

УДК 004.8

Афанасьев Александр Диомидович, доктор физико-математических наук профессор, Иркутский национальный исследовательский технический университет, г. Иркутск

Голубев Лев Игоревич, магистрант, Институт информационных технологий и анализа данных, Иркутский национальный исследовательский технический университет, г. Иркутск

ВЛИЯНИЕ АУГМЕНТАЦИИ ДАННЫХ НА КАЧЕСТВО КЛАССИФИКАЦИИ МЕДИЦИНСКИХ СНИМКОВ

Аннотация

В статье рассматривается влияние различных методов аугментации данных на качество классификации медицинских изображений с использованием архитектуры VGG16 и модифицированного классификатора. Целью исследования является сопоставление традиционных преобразований изображений и генеративных подходов на основе DCGAN в условиях ограниченного объема данных. Эксперименты проводились на двух наборах Alzheimer Detection с четырьмя уровнями когнитивных нарушений и BR35H для бинарной классификации опухолей головного мозга. Для традиционной аугментации использовались повороты, сдвиги, масштабирование, отражения и регулировка яркости. Генеративная аугментация реализована путем синтеза изображений для слабо представленных классов. Анализ метрик показал, что традиционные методы позволяют повысить точность моделей и снизить риск переобучения. Для Alzheimer Detection точность увеличилась с 0.88 до 0.91, для BR35H с 0.90 до 0.95. Использование DCGAN показало неоднозначный результат, что сопровождалось ростом устойчивости обучения, но снижением точности для отдельных классов из-за низкой реалистичности синтетических

изображений. Полученные данные подтверждают эффективность традиционной аугментации и указывают на необходимость дальнейшей оптимизации генеративных подходов для применения в классификации медицинских снимков.

Ключевые слова: аугментация данных, глубокое обучение, DCGAN, VGG-16, классификация изображений, болезнь Альцгеймера, опухоль мозга, переобучение.

Annotation

The article examines the impact of various data augmentation methods on the performance of medical image classification using the VGG16 architecture with a modified classifier. The study aims to compare traditional image transformations with generative approaches based on DCGAN under conditions of limited data availability. Experiments were conducted on two datasets: Alzheimer Detection, encompassing four levels of cognitive impairment, and BR35H, for binary classification of brain tumors. Traditional augmentation techniques included rotations, shifts, scaling, flips, and brightness adjustments. Generative augmentation was implemented by synthesizing images for underrepresented classes. Metric analysis demonstrated that traditional methods improved model accuracy and reduced the risk of overfitting. For the Alzheimer Detection dataset, accuracy increased from 0.88 to 0.91, and for BR35H, from 0.90 to 0.95. The use of DCGAN yielded mixed results, showing increased training stability but decreased accuracy for certain classes due to the low realism of synthetic images. The findings confirm the effectiveness of traditional augmentation and highlight the need for further optimization of generative approaches for medical image classification.

Keywords: data augmentation, deep learning, DCGAN, VGG-16, image classification, Alzheimer's disease, brain tumor, overfitting.

Введение

Современные методы машинного обучения, особенно глубокие нейронные сети, показывают высокую эффективность в решении задач

классификации изображений. Их способность автоматически выявлять сложные паттерны в данных способствует развитию интеллектуальных систем принятия решений и автоматизации процессов. Однако успешное применение этих моделей зачастую сталкивается с критическим ограничением — недостаточным объёмом размеченных данных. Это особенно актуально для медицины, где процесс сбора, подготовки и аннотирования снимков требует значительных временных затрат и участия квалифицированных специалистов.

Одним из наиболее перспективных и широко используемых подходов для преодоления дефицита данных является аугментация — процесс искусственного расширения обучающей выборки за счёт создания дополнительных примеров путём различных преобразований исходных изображений. В глубоком обучении применяются как классические методы аугментации такие как повороты, отражения, масштабирование и изменение яркости, так и более сложные, в частности, генерация новых изображений с помощью генеративно-сопоставительных сетей (generative adversarial network, GAN) [1, 2].

В обзоре Chlap и соавт. [3] показано, что комбинация пространственных и цветовых преобразований способна повысить точность классификации рентгеновских снимков на 8–12%. Тем не менее данная работа носит обобщающий характер и практически не затрагивает современные генеративные методы. В свою очередь, исследование Perez и Wang [4] демонстрирует, что избыточное использование трансформаций может искажать ключевые признаки изображений, снижая диагностическую ценность моделей. Также следует подчеркнуть, что в большинстве работ, посвящённых аугментации медицинских изображений, отсутствует сравнение традиционных методов и генеративных сопоставительных сетей (GAN) в едином контексте.

Таким образом, несмотря на наличие положительных результатов, влияние аугментации на качество классификации медицинских изображений остаётся предметом активного обсуждения и требует более глубокого

изучения. Это обусловлено тем, что чрезмерные или неподходящие трансформации способны исказить диагностически значимые признаки и ухудшить итоговые результаты.

Целью настоящего исследования является анализ влияния различных методов аугментации данных на качество классификации медицинских изображений. В рамках исследования будет проведено сравнение традиционных техник аугментации и GAN-аугментации [5], с вниманием к применимости этих методов в условиях ограниченного объёма доступных данных.

Используемые наборы данных

В качестве материала для обучения нейросетевых моделей использовались два набора данных. Первый набор, Alzheimer Detection [6], включал 8663 МРТ-изображения головного мозга, которые были разделены на обучающую выборку с 6931 изображением, валидационную выборку с 1732 изображениями и тестовую выборку с 1103 изображениями. Классификация проводилась по четырём уровням когнитивных нарушений, от отсутствия патологий до умеренных нарушений. Пример распределения классов приведён на рисунке 1.

