

Пилипенко Владислав Вадимович, магистрант,
Белгородский государственный национальный
исследовательский университет, Белгород

Михелев Владимир Михайлович, к.т.н.,
Белгородский государственный национальный
исследовательский университет, Белгород

ЭФФЕКТИВНОСТЬ РАСПОЗНАВАНИЯ И МНОГОМЕТОЧНОЙ КЛАССИФИКАЦИИ ЗАБОЛЕВАНИЙ ГЛАЗ ПРИ ИСПОЛЬЗОВАНИИ РАЗЛИЧНЫХ МОДЕЛЕЙ НЕЙРОННЫХ СЕТЕЙ

Аннотация: в рамках данной работы исследуется эффективность нейронных сетей для многометочной классификации офтальмологических заболеваний по изображениям глазного дна. Проанализированы шесть архитектур и три оптимизатора на датасете из 2164 изображений с 5-кратной кросс-валидацией. Выявлены наиболее точные модели и сложности при распознавании отдельных патологий.

Ключевые слова: нейронные сети, офтальмология, многометочная классификация, глубокое обучение, медицинские изображения.

Офтальмологические патологии остаются одной из ключевых медицинских проблем, влияющих на зрительные функции и общее состояние здоровья населения. Согласно статистическим данным, нарушения зрения затрагивают значительную часть глобального сообщества, подчеркивая необходимость внедрения инновационных методов диагностики. Использование методов машинного обучения, в частности нейронных сетей, предоставляет возможность оптимизировать процессы выявления заболеваний, обеспечивая более быструю и точную оценку состояния глазного дна. Настоящее исследование посвящено анализу вычислительного эксперимента, направленного на оценку эффективности различных архитектур нейронных сетей при решении задачи многометочной классификации глазных патологий на основе датасета из 2164 изображений глазного дна. Целью работы является сравнение производительности моделей с использованием трёх оптимизаторов и выявление факторов, влияющих на точность распознавания сложных офтальмологических заболеваний.

В качестве исходных данных использовался датасет, полученный с платформы Kaggle, в котором были исключены редкие классы заболеваний и включающий 2164 изображения (рисунк 1) глазного дна с мультитейбл-аннотациями по пяти классам: диабетическая, макулярный разрыв, диабетическая миопия, тесселированный фон сетчатки и экскавация диска зрительного нерва. Исключение классов с долей менее 5% от общего числа изображений позволило снизить дисбаланс данных, уменьшить риск переобучения и повысить надёжность классификации.



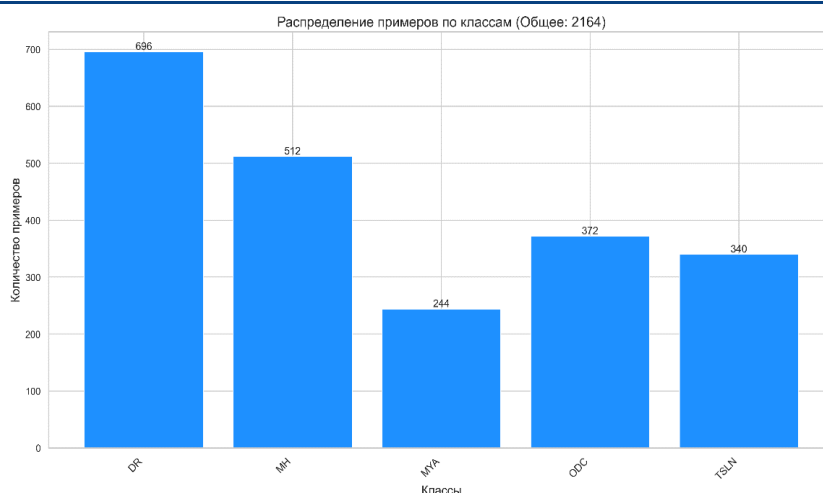


Рис. 1 Распределение изображений по классам

Данные были разделены на обучающую (80%) и валидационную (20%) выборки. Для оценки обобщающей способности моделей применялась 5-кратная кросс-валидация, при которой каждый из пяти поднаборов последовательно использовался в качестве валидационного.

