



УДК 004.8:61

РЕШЕНИЕ ЗАДАЧИ КЛАССИФИКАЦИИ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Карпушкина Ирина Сергеевна

Студент магистратуры

2 курс, факультет «Информационные технологии»

Кафедра «Математическая кибернетика и информационные технологии»

Московский технический университет связи и информатики

e-mail: karpushkinais@yandex.ru

Скоморохина Екатерина Романовна

Студент бакалавриата

4 курс, факультет «Информационные технологии»

Кафедра «Сетевые информационные технологии и сервисы»

Московский технический университет связи и информатики

email: skomorokhina98@mail.ru

Чикенев Станислав Дмитриевич

Студент бакалавриата

4 курс, факультет «Информационные технологии»

Кафедра «Сетевые информационные технологии и сервисы»

Московский технический университет связи и информатики

email: chickenev2001@mail.ru

Аннотация

В данной статье описывается теоретическое решение задачи классификации медицинских изображений с помощью методов машинного обучения. Для этого предлагается использовать комбинацию методов опорных векторов (SVM) и метода главных компонент (PCA), которые позволяют классифицировать изображения на основе их признаков, извлеченных из исходных данных. В работе рассматриваются основные преимущества и ограничения каждого из методов машинного обучения, а также представлена методика построения модели и ее применения. Описанный подход позволяет достигнуть высокой точности классификации медицинских изображений и может быть применен в области диагностики онкологических заболеваний и других медицинских проблем, связанных с анализом изображений.

Ключевые слова: машинное обучение, SVM, PCA, классификация медицинских изображений, обработка изображений, диагностика онкологических заболеваний.

SOLVING THE PROBLEM OF CLASSIFICATION OF MEDICAL IMAGES USING MACHINE LEARNING METHODS

Irina S. Karpushkina

Master's degree student

2nd year, faculty "Information technologies"

Department of Mathematical Cybernetics and Information Technologies

Moscow Technical University of Communications and Informatics

Ekaterina R. Skomorokhina

Bachelor

4th year, faculty "Information technologies"

Department of "Network Information Technologies and Services"

Moscow Technical University of Communications and Informatics

Stanislav D. Chikenev

Bachelor

4th year, faculty "Information technologies"

Department of "Network Information Technologies and Services"

Moscow Technical University of Communications and Informatics

ABSTRACT

This article describes a theoretical solution to the problem of classifying medical images using machine learning methods. A combination of support vector machines (SVM) and principal component analysis (PCA) is proposed to classify images based on their features extracted from the original data. The main advantages and limitations of each machine learning method are discussed, and a methodology for model construction and application is presented. The proposed approach achieves high accuracy in classifying medical images and can be applied in the field of oncological disease diagnosis and other medical problems related to image analysis.

Keywords: machine learning, SVM, PCA, medical image classification, image processing, oncological disease diagnosis.

Введение

Общее описание проблемы

В последние годы наблюдается увеличение числа обращений за медицинской помощью в связи с онкологическими заболеваниями [1]. Одним из основных факторов, который затрудняет своевременную диагностику и лечение рака, является сложность интерпретации медицинских изображений. Современные методы диагностики и лечения онкологических заболеваний требуют точной и быстрой обработки медицинских изображений, что может быть достигнуто с помощью методов машинного обучения [2].

Необходимость решения этой проблемы с помощью методов машинного обучения

Методы машинного обучения позволяют ускорить и улучшить обработку медицинских изображений, что повышает точность диагностики и эффективность лечения онкологических заболеваний [3, 4]. В настоящее время разработаны множество алгоритмов и методов машинного обучения, которые успешно применяются в медицине для решения различных задач, связанных с диагностикой и лечением онкологических заболеваний [5].

Цель работы и задачи

Цель данной работы – исследовать возможность применения методов машинного обучения для улучшения обработки медицинских изображений при диагностике и лечении онкологических заболеваний. Для достижения этой цели поставлены следующие задачи:

Анализ существующих методов обработки медицинских изображений с использованием методов машинного обучения;

Разработка метода обработки медицинских изображений на основе выбранного метода машинного обучения;

Методы машинного обучения для решения проблемы

Методы машинного обучения для классификации медицинских изображений представлены в таблице 1.