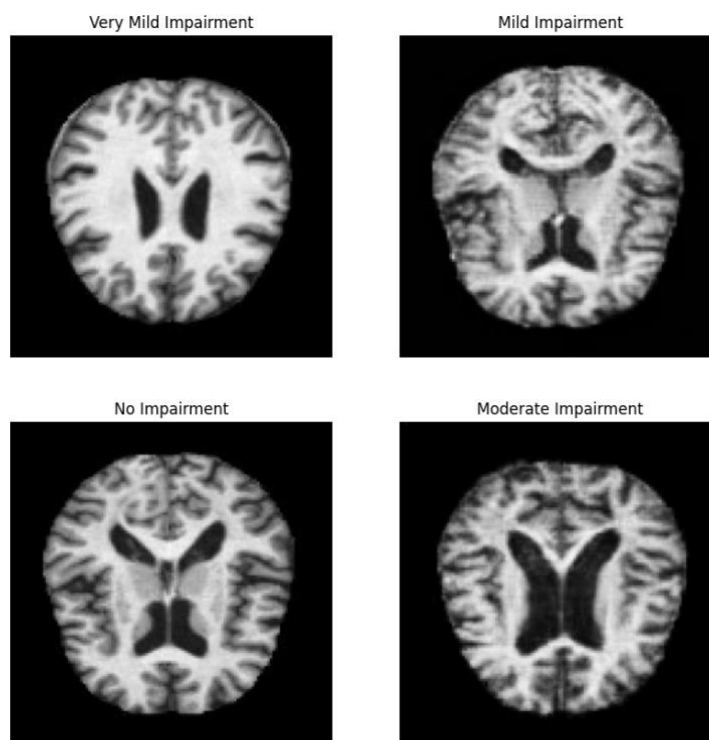


Рисунок 1 – Классы набора данных Alzheimer Detection

Второй набор данных, BR35H: Brain Tumor Detection [7], содержал 463 изображения, из которых 423 использовались для обучения, 40 для валидации и 20 для тестирования. Разметка была бинарной и отражала наличие опухоли головного мозга (YES) или её отсутствие (NO). Все изображения из первого и второго набора данных были приведены к размеру 224×224 пикселя, что соответствует входным параметрам используемой архитектуры [8]. Пример классов второго набора данных представлен на рисунке 2.

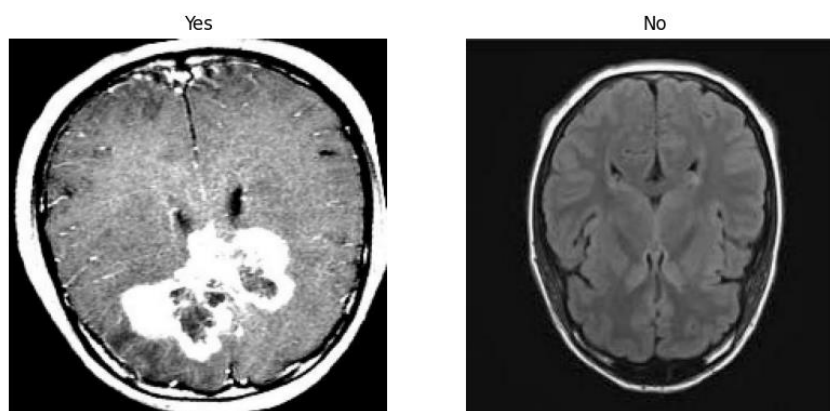


Рисунок 2 – Классы набора данных br35h

VGG16 с модифицированным классификатором

В качестве базовой модели для проведения экспериментов была выбрана сверточная нейронная сеть VGG16, обученная на большом наборе изображений ImageNet [9]. Так как стандартный классификатор VGG16 рассчитан на 1000 классов ImageNet, его финальная часть была заменена модифицированной архитектурой. Это позволило сосредоточить процесс адаптации сети на специфических особенностях медицинских данных, реализуя подход трансферного обучения [10]. Новый классификатор включал дополнительный сверточный слой с 256 фильтрами, функцию активации ReLU, слой нормализации по батчам и операцию MaxPooling. Для снижения риска переобучения применялся механизм Dropout (0,25). На полносвязной стадии добавлен слой из 256 нейронов с активацией ReLU, батч-нормализация и Dropout (0,5). Завершающий слой Dense с функцией активации softmax формирует вероятностное распределение по классам. Полная схема архитектуры нового классификатора представлена на рисунке 3.

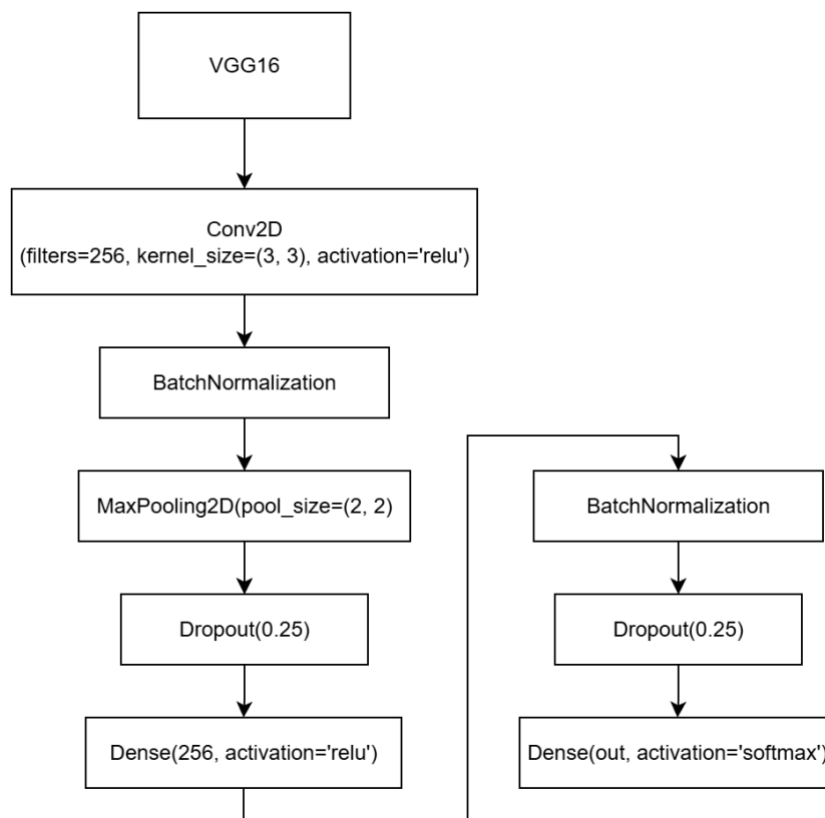


Рисунок 3 – Архитектура нового классификатора для VGG16

Для оптимизации модели использовался алгоритм Adam со стандартными настройками, функция потерь - sparse categorical crossentropy.

Основной метрикой качества была выбрана точность (Accuracy), показывающая долю правильных ответов. Однако при несбалансированных классах точность может быть недостаточно информативной, так как редкие категории рискуют остаться недооценёнными. Для более детального анализа применялась матрица ошибок, которая количественно оценивает распределение ошибок по классам и определяет долю ложноположительных и ложноотрицательных результатов на тестовой выборке.

Техники традиционной аугментации

Поскольку в задаче классификации медицинских изображений особенно важно сохранять их диагностическую достоверность, для аугментации использовались только небольшие преобразования, которые имитируют естественные вариации медицинских снимков, что позволяет сохранять их реалистичность [11].

Повороты до $\pm 5^\circ$ имитировали естественные изменения положения головы, горизонтальные и вертикальные сдвиги до 5% учитывали смещения пациента во время сканирования. Масштабирование и сдвиги по срезам компенсировали различия между моделями МРТ. Горизонтальное отражение повышало симметрию выборки, вертикальное не применяли из-за риска искажения анатомии. Корректировка яркости в пределах 90–110% учитывала вариации настроек оборудования. Пустые области после поворотов и сдвигов заполнялись значениями ближайших пикселей, что предотвращало артефакты.