Предобработка изображений была направлена на приведение входных данных к требованиям архитектур и на снижение чувствительности моделей к шумам и вариативности. В процессе обучения применялись аугментации: случайные повороты до 15° , горизонтальное и вертикальное отражение с вероятностью 50%, а также изменение яркости, контрастности, насыщенности и оттенка в пределах 20%. Эти преобразования моделировали реалистичные вариации условий получения медицинских изображений и способствовали улучшению устойчивости моделей к визуальному разнообразию. Дополнительно изображения подвергались случайной обрезке с масштабированием от 80% до 100% исходного размера, что позволяло фокусироваться на различных участках глазного дна. На этапе валидации использовались только масштабирование до фиксированного размера, преобразование в тензоры и нормализация. Размер изображений составлял 224×224 пикселей для моделей ResNet, EfficientNet, VGG16, Swin Transformer и ConvNeXt и 299×299 пикселей для Inception V3.

В исследовании были проанализированы шесть архитектур нейронных сетей, выбранных за их разнообразие и доказанную эффективность в задачах анализа изображений:

1. ResNet (18 и 152): архитектуры с остаточными связями, устраняющими проблему исчезающего градиента. ResNet18 представляет собой компактную модель, оптимальную для ограниченных вычислительных ресурсов, тогда как ResNet152 обеспечивает высокую точность благодаря глубокой структуре.

2. EfficientNet (B0 и B7): архитектуры, оптимизирующие соотношение глубины, ширины и разрешения. EfficientNetB7 демонстрирует лучшие результаты за счёт более сложной конфигурации по сравнению с более лёгкой версией B0.

3. VGG16: модель с последовательными сверточными слоями, способная эффективно выделять локальные признаки, но обладающая высокой вычислительной сложностью.

4. Inception V3: использует Inception-модули для факторизации свёрток, что позволяет эффективно обрабатывать сложные визуальные паттерны при умеренной нагрузке на ресурсы.

5. Swin Transformer: архитектура с механизмом внимания в скользящих окнах, объединяющая локальную и глобальную обработку признаков, особенно актуальна для задач медицинской визуализации.



6. ConvNeXt: гибридный подход, сочетающий свёрточные сети и принципы трансформеров. Отличается улучшенной обобщающей способностью за счёт применения увеличенных ядер свёртки и современных методов нормализации.

Для адаптации моделей к задаче многометочной классификации были модифицированы выходные слои, а часть слоёв была разморожена для дообучения. Подробности приведены в таблице 1:

Таблица 1

Модификации выходных слоёв и размороженные слои моделей

Модель	Модификации выходного слоя	Размороженные слои
ResNet	Linear, BatchNorm1d, ReLU, Dropout (0.3)	layer4
EfficientNet	Dropout (0.5), Linear, SiLU, BatchNorm1d	Три блока
VGG16	Linear, ReLU, Dropout (0.5)	Четыре слоя
Inception V3	Linear, BatchNorm1d, ReLU, Dropout (0.4)	Mixed_7c
Swin Transformer	LayerNorm, Linear, GELU, Dropout (0.5)	Две стадии
ConvNeXt	LayerNorm, Linear, GELU, Dropout (0.2)	Три блока

Обучение проводилось с использованием оптимизаторов AdamW, Adam и Adagrad с начальной скоростью обучения 0.0001 и L2-регуляризацией весов $1e-4$. AdamW внедряет регуляризацию весов непосредственно в процесс оптимизации, что способствует более стабильной сходимости по сравнению с Adam, основанным на адаптивной оценке градиентов. Adagrad адаптирует скорость обучения для каждого параметра, но может уступать по эффективности на глубокой архитектуре.

Оценка производительности моделей осуществлялась по следующим метрикам:

Precision – доля правильно классифицированных положительных примеров среди всех предсказанных как положительные:

$$Specifity (Precision) = \frac{TP}{TP+FP} \quad (1)$$

Recall – доля верно определённых положительных примеров среди всех истинно положительных:

$$TPR (Recall) = \frac{TP}{TP+FN} \quad (2)$$

F1-Score – гармоническое среднее между точностью и полнотой:

$$F1 = 2 * \frac{precision * recall}{(precision + recall)} \quad (3)$$

Рассматривались как Macro-вариант (усреднение по классам), так и Micro (по всем предсказаниям), что позволило учесть как редкие, так и преобладающие классы. Дополнительно использовалась метрика ROC-AUC, отражающая соотношение между TPR и FPR:

$$AUC = \int_0^1 TPRd (FPR) \quad (4)$$

Значения ROC-AUC варьируются от 0.5 (случайная классификация) до 1.0 (идеальная), характеризую общее качество классификации на разных порогах.

Полученные данные в ходе проведения вычислительного эксперимента представлены на рисунке 2.