Таблица 1. Сравнение выбранных методов машинного обучения для классификации медицинских изображений

Методы машинного обучения	Преимущества	Ограничения
SVM + PCA	Высокая точность классификации. Метод PCA позволяет сократить количество признаков.	При большом количестве признаков и изображений требуется значительное количество времени для обучения модели.
Нейронные сети	Могут обрабатывать большие объемы данных и обучаться на сложных паттернах.	Требуется большое количество данных для обучения и высокая вычислительная мощность для обучения и применения модели.
Деревья решений	Простота интерпретации полученной модели. Можно использовать для классификации в режиме реального времени.	Низкая точность классификации при использовании большого количества признаков.
К-ближайших соседей	Простота и скорость обучения модели. Можно использовать для классификации в режиме реального времени.	Требуется наличие предварительно размеченных данных для обучения модели. Низкая точность классификации при использовании большого количества признаков.

Из таблицы видно, что выбранные методы машинного обучения – SVM и PCA – имеют высокую точность классификации и позволяют сократить количество признаков для обработки изображений, что особенно важно для задачи классификации медицинских изображений, где имеется множество различных признаков. SVM хорошо работает с нелинейными данными и имеет высокую точность классификации, а PCA позволяет уменьшить размерность данных, сохраняя при этом максимально возможную информацию. Вместе эти методы позволяют эффективно обрабатывать медицинские изображения, имеющие сложные паттерны и большое количество признаков. Нейронные сети могут использоваться для классификации медицинских изображений, но требуют большого количества данных и вычислительных ресурсов, что может быть ограничивающим фактором в реальных условиях. Деревья решений могут быть использованы для классификации медицинских изображений, но их точность может быть низкой при использовании большого количества признаков. К-ближайших соседей могут

быть использованы для классификации медицинских изображений, но точность модели может быть низкой при использовании большого количества признаков. Кроме того, для обучения модели требуется наличие предварительно размеченных данных, что может быть проблематично в случае отсутствия таких данных или их ограниченного объема. В связи с этим, для данной задачи был выбран метод SVM в сочетании с PCA, который позволяет эффективно обрабатывать медицинские изображения, имеющие сложные паттерны и большое количество признаков.

Анализ данных

Описание используемых данных и их источников

Для проведения исследования будет использоваться набор медицинских изображений, полученных с помощью КТ. Набор данных состоит из 500 изображений, каждое из которых представлено в виде трехмерного массива пикселей.

Источником данных является база данных медицинских изображений, предоставленная Институтом медицинской физики и биоинженерии при Национальном институте здравоохранения США.

Предварительный анализ данных

Перед началом работы с данными необходимо провести их предварительный анализ, включающий в себя оценку качества изображений, проверку наличия пропущенных значений и выбросов, а также проверку распределения данных.

Очистка данных

После проведения предварительного анализа необходимо произвести очистку данных от пропущенных значений и выбросов, а также устранить шум на изображениях с помощью фильтрации.

Обработка данных перед применением методов машинного обучения

Для применения методов машинного обучения необходимо произвести предварительную обработку данных, включающую в себя приведение изображений к стандартному размеру, нормализацию значений пикселей и извлечение признаков с помощью метода главных компонент (Principal Component Analysis, PCA).

Теоретическое решение проблемы

Описание выбранного теоретического решения проблемы

Для решения проблемы классификации медицинских изображений предлагается использовать SVM в сочетании PCA.

Метод опорных векторов будет использоваться для классификации изображений на основе их признаков, полученных с помощью метода главных компонент.

Построение модели и методика ее применения

Для построения модели будет использоваться библиотека машинного обучения scikit-learn на языке программирования Python. Модель будет обучаться на наборе данных, описанных в разделе «Анализ данных», с помощью SVM и PCA.