Аугментация проводилась с помощью ImageDataGenerator: для каждого исходного изображения создавалось три новых со случайными комбинациями трансформаций, увеличивая выборку в четыре раза. Первый датасет расширен с 6931 до 27724 изображений, а второй с 423 до 1692. На рисунке 4 представлены отдельные примеры преобразований, с помощью которых осуществлялась аугментация исходных наборов данных.

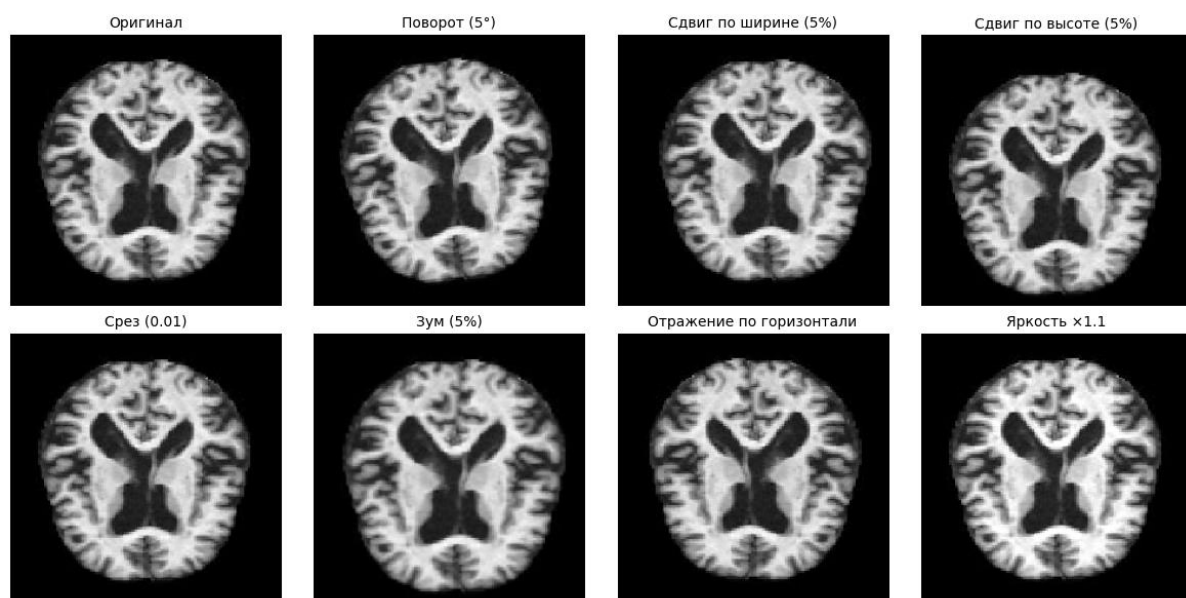


Рисунок 4 – Примеры используемых техник традиционной аугментации

Метод GAN-аугментации

Помимо традиционных методов применялась аугментация данных с использованием генеративных состязательных сетей, в частности архитектуры DCGAN (Deep Convolutional Generative Adversarial Network). Данная модель включает два взаимосвязанных компонента — генератор и дискриминатор.

Генератор получает на вход случайный вектор и преобразует его в изображение размером 224×224 пикселя. На первом этапе вектор проецируется в тензор небольшого пространственного разрешения и большой глубины, после чего изображение поэтапно увеличивается с помощью транспонированных свёрточных слоёв. Каждый блок масштабирования сопровождается пакетной нормализацией и функцией активации LeakyReLU, что обеспечивает устойчивость обучения и формирование реалистичных изображений. Финальный слой использует функцию активации гиперболический тангенс (\tanh), позволяя генерировать изображения с распределением, близким к исходным данным. Архитектура генератора представлена на рисунке 5.

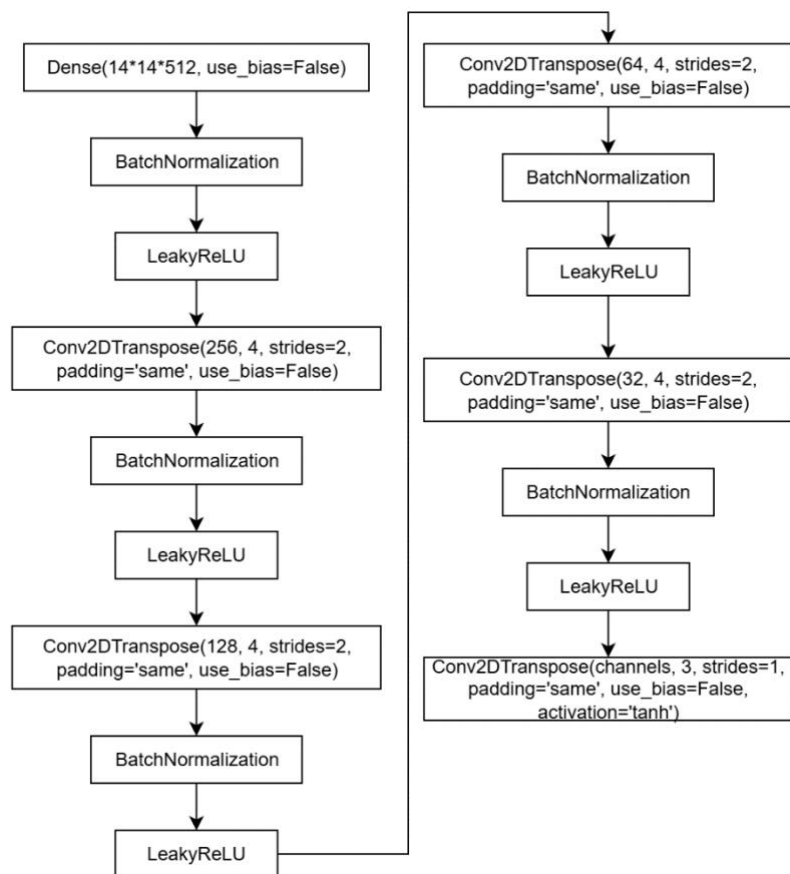


Рисунок 5 – Архитектура генератора

Дискриминатор принимает изображение и последовательно уменьшает его пространственные размеры с помощью свёрточных слоёв, используя LeakyReLU и Dropout для повышения устойчивости к переобучению. На выходе сеть возвращает вероятность принадлежности изображения к классу реальных. В процессе совместного обучения генератор стремится синтезировать изображения, способные «обмануть» дискриминатор, а дискриминатор обучается различать настоящие и сгенерированные данные. Архитектуру дискриминатора можно увидеть на рисунке 6.

