

УДК 004.85

ОШИБКИ В РАСПОЗНАВАНИИ МЕДИЦИНСКИХ СНИМКОВ ГОЛОВНОГО МОЗГА ПРИ ОБУЧЕНИИ МОДЕЛИ

Петров Артем Ильич,

бакалавр

Калужский филиал федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»
artemchik-2000@mail.ru

Белов Юрий Сергеевич,

доцент

Калужский филиал федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»
ysbelov@bmstu.ru

Аннотация

С использованием современных методов машинного обучения и глубокого обучения возможно автоматизированное распознавание опухолей головного мозга на медицинских изображениях. Однако при разработке и обучении моделей машинного обучения могут возникнуть определенные ошибки, которые важно учитывать для повышения точности диагностики и эффективности лечения.

Исследование фокусируется на выявлении проблемы переобучения модели в процессе классификации опухолей головного мозга при использовании методов машинного обучения. Анализируются потенциальные источники ошибок в обучении модели, включая факторы, такие как нестабильность, недостаточность данных, неправильное представление классов, недообучение и выбор неоптимальных алгоритмов. Рассматриваются различные методы визуализации и анализа результатов обучения, включая графики потерь, точности, кривые ROC и матрицы ошибок. Обсуждаются стратегии и подходы к улучшению качества классификации опухолей головного мозга при помощи современных методов машинного обучения. Предоставляются рекомендации в области медицинского образования и машинного обучения с целью оптимизации процесса диагностики опухолей головного мозга и повышения точности прогнозирования.

Ключевые слова: классификация медицинских изображений, обучение модели, ошибки при обучении.

ERRORS IN RECOGNIZING MEDICAL IMAGES OF THE BRAIN DURING MODEL TRAINING

Petrov Artyom Ilich,

Federal State Budgetary Educational Institution of Higher Education «Bauman Moscow State Technical University» (Kaluga Branch)

Belov Yuri Sergeevich,

Federal State Budgetary Educational Institution of Higher Education «Bauman Moscow State Technical University» (Kaluga Branch)

ABSTRACT

Using modern methods of machine learning and deep learning, automated recognition of brain tumors in medical images is possible. However, certain errors may occur during the development and training of machine learning models, which are important to take into account in order to improve the accuracy of diagnosis and the effectiveness of treatment.

The study focuses on identifying the problem of retraining the model in the process of classifying brain tumors using machine learning methods. Potential sources of errors in model training are analyzed, including factors such as instability, insufficient data, incorrect representation of classes, under-learning and the choice of suboptimal algorithms. Various methods of visualization and analysis of learning outcomes are considered, including loss graphs, accuracy graphs, ROC curves and error matrices. Strategies and approaches to improving the quality of classification of brain tumors using modern machine learning methods are discussed. Recommendations in the field of medical education and machine learning are provided in order to optimize the process of diagnosing brain tumors and improve the accuracy of forecasting.

Keywords: classification of medical images, model training, learning errors.

Распознавание опухолей головного мозга с помощью машинного обучения является перспективным направлением в медицине, которое может привести к более точной диагностике и улучшению результатов лечения пациентов [1].

Однако, как и в любой системе машинного обучения [2], при обучении модели для распознавания опухолей головного мозга могут возникать ошибки, которые могут привести к неверным диагнозам.

Одна из наиболее распространенных ошибок, с которой можно столкнуться при обучении модели для распознавания опухолей головного мозга, — это неустойчивость модели. Неустойчивая модель — это модель, которая очень чувствительна к небольшим изменениям в обучающих данных. Это может привести к тому, что модель будет хорошо работать на обучающих данных, но плохо на тестовых данных.

Существует несколько причин, по которым модель может быть неустойчивой [3].

- Если модель слишком сложная, она может начать запоминать обучающие данные, а не учиться на них.
- Если у модели недостаточно обучающих данных, она не сможет научиться всем необходимым паттернам.
- Если классы в обучающих данных не представлены должным образом, модель может не смочь их правильно распознать.

Существует несколько способов предотвращения неустойчивости модели:

- Использование регуляризации
- Использование кросс-валидации
- Использование более простых моделей

Недостаточный объем обучающих данных — это еще одна распространенная ошибка, которая может привести к ошибкам при распознавании опухолей головного мозга [4].

Если у модели недостаточно обучающих данных, она не сможет научиться всем необходимым паттернам, что может привести к тому, что она будет неправильно классифицировать новые данные.

Существует несколько признаков того, что модель обучена на недостаточном объеме данных:

- Низкая точность на тестовых данных
- Нестабильность обучения
- Появление артефактов

Существует несколько способов решить проблему недостаточного объема обучающих данных:

- Сбор большего количества данных
- Использование методов искусственного увеличения данных
- Использование переноса обучения
- Использование более простых моделей

Неправильное представление классов — это еще одна распространенная ошибка, которая может привести к ошибкам при распознавании опухолей головного мозга.

Если классы в обучающих данных не представлены должным образом, модель может не смочь их правильно распознать.

