эта группа провела оценку с целью определения вероятности метастазов для каждого изображения. Этот процесс проходил в рамках гибкого двухчасового сеанса, имитирующего рутинный рабочий процесс. Также в оценке участвовал 1 специалист, который не был ограничен по времени. Перед алгоритмом было поставлено две

задачи: 1) Идентификация отдельных метастазов, 2) Классификация метастазов. Основным показателем точности алгоритмов является AUC (Area Under Curve). Самый эффективный алгоритм достиг доли истинно положительных (True Positive Rate) результатов на том же уровне, что и наблюдались у врача (0,0125 ложноположительных результатов). У 5 лучших алгоритмов средняя AUC была сравнима с врачом, интерпретирующим слайды, при отсутствии ограничений по времени. Однако стоит отметить, что исследование проведено в рамках симуляционного упражнения, не полностью отражающего реальный патологический процесс, а тестовая выборка, в свою очередь, была обогащена метастазами, что отличается от реальных клинических сценариев, а также используемые в исследовании алгоритмы могут быть менее эффективны в выявлении редких патологий, не включенных в обучающий набор. Также авторы отмечают, что исследование ограничено оценкой метастазов рака молочной железы, не учитывая другие патологии и дополнительные окраски. Таким образом, авторы приходят к выводу, что некоторые алгоритмы глубокого обучения достигли более высоких диагностических показателей, чем группа из 11 врачей, участвовавших в симуляционном упражнении, призванном имитировать рутинный рабочий процесс патологии, а производительность алгоритма была сопоставима с интерпретацией изображений экспертом без ограничений по времени. Однако, для более четкого понимания возможного применения данных алгоритмов в клинических условиях необходимо более тщательное исследование и устранение всех допущений, указанных выше.

Особенно активно данные методы получили развитие в исследовании методологии диагностики рака кожи:

К.Барата и др. [2] в статье 2013 года рассматривают две разные системы обнаружения меланомы на дерматоскопических изображениях. Первая система использует глобальные методы для классификации поражений кожи, тогда как вторая система использует локальные признаки и "Bag-of-features" (BoF) (это метод в обработке изображений, который представляет изображение как набор ключевых точек или дескрипторов, не учитывая их пространственное расположение).

В статье сравнивается роль особенностей цвета и текстуры в классификации кожных поражений и ставится цель определить, какой набор признаков является более важным для конечного предсказания. Результаты классификации были получены на наборе данных из 176 дерматоскопических изображений из больницы Педро Эспано. После обучения модели авторами был сделан вывод, что цветовые характеристики превосходят характеристики текстуры при использовании для предсказания является ли образование меланомой.

С годами исследования и применяемые методы становились более сложными: они стали разделяться на несколько этапов для более качественной подготовки изображений для дальнейшего анализа, что делает алгоритмы более устойчивыми к работе с изображениями разного качества, что, в свою очередь, повышает точность алгоритма и делает потенциальную область его применения шире.

Например, в 2021 году в своей статье Дж. Акоста и др. [11] в целях диагностики рака кожи, используя фотографии кожных образований, использовали комбинированный подход, включающий методы сверточных нейронных сетей (CNN) с предварительно обученной структурой ResNet152. Набор данных был взят из конкурса Международного симпозиума по биомедицинскому изображению 2017 года. Предложенный алгоритм реализуется в два этапа. На первом этапе использовалась CNN для создания ограничительной рамки вокруг кожного образования, а на втором этапе ResNet152 использовалась для классификации этого образования как доброкачественного или злокачественного. Предложенная модель, eVida M6, продемонстрировала accuracy 90,4% и precision 90,4%. Ее результаты стали на 3,66% лучше других моделей ISIC 2017.