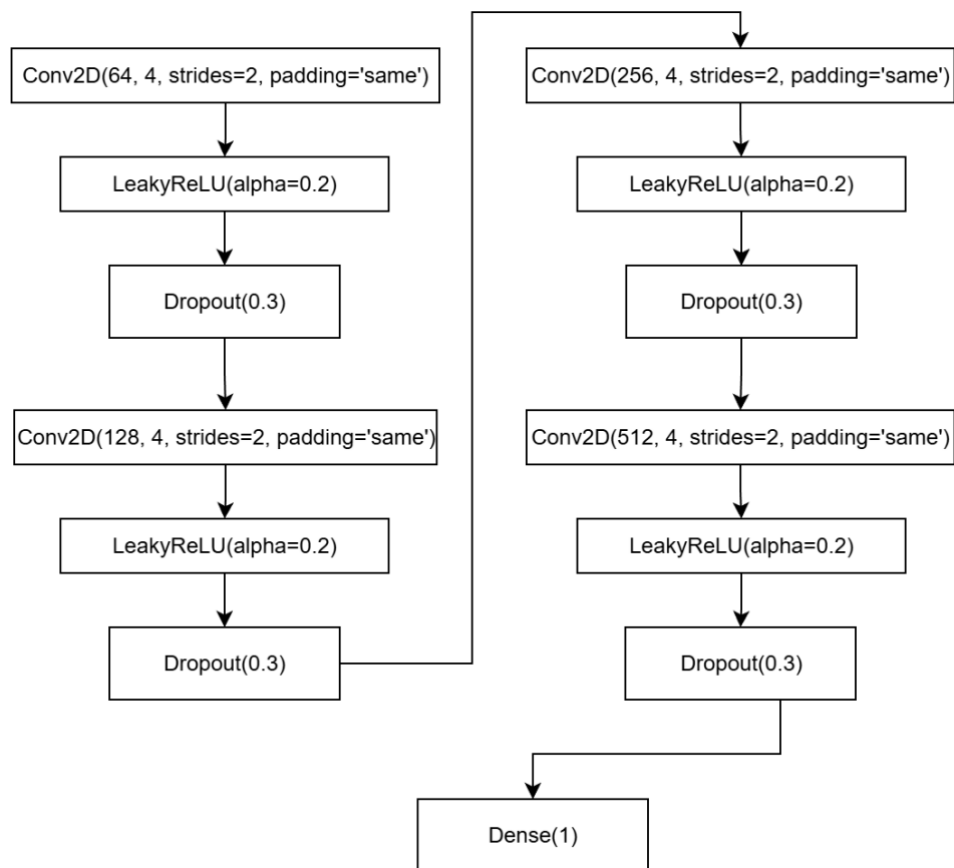


Рисунок 6 – Архитектура дискриминатора

Функции потерь определяются следующим образом. Для генератора использовалась бинарная кросс-энтропия (binary cross-entropy), направленная на максимизацию вероятности того, что дискриминатор классифицирует синтетические изображения как реальные. Для дискриминатора применялась сумма ошибок при классификации подлинных и искусственных изображений. Оптимизация весов обеих моделей выполнялась с помощью алгоритма Adam с параметрами, обеспечивающими стабильное обучение в течение 1000 эпох.

Результаты

В рамках исследования было проведено шесть экспериментальных прогонов для оценки влияния аугментации на качество классификации. Первый прогон реализовывал обучение на первом наборе данных без применения аугментации, второй на том же наборе с использованием традиционной аугментации. Третий и четвёртый прогоны были реализованы аналогичным образом для второго набора данных. Пятый прогон

предусматривал комбинированное использование традиционной и GAN аугментаций на первом наборе данных, шестой рассматривал аналогичный подход для второго набора.

На рисунке 7 представлены результаты обучения модели на первом наборе данных без аугментации. Левые графики отображают изменение метрики ассурасу (точности) и функции потерь на обучающей и валидационной выборках в зависимости от эпохи. Точность на валидации выросла с 0,75 на первой эпохе до 0,94 на двадцатой эпохе. Точность на обучении достигла значения около 0,99. После восьмой эпохи наблюдается расхождение кривых, что указывает на начало переобучения модели. Справа представлена матрица ошибок, где по горизонтали расположены предсказанные классы, а по вертикали истинные метки. Интенсивность цвета отражает количество образцов в каждой ячейке. На диагонали показано число правильно классифицированных изображений. Для класса No Impairment верно предсказаны 514 из 559 изображений, ошибки в основном приходятся на категорию Very Mild Impairment. Для класса Very Mild Impairment правильно определены 332 из 383 изображений, ошибки смещены к No Impairment. Класс Mild Impairment распознан верно в 117 из 150 случаев. Все 11 изображений класса Moderate Impairment классифицированы без ошибок.

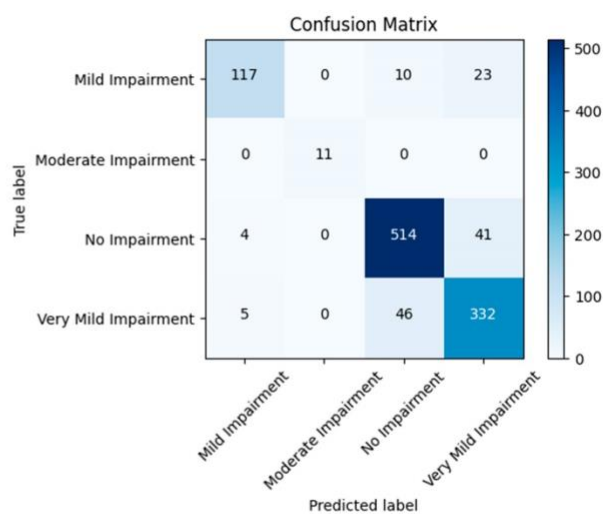
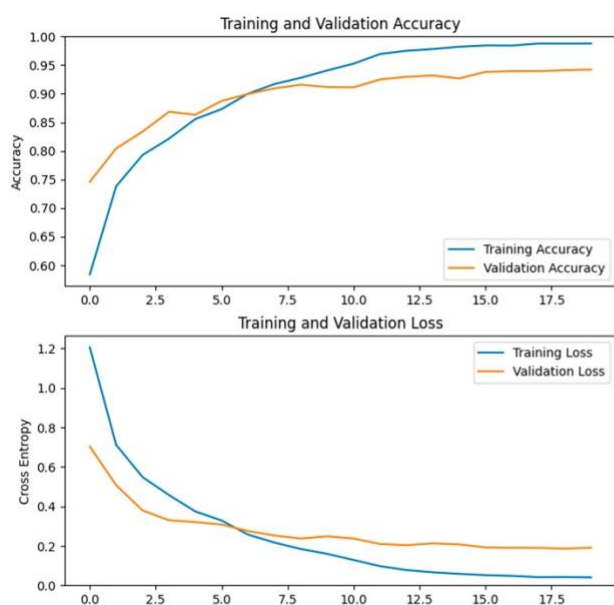


Рисунок 7 – Динамика обучения модели на первом датасете без аугментациями: графики точности и потерь (слева), матрица ошибок (справа)

Далее на рисунке 8 показаны результаты обучения с традиционной аугментацией. Точность на валидации выросла с 0,82 на первой эпохе до 0,98 на двадцатой. Точность на обучении достигла 0,99. Кривые точности начали двигаться параллельно, что говорит о снижении переобучения. Матрица ошибок подтверждает улучшение качества по сравнению с первым прогоном. Для класса No Impairment число верных предсказаний увеличилось с 514 до 535. Для класса Very Mild Impairment правильных ответов стало 340 вместо 332. Для Mild Impairment этот показатель вырос с 117 до 127.