Существует несколько признаков того, что классы представлены неадекватно:

- Если модель хорошо работает на некоторых классах, но плохо на других, это означает, что эти классы представлены недостаточно хорошо.
- Если модель часто путает между собой разные классы, это означает, что между этими классами нет четкого различия.
- Если обучающие данные не представляют собой реальный мир, модель может не смочь правильно работать на реальных данных.

Существует несколько способов решить проблему неадекватного представления классов [5]:

- Сбор более сбалансированных данных
- Использование методов очистки данных
- Использование методов выборки
- Использование метрического обучения

Недостаточное обучение модели — это еще одна распространенная ошибка, которая может привести к ошибкам при распознавании опухолей головного мозга [6].

Если модель не обучена должным образом, она не сможет научиться всем необходимым паттернам, что может привести к тому, что она будет неправильно классифицировать новые данные.

Существует несколько признаков того, что модель обучена недостаточно:

- Если модель хорошо работает на обучающих данных, но плохо на тестовых данных, это означает, что она переобучена и не может обобщать новые данные.
- Если кривая обучения модели нестабильна и скачет вверх и вниз, это означает, что модель не может найти хорошую модель решения и, возможно, ей не хватает данных для обучения.
- Если на изображениях, на которых модель делает неправильные прогнозы, есть артефакты, это означает, что модель, возможно, пытается запомнить обучающие данные, а не учиться на них.

Существует несколько способов решить проблему недостаточного обучения модели:

- Увеличение количества эпох обучения может дать модели больше времени для обучения.
- Снижение скорости обучения может помочь модели сходиться к оптимальному решению.
- Использование различных методов оптимизации для обучения модели.
- Использование регуляризации, которая помогает предотвратить запоминание обучающих данных моделью.

Выбор неподходящего алгоритма машинного обучения может привести к ошибкам при распознавании опухолей головного мозга.

Существует множество различных алгоритмов машинного обучения, и каждый из них имеет свои сильные и слабые стороны [7]. Важно выбрать алгоритм, который подходит для конкретной задачи.

При выборе алгоритма машинного обучения для распознавания опухолей головного мозга следует учитывать следующие факторы:

- Тип задачи
- Размер и качество обучающих данных
- Вычислительные ресурсы

Примеры неподходящих алгоритмов [8]:

- Линейные модели

Линейные модели могут быть неэффективны для задач распознавания опухолей головного мозга, поскольку эти задачи часто являются нелинейными.

- К-ближайших соседей

Данный алгоритм может быть неэффективен для задач с большим количеством признаков, поскольку вычисление расстояния между точками данных может быть дорогостоящим.

- Деревья решений

Деревья решений могут быть переобучены на обучающих данных, что может привести к плохой производительности на тестовых данных.

Таким образом, были рассмотрены некоторые из возможных ошибок, которые могут возникнуть в процессе распознавания опухолей головного мозга при обучении модели машинного обучения.

Эти ошибки могут затруднять точное и надежное распознавание опухолей головного мозга, что имеет критическое значение для диагностики и лечения пациентов. Понимание

этих потенциальных проблем и усилия по их преодолению могут существенно улучшить эффективность моделей машинного обучения в области медицинского образования.

С постоянным развитием методов машинного обучения и тщательным подходом к функционированию моделей можно повысить качество диагностики и улучшить результаты лечения пациентов с опухолями головного мозга. Будущие исследования и разработки в этой области могут привести к улучшенным методам обучения моделей и более точному распознаванию опухолей головного мозга, что будет иметь значительное значение для медицинского сообщества и пациентов.

Список литературы:

1. Jakesh Bohaju, "Brain Tumor." Kaggle, 2020, doi: 10.34740/KAGGLE/DSV/1370629
2. D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, et al. Deep learning for health informatics IEEE J Biomed Health Inform, 21 (2017), pp. 4-21
3. M. Тоғаçar, B. Ergen, Z. Cömert, F. Özyurt A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models IRBM, 41 (4) (2020), pp. 212-222
4. Петров А. И., Белов Ю. С. Модели классификации медицинских изображений при трансферном обучении. 2023, <http://e-scio.ru/wp-content/uploads/2023/01/Петров-А.-И.-Белов-Ю.-С.pdf>
5. J. Pardede Department of informatics engineering, institut Teknologi Nasional Bandung, Bandung, Indonesia, B. Sitohang, S. Akbar, and M. L. Khodra, "Implementation of transfer learning using VGG16 on fruit ripeness detection Int. J. Intell. Syst. Appl., 13 (2) (2021), pp. 52-61
6. P. Xu, J. Zhao, J. Zhang Identification of intrinsically disordered protein regions based on deep neural network-VGG16 Algorithms, 14 (4) (2021), p. 107
7. Петров А. И., Белов Ю. С. Трансферное обучение в нейросетевой обработке данных медицинских изображений на основе сверточных сетей // E-Scio [Электронный ресурс]: Электронное периодическое издание «E-Scio.ru» – Эл № ФС77-66730
8. D. Jayadevappa, S. Srinivas Kumar, and D. Murty, "Medical image segmentation algorithms using deformable models: a review," IETE Technical review, vol. 28, no. 3, pp. 248-255, 2011.