А в исследовании "A machine learning approach for skin disease detection and classification using image segmentation"(2022) М.Ахаммеда и др. [1] была представлена автоматическую модель классификации для распознавания нескольких типов заболеваний кожи. Эта модель создана для использования обширного набора соответствующих признаков и достигает высокой точности в процессе классификации. Для получения результатов авторы предложили алгоритм, состоящий из

данных этапов: первый этап — предварительная обработка изображения (удаление, например, волос или шумов с изображения), второй этап — сегментация изображения, третий этап — извлечение признаков и последний этап — классификация. Алгоритмы классификации в машинном обучении не могут обрабатывать изображения напрямую. Из-за этого авторам необходима разработка функции для преобразования изображений в набор признаков, необходимых для его классификации (например, дерматоскопическое изображение имеет различные характеристики, которые используются для описания). Однако не все характеристики применимы к классификации кожных заболеваний. Именно поэтому авторы предложили сначала каждое изображение заменять на матрицу совпадения уровней серого (GLCM). Затем на основе этой матрицы рассчитываются характеристики контраста, энтропии, однородности. Конечной задачей работы является классификация. В данной работе были использованы три классификатора: метод опорных векторов, К ближайших соседей, дерево решений. Для тренировки и валидации алгоритмов были использованы датасеты “International Skin Imaging Collaboration (ISIC) 2019”, “HAM10000 (Human-Against-Machine with 10000 training images)”. Перед использованием авторы использовали метод случайной передискретизации для того, чтобы сделать данные более сбалансированными. Так как несбалансированное количество объектов разных классов (в данном случае разных кожных заболеваний) может привести к проблемам в обучении модели, так как она может быть более склонна к предсказанию чаще встречающегося класса. По итогам тестирования алгоритмов на сбалансированном и несбалансированном наборе данных классификатор метода опорных векторов (SVM) имеет небольшое значение потерь для обоих датасетов, что указывает на лучшую точность классификации. Точности (ассигасу) предложенных алгоритмов на сбалансированном датасете составили рекордные значения: алгоритм SVM - 95%, дерево решений - 94% и метод KNN - 93%, что указывает на превосходство модели с использованием алгоритма SVM на большом наборе данных.

Также отличным примером является исследование А.Естева и

Б.Курпела (2022) [8] представляет метод классификации заболеваний кожи с использованием глубоких сверточных нейронных сетей (CNN). Модель обучена на большом наборе из 129,450 клинических изображений различных заболеваний кожи. Оценка производится на биопсийно подтвержденных клинических изображениях в двух ключевых бинарных задачах: карциномы кератиноцитов против доброкачественных себорейных кератозов и злокачественных меланом против доброкачественных родинок. Эффективность алгоритма была проверена двумя способами. Сначала болезни были разделены на три класса — доброкачественные поражения, злокачественные поражения и неопухолевые поражения. В этой задаче CNN достигает  $72,1 \pm 0,9\%$  общей точности (среднее значение точности отдельных классов вывода), а два дерматолога достигают точности 65,56% и 66% на подмножестве проверочного набора. Далее авторы проверяют алгоритм, используя разделение болезней на девять классов таким образом, что заболевания каждого класса имели схожие планы лечения. CNN достигает общей точности  $55,4 \pm 1,7\%$ , тогда как те же два дерматолога достигают точности 53,3% и 55%. Таким образом модель достигает достаточной точности, подтверждая способность искусственного интеллекта классифицировать рак кожи на уровне, сопоставимом с профессиональными врачами.

Таким образом, за последнее десятилетие алгоритмы были улучшены настолько сильно, что могут заменить полноценную консультацию врача.

Однако, в медицинских кругах достаточно активно поднимается вопрос об этической составляющей применения машинного обучения и искусственного интеллекта в медицине.

В своей статье "Should we be afraid of medical ai?" 2019 года Е.Д.Нуччи [7] рассматривает тезис, выдвинутый Розалинд МакДугалл в Журнале медицинской этики, относительно потенциальной угрозы независимости пациентов со стороны медицинского искусственного интеллекта (ИИ). Основным кейсом здесь служит пример IBM Watson for Oncology (IBM Watson for Oncology - это система искусственного интеллекта, разработанная IBM для поддержки врачей

в принятии решений при лечении онкологических заболеваний. Эта технология использует машинное обучение и анализ данных для обработки обширной медицинской информации, включая медицинские записи, научные исследования, клинические протоколы и другие источники данных), который, по мнению МакДугалл, может привести к игнорированию индивидуальных ценностей и предпочтений пациентов.