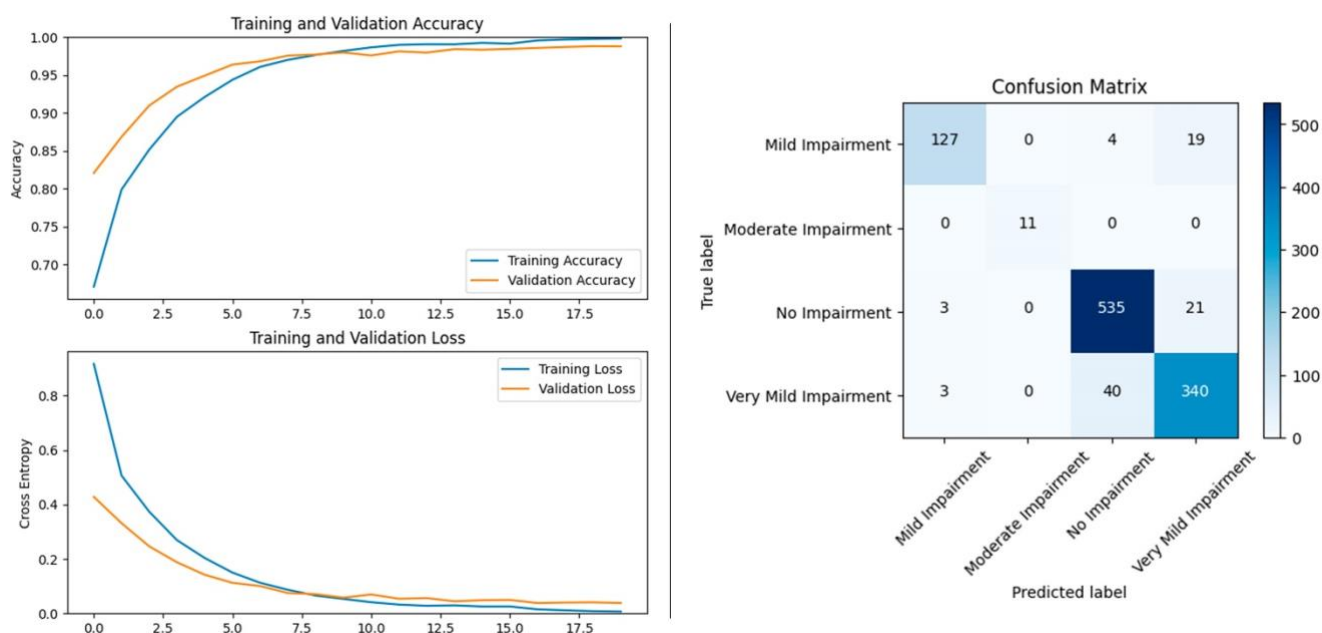


Рисунок 8 – Динамика обучения модели на первом датасете с аугментациями: графики точности и потерь (слева), матрица ошибок (справа)

На рисунке 9 представлена динамика обучения модели на втором наборе данных без аугментаций. В первые эпохи наблюдался уверенный рост

точности на обучающей выборке, при этом точность на валидационной выборке увеличилась с 0,80 на первой эпохе до 0,97 на третьей. Однако начиная с шестой эпохи темп роста замедлился, а после десятой эпохи стало заметно расхождение между Training Accuracy и Validation Accuracy, что указывает на формирование признаков переобучения. На матрице ошибок видно, что модель некорректно классифицировала два изображения из двадцати.

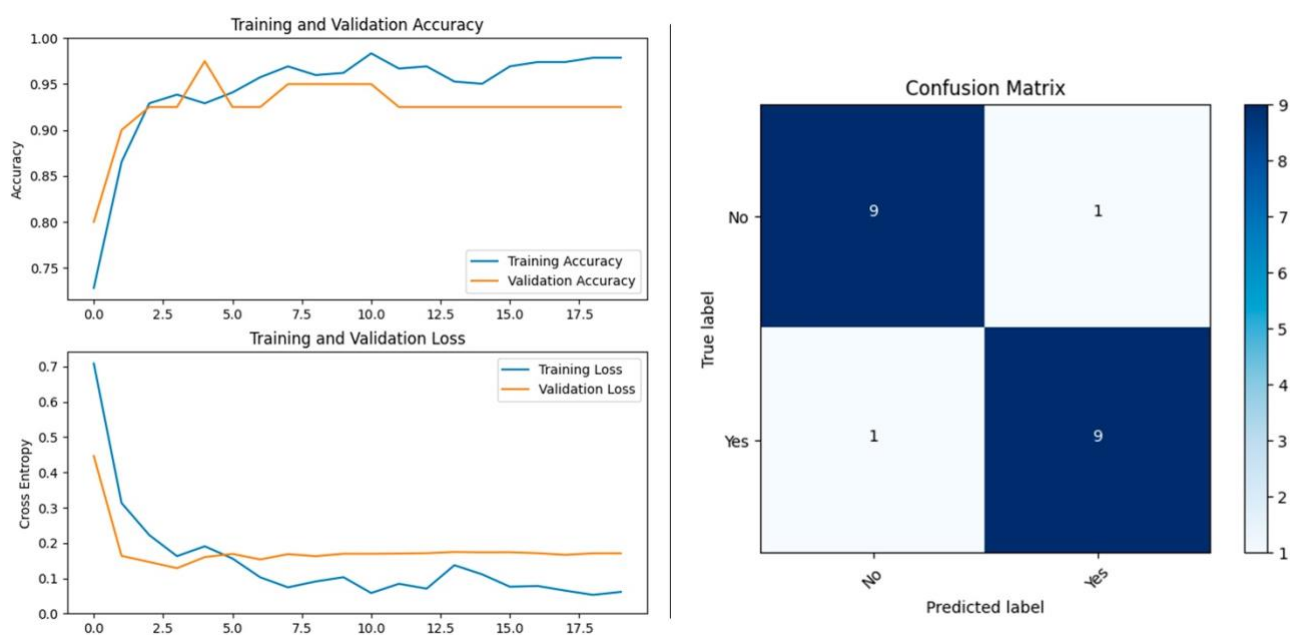


Рисунок 9 – Динамика обучения модели на втором датасете без аугментаций: графики точности и потерь (слева), матрица ошибок (справа)

На рисунке 10 представлены результаты обучения модели на втором наборе данных с применением традиционных аугментаций. В начальных эпохах точность на валидационной выборке уверенно росла с 0,85 до 0,95 к седьмой эпохе. После этого наблюдался резкий скачок, однако начиная с

восьмой эпохи на графике функций (Левый нижний график) потерь зафиксированы скачки увеличения ошибки Validation Loss, что сигнализирует о начале переобучения. Несмотря на некоторое снижение точности на валидации, на рисунке 14 видно, что на тестовой выборке модель справилась лучше. Она корректно классифицировала на одно изображение больше по сравнению с предыдущим прогоном.