Методика применения модели заключается в следующих шагах:

Подготовка медицинского изображения для обработки, включающая в себя приведение изображения к стандартному размеру и нормализацию значений пикселей.

Извлечение признаков из изображения с помощью PCA, а классификация изображения на основе извлеченных признаков с помощью SVM.

После обучения модели необходимо провести ее тестирование на отложенной выборке. Для оценки качества модели необходимо использовать следующие метрики:

1. Точность (Accuracy);
2. Точность (Precision);
3. Полнота (Recall);
4. F1-мера (F1-score).

Заключение

В данной статье были рассмотрены методы машинного обучения для решения задачи классификации медицинских изображений. Была предложена методика, основанная на комбинации методов SVM и PCA. SVM использовался для классификации изображений на основе их признаков, а PCA позволял уменьшить размерность данных, сохраняя при этом максимально возможную информацию.

Предложенная методика, основанная на комбинации методов SVM и PCA, может быть эффективным теоретическим решением проблемы классификации медицинских изображений. SVM позволяет обрабатывать медицинские изображения, имеющие сложные паттерны и большое количество признаков, в то время как PCA позволяет уменьшить размерность данных, сохраняя при этом максимально возможную информацию.

Список литературы:

1. Ян Лекун. Как учится машина. Революция в области нейронных сетей и глубокого обучения. (Библиотека Сбера: Искусственный интеллект). – М.: Альпина нон-фикшн, 2021. – ISBN 978-5-907394-29-2.
2. Андреас, Мюллер Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными: моногр. / Мюллер Андреас. - М.: Альфа-книга, 2017. - 697 с.
3. Домингос, П. Верховный алгоритм. Как машинное обучение изменит наш мир / П. Домингос. - М.: Манн, Иванов и Фербер, 2016. - 190 с.
4. Волков, Г. А. Исторические аспекты развития методов распознавания медицинских изображений / Г. А. Волков, К. Р. Волкова // Наука. Исследования. Практика : Сборник избранных статей по материалам Международной научной конференции, Санкт-Петербург, 25 апреля 2020 года. – Санкт-Петербург: Частное научно-образовательное учреждение дополнительного профессионального образования Гуманитарный национальный исследовательский институт «НАЦРАЗВИТИЕ», 2020. – С. 109-111. – EDN VGILPK.
5. Усов, А. А. Машинное обучение в задачах медицинской диагностики / А. А. Усов // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем : Материалы Всероссийской конференции с международным участием, Москва, 18–22 апреля 2022 года. – Москва: Российский университет дружбы народов (РУДН), 2022. – С. 430-432. – EDN OXNXIC.

References:

1. Jan LeCun. How the machine learned. A revolution in neural networks and deep learning. (Sber Library: Artificial Intelligence). – М.: Alpina Nonfiction, 2021. – ISBN 978-5-907394-29-2.
2. Andreas, Müller Introduction to Machine Learning with Python. A guide for data scientists: monograph. / Müller Andreas. - М.: Alfa-kniga, 2017. - 697 p.
3. Domingos, P. Master algorithm. How machine learning will change our world / P. Domingos. - М.: Mann, Ivanov and Ferber, 2016. - 190 p.
4. Volkov G.V. A. Volkov G. A., Volkova K. R. Historical aspects of the analysis of detected images // Nauka. research. Practice: Collection of selected articles based on the materials

of the international scientific conference, St. Petersburg, April 25, 2020. - St. Petersburg: Private scientific and educational institution of professional additional education Humanitarian National Research Institute "NATIONAL RAZVITIE", 2020. - P. 109-111. - EDN VGILPK.

5. Usov, A. A. Machine learning in the problems of medical diagnostics / A. A. Usov // Information and telecommunication technologies and mathematical modeling of high-tech systems: Proceedings of the All-Russian Conference with the participation of RIAC, Moscow, April 18–22, 2022. - Moscow: Peoples' Friendship University of Russia (RUDN), 2022. - P. 430-432. — EDN OXNIX.