Автор в свою очередь аргументирует, что в часто в таких тезисах происходит путаница между терминами "ИИ" и "машинное обучение". Машинное обучение в принципе гораздо мощнее в вычислительном отношении, чем ИИ, потому что его возможности значительно превосходят человеческие возможности; но ИИ может быть намного страшнее для простых пациентов, потому что он имитирует людей. Эта путаница влечет к тому, что люди упускают потенциал машинного обучения в области персонализированной медицины, основанной на обширных данных. Также в обсуждаемом тезисе не берется во внимание различие между предоставлением советов на основе фактических данных и фактическим принятием решений в области здравоохранения.

В итоге автор делает вывод, что вопрос о том, какие задачи и в каком объеме мы должны делегировать машинному обучению и другим технологиям в сфере здравоохранения, а также за её пределами, является одним из важнейших вопросов нашего времени. Однако для ответа на него необходимо тщательно анализировать и различать различные системы и задачи.

## **4 Вывод:**

Таким образом, на основе приведенных выше статей и результатов в них, иллюстрирующих прогресс в увеличении точности алгоритмов машинного обучения за последние десятилетия, можно сделать вывод, что машинное обучение имеет большую перспективу для становления одним из основных инструментов в диагностировании и предсказании возникновения различных заболеваний.

## 5 Используемая литература

### Список литературы

- [1] Mostafiz Ahammed, Md Al Mamun, and Mohammad Shorif Uddin. A machine learning approach for skin disease detection and classification using image segmentation. *Healthcare Analytics*, 2:100122, 2022.
- [2] Catarina Barata, Margarida Ruela, Mariana Francisco, Teresa Mendonça, and Jorge S Marques. Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE systems Journal*, 8(3):965–979, 2013.
- [3] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017.
- [4] Rosa Blanco, Iñaki Inza, Marisa Merino, Jorge Quiroga, and Pedro Larrañaga. Feature selection in bayesian classifiers for the prognosis of survival of cirrhotic patients treated with tips. *Journal of Biomedical Informatics*, 38(5):376–388, 2005.
- [5] Gregory F Cooper, Vijoy Abraham, Constantin F Aliferis, John M Aronis, Bruce G Buchanan, Richard Caruana, Michael J Fine, Janine E Janosky, Gary Livingston, Tom Mitchell, et al. Predicting dire outcomes of patients with community acquired pneumonia. *Journal of biomedical informatics*, 38(5):347–366, 2005.
- [6] Resul Das, Ibrahim Turkoglu, and Abdulkadir Sengur. Diagnosis of valvular heart disease through neural networks ensembles. *Computer methods and programs in biomedicine*, 93(2):185–191, 2009.



- [7] Ezio Di Nucci. Should we be afraid of medical ai? *Journal of Medical Ethics*, 45(8):556–558, 2019.
- [8] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118, 2017.
- [9] Terrence S Furey, Nello Cristianini, Nigel Duffy, David W Bednarski, Michel Schummer, and David Haussler. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16(10):906–914, 2000.
- [10] Samuel Hawkins, Hua Wang, Ying Liu, Alberto Garcia, Olya Stringfield, Henry Krewer, Qian Li, Dmitry Cherezov, Robert A Gatenby, Yoganand Balagurunathan, et al. Predicting malignant nodules from screening ct scans. *Journal of Thoracic Oncology*, 11(12):2120–2128, 2016.
- [11] Mario Fernando Jojoa Acosta, Liesle Yail Caballero Tovar, Maria Begonya Garcia-Zapirain, and Winston Spencer Percybrooks. Melanoma diagnosis using deep learning techniques on dermoscopic images. *BMC Medical Imaging*, 21(1):1–11, 2021.
- [12] Mert Ozcan and Serhat Peker. A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics*, 3:100130, 2023.