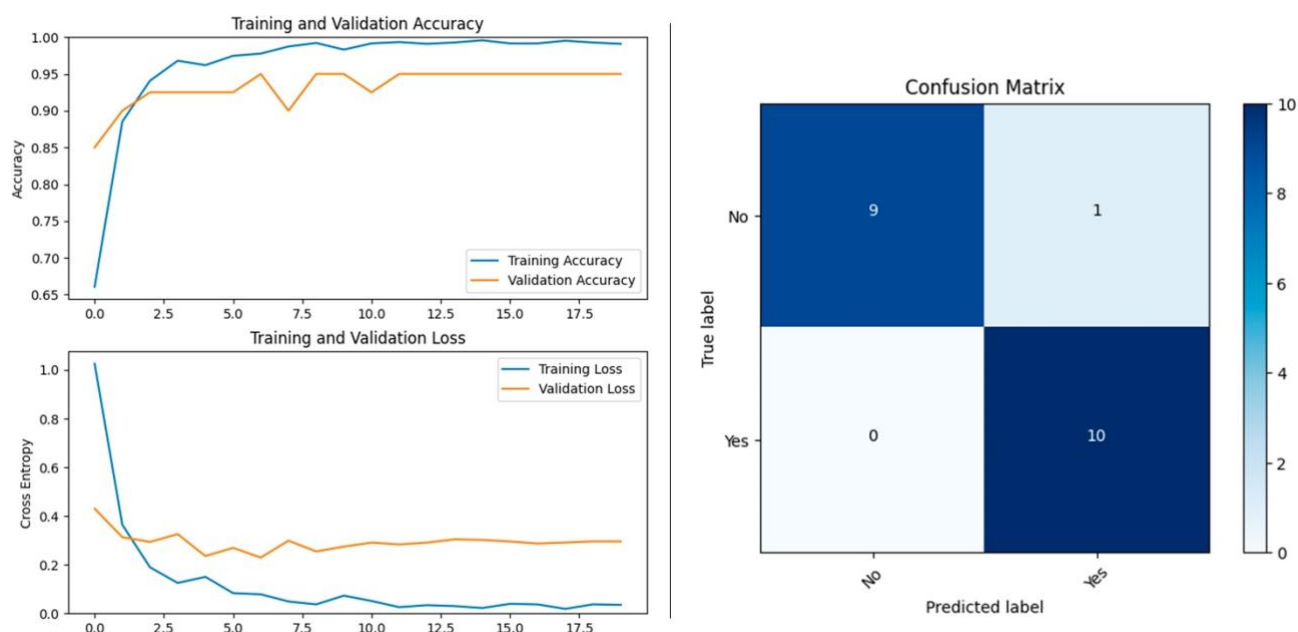


Рисунок 10 – Динамика обучения модели на втором датасете с аугментациями: графики точности и потерь (слева), матрица ошибок (справа)

Рассмотрим результаты применения генеративно-состязательных сетей (GAN) для аугментации изображений. Для каждого набора данных было сгенерировано по 261 синтетическому изображению для наиболее проблемных классов Very Mild Impairment в первом датасете и NO во втором. Эти изображения были добавлены к обучающей выборке с целью улучшения

представления редких или плохо предсказанных категорий и оценки влияния генеративной аугментации на качество классификации. На рисунке 11 можно увидеть пример сгенерированных изображений.

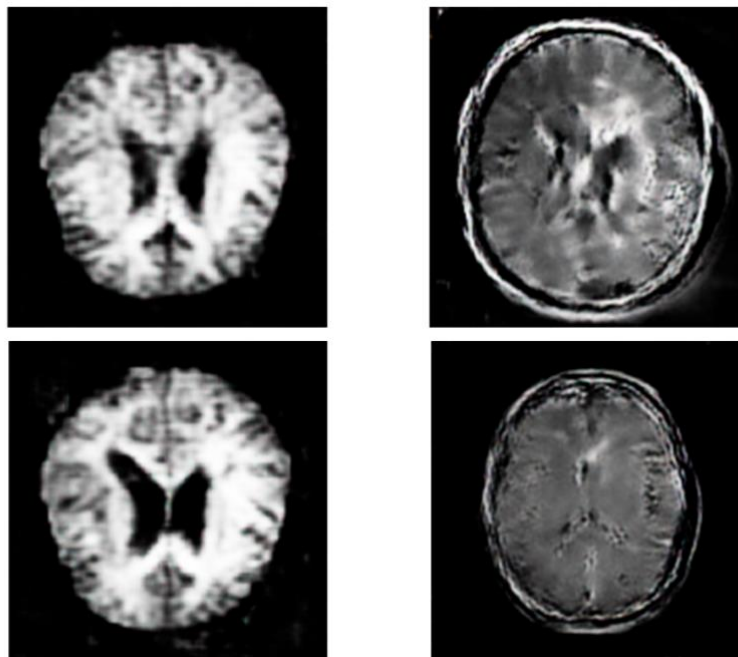


Рисунок 11 – Примеры сгенерированных изображений для первого и второго набора данных

На рисунке 12 показаны результаты обучения на первом наборе данных с комбинацией традиционной и GANаугментаций. Точность на валидации выросла с 0,84 до 0,97 к восьмой эпохе, после чего рост замедлился, а с десятой эпохи началось расхождение кривых Training Accuracy и Validation Accuracy, сигнализирующее о переобучении. Матрица ошибок справа демонстрирует следующие результаты. Для класса No Impairment верно предсказаны 547 из 559 изображений. Класс Very Mild Impairment распознан в 314 из 383 случаев, при этом 69 ошибок пришлось на категорию No Impairment. Класс Mild Impairment правильно определен в 135 из 150 случаев. Все 11 изображений класса Moderate Impairment классифицированы без ошибок. Снижение точности для Very Mild Impairment, вероятно, связано с недостаточно реалистичными сгенерированными изображениями.

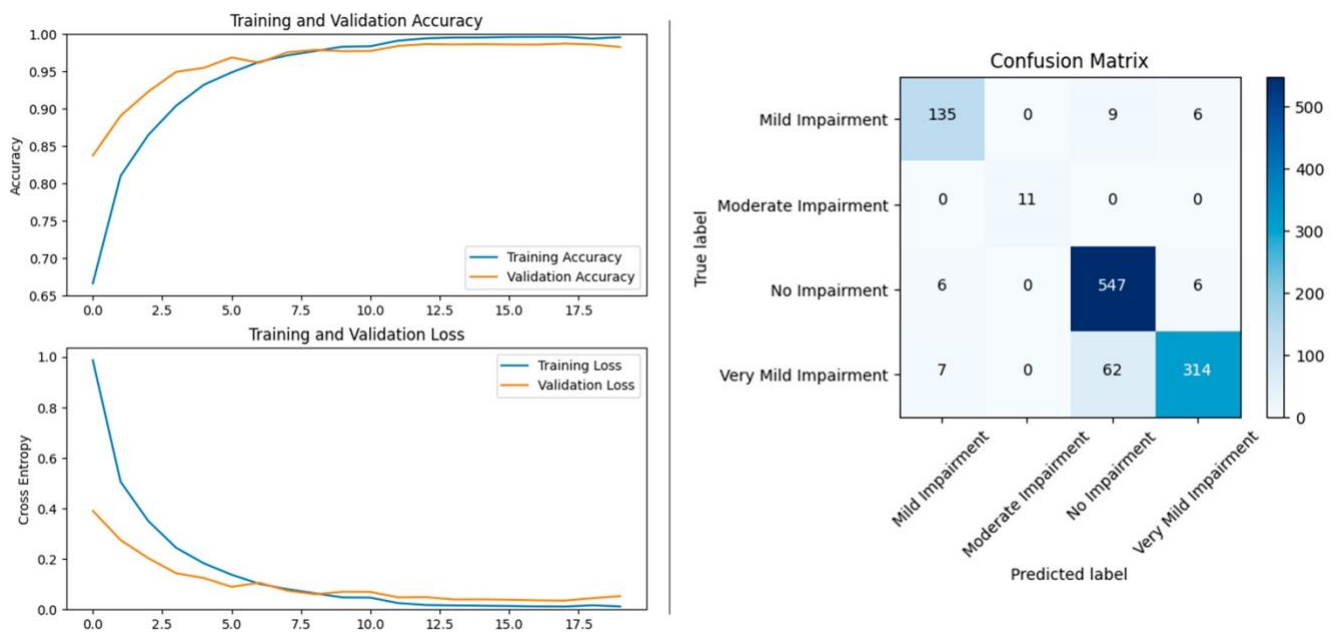


Рисунок 12 – Динамика обучения модели на первом датасете с GAN аугментацией: графики точности и потерь (слева), матрица ошибок (справа)

Рисунок 13 демонстрирует результаты обучения на втором наборе данных с традиционной и GAN-аугментацией. Которые продемонстрировали рост точности на валидационной выборке с 0,60 на первой эпохе до 0,975 к восьмой, после девятой эпохи рост замедлился. Параллельное снижение функций потерь Training Loss и Validation Loss подтверждает стабильность обучения. Матрица ошибок показывает, что модель верно классифицировала все 10 изображений класса YES, а для класса NO правильно предсказаны 9 из 10 образцов, что также соответствует результату традиционной аугментации (допущена одна ошибка). Даже с учётом этой единственной неточности общая динамика обучения свидетельствует о хорошей адаптации модели к данным. Стабильное снижение потерь и высокие показатели точности на валидации говорят о том, что модель эффективно обобщает закономерности и минимизирует риск переобучения.

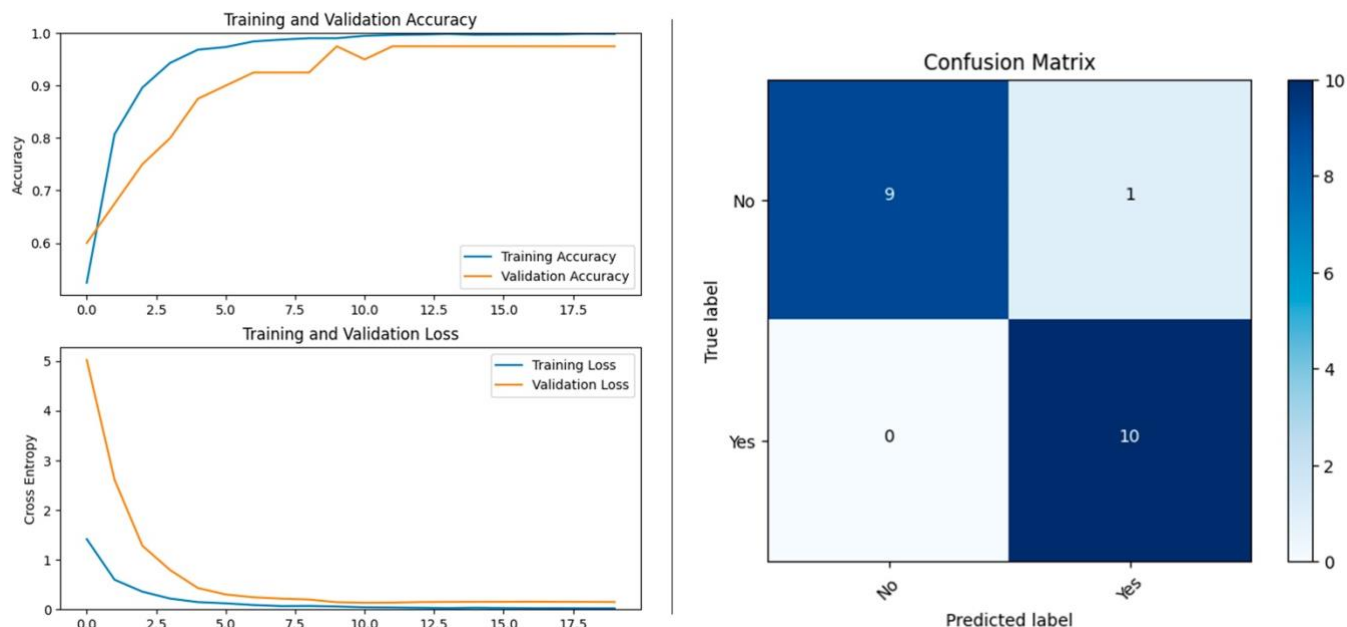


Рисунок 13 – Динамика обучения модели на втором датасете с GAN-аугментацией: графики точности и потерь (слева), матрица ошибок (справа)

В таблице 1 представлены значения точности (accuracy) на тестовых данных для всех прогонов модели при разных методах обработки данных. Для двух датасетов (Alzheimer Detection и BR35H: Brain Tumor Detection) показаны результаты без аугментации, со стандартной аугментацией и с комбинированной аугментацией (традиционные методы + GAN). Из таблицы видно, что применение аугментации повышает точность модели на обоих датасетах. Для Alzheimer Detection точность увеличилась с 0.88 до 0.91, а для BR35H — с 0.90 до 0.95. При этом комбинированная аугментация (традиционные методы + GAN) не дала дополнительного прироста по сравнению с обычной аугментацией. Вероятно, сгенерированные изображения либо мало отличаются от исходных данных, либо, как показал пятый прогон, содержат заметные артефакты, что снижает их полезность для обучения.

Таблица 1 – Accuracy(Точность) на всех прогонах модели

	Alzheimer Detection	BR35H: Brain Tumor Detection
Без аугментации	0.88	0.90
С аугментацией	0.91	0.95
Традиционная и GAN аугментация	0.91	0.95

Заключение

Проведённое исследование показало положительное влияние традиционных методов аугментации данных на качество классификации медицинских изображений с использованием модели VGG16. Применение поворотов, сдвигов, масштабирования и изменения яркости улучшило как Training Accuracy, так и Validation Accuracy, а также снизило риск переобучения. На тестовых данных точность для датасета Alzheimer Detection увеличилась с 0,88 до 0,91, а для BR35H: Brain Tumor Detection — с 0,90 до 0,95.

Использование генеративной аугментации на основе DCGAN дало смешанные результаты. Хотя синтетические изображения помогли расширить представление редких классов, для некоторых классов наблюдалось снижение точности, что указывает на недостаточную реалистичность части сгенерированных данных. В одном случае модель также ошибочно классифицировала изображение класса NO. Тем не менее анализ графиков точности и функции потерь показал повышение устойчивости модели к переобучению.

Таким образом, традиционная аугментация демонстрирует стабильное улучшение точности и качества обучения моделей при ограниченном объёме данных. Генеративные методы аугментации имеют потенциал, однако требуют дополнительной настройки и контроля качества синтезируемых изображений.

В дальнейшем целесообразно исследовать комбинированные подходы с оптимизацией гиперпараметров генеративных моделей и более строгой оценкой реалистичности изображений, а также проверить эффективность методов аугментации на других типах медицинских данных и архитектурах нейронных сетей.

Литература

1. Shorten C., Khoshgoftaar T. M. A survey on image data augmentation for deep learning // *Journal of Big Data*. – 2019. – Vol. 6, No. 1. – P. 1–48.
2. Frid-Adar M., Klang E., Amitai M., Goldberger J., Greenspan H. GAN-based synthetic medical image augmentation for improved CNN performance // *IEEE Access*. – 2018. – Vol. 6. – P. 55916–55931.
3. Chlap P., Min H., Vandenberg N., Dowling J. A., Holloway L., Haworth A. A review of medical image data augmentation techniques for deep learning applications // *Journal of Medical Imaging and Radiation Oncology*. – 2021. – Vol. 65, No. 5. – P. 545–563.
4. Wang J., et al. The effectiveness of data augmentation in image classification using deep learning // *Convolutional Neural Networks Vis. Recognit.* – 2017. – T. 11. – №. 2017. – С. 1-8.
5. Volpi R., Namkoong H., Sener O., Duchi J., Murino V., Savarese S. Generalizing to unseen domains via adversarial data augmentation // *Advances in Neural Information Processing Systems (NeurIPS)*. – 2018. – Vol. 31. – P. 5339–5349.
6. Roboflow. Alzheimer’s Disease Detection Dataset / URL: <https://universe.roboflow.com/alzheimer-h49wu/alzheimer-detection-ra7oh> (дата обращения: 12.05.2025). – Текст. Изображение : электронные.
7. Roboflow. Brain Tumor Detection BR35H Dataset / URL: <https://universe.roboflow.com/machine-learning-projects-5rkmx/br35h-brain-tumor-detection> (дата обращения: 15.05.2025). – Текст. Изображение: электронные.

8. Chen J., Liu Y., Wei S., Bian Z., Subramanian S., Carass A., Prince J. L., Du Y. A survey on deep learning in medical image registration: new technologies, uncertainty, evaluation metrics, and beyond // *Medical Image Analysis*. – 2025. – Vol. 100. – P. 103385. DOI: 10.1016/j.media.2024.103385.

9. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // *arXiv preprint arXiv:1409.1556*. – 2014.

10. Tajbakhsh N., Shin J. Y., Gurudu S. R., Hurst R. T., Kendall C. B., Gotway M. B., Liang J. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? // *IEEE Transactions on Medical Imaging*. – 2016. – Vol. 35, No. 5. – P. 1299–1312.

11. Shorten C., Khoshgoftaar T. M., Chowdhury S. Imbalanced medical image classification: Data augmentation approaches and performance evaluation // *Computers in Biology and Medicine*. – 2020. – Vol. 122. – P. 103803.

Literature

1. Shorten C., Khoshgoftaar T. M. A survey on image data augmentation for deep learning // *Journal of Big Data*. – 2019. – Vol. 6, No. 1. – P. 1–48.

2. Frid-Adar M., Klang E., Amitai M., Goldberger J., Greenspan H. GAN-based synthetic medical image augmentation for improved CNN performance // *IEEE Access*. – 2018. – Vol. 6. – P. 55916–55931.

3. Chlap P., Min H., Vandenberg N., Dowling J. A., Holloway L., Haworth A. A review of medical image data augmentation techniques for deep learning applications // *Journal of Medical Imaging and Radiation Oncology*. – 2021. – Vol. 65, No. 5. – P. 545–563.

4. Wang J., et al. The effectiveness of data augmentation in image classification using deep learning // *Convolutional Neural Networks Vis. Recognit.* – 2017. – T. 11. – №. 2017. – C. 1-8.

5. Volpi R., Namkoong H., Sener O., Duchi J., Murino V., Savarese S. Generalizing to unseen domains via adversarial data augmentation // *Advances in*

Neural Information Processing Systems (NeurIPS). – 2018. – Vol. 31. – P. 5339–5349.

6. Roboflow. Alzheimer's Disease Detection Dataset / URL: <https://universe.roboflow.com/alzheimer-h49wu/alzheimer-detection-ra7oh> (Accessed: 12.05.2025). – The text. Image: electronic.

7. Roboflow. Brain Tumor Detection BR35H Dataset / URL: <https://universe.roboflow.com/machine-learning-projects-5rkmx/br35h-brain-tumor-detection> (Accessed: 15.05.2025). – The text. Image: electronic.

8. Chen J., Liu Y., Wei S., Bian Z., Subramanian S., Carass A., Prince J. L., Du Y. A survey on deep learning in medical image registration: new technologies, uncertainty, evaluation metrics, and beyond // *Medical Image Analysis*. – 2025. – Vol. 100. – P. 103385. DOI: 10.1016/j.media.2024.103385.

9. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // *arXiv preprint arXiv:1409.1556*. – 2014.

10. Tajbakhsh N., Shin J. Y., Gurudu S. R., Hurst R. T., Kendall C. B., Gotway M. B., Liang J. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? // *IEEE Transactions on Medical Imaging*. – 2016. – Vol. 35, No. 5. – P. 1299–1312.

11. Shorten C., Khoshgoftaar T. M., Chowdhury S. Imbalanced medical image classification: Data augmentation approaches and performance evaluation // *Computers in Biology and Medicine*. – 2020. – Vol. 122. – P. 103803.