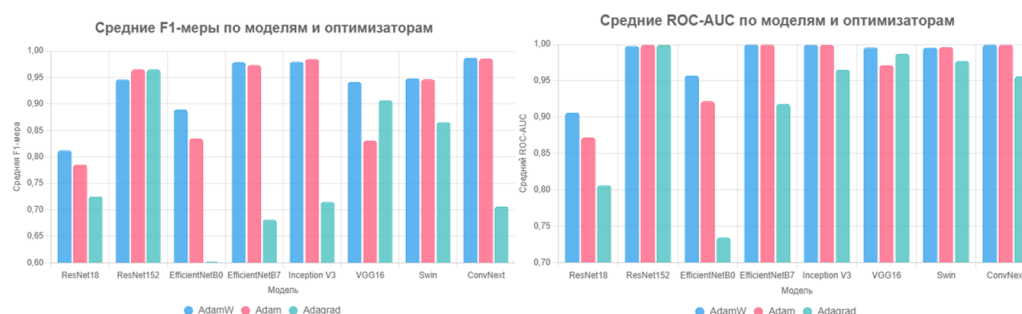


Рис. 2 Гистограммы F1-меры и ROC-AUC

Эксперимент выявил существенные различия в результатах, обусловленные как архитектурой моделей, так и выбором оптимизатора. При использовании AdamW модель ConvNeXt продемонстрировала выдающиеся показатели: $F1 \approx 0.987$, $Precision \approx 0.985$, $Recall \approx 0.990$ и $ROC-AUC \approx 0.999$. Inception V3 также показала высокую эффективность: $F1 \approx 0.979$, $Precision \approx 0.987$, $Recall \approx 0.971$, $ROC-AUC \approx 0.999$. ResNet152 и EfficientNetB7 обеспечили надёжные результаты ($F1 \approx 0.946$ и 0.959 соответственно), а VGG16 и Swin Transformer показали умеренную эффективность. Наименее успешными оказались ResNet18 и EfficientNetB0.

Оптимизатор Adam подтвердил высокую эффективность: ConvNeXt ($F1 \approx 0.981$) и Inception V3 ($F1 \approx 0.985$) вновь стали лидерами. Остальные модели показали результаты аналогичной динамики. При использовании Adagrad производительность существенно снижалась: наиболее устойчивыми остались ResNet152 и VGG16, в то время как ConvNeXt, EfficientNetB0 и ResNet18 показали значительное падение метрик.

Что же касается классового распознавания, то класс «экскавация диска зрительного нерва» оказался наиболее трудным для классификации, особенно при использовании Adagrad. В то же время диабетическая ретинопатия и макулярный разрыв классифицировались с высокой точностью, что может быть связано с ярко выраженными визуальными признаками. Диабетическая миопия и тесселированный фон сетчатки заняли промежуточное положение по уровню сложности.

Полученные результаты свидетельствуют о высокой чувствительности задачи многометочной классификации офтальмологических заболеваний к выбору архитектуры и оптимизатора. ConvNeXt и Inception V3 при оптимизации с помощью AdamW показали наивысшую эффективность, при этом ResNet152 и EfficientNetB7 также продемонстрировали хорошие показатели, но за счёт больших вычислительных затрат. Наименее эффективными оказались ResNet18 и EfficientNetB0, а использование Adagrad в большинстве случаев оказывалось неоптимальным. Особое внимание в дальнейшем должно быть уделено улучшению распознавания сложных патологий, таких как экскавация диска зрительного нерва, путём доработки архитектур и методов обучения.

Список литературы:

1. Ting D.S.W. Artificial intelligence and deep learning in ophthalmology [Электронный ресурс] / Ting D.S.W., Pasquale L. R., Peng L., et al. // British Journal of Ophthalmology. – 2019. – Vol. 103, No. 2. – pp. 167-175. – Режим доступа: <https://bj.o.bmj.com/content/103/2/167>
2. Grassmann, F. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography [Электронный ресурс] / F. Grassmann, J. Mengelkamp, C. Brandl, и др. // Ophthalmology. – 2018. – Vol. 125, № 9. – P. 1410–1420. – Режим доступа: <https://pubmed.ncbi.nlm.nih.gov/29653860/>



3. Sahlsten, J. Deep learning fundus image analysis for diabetic retinopathy and macular edema grading [Электронный ресурс] / J. Sahlsten, J. Jaskari, J. Kivinen, и др. // Sci. Rep. – 2019. – Vol. 9. – Article 10750. – Режим доступа: <https://pubmed.ncbi.nlm.nih.gov/31341220/>
4. Zhou H.H. Design of Intelligent Diagnosis and Treatment System for Ophthalmic Diseases Based on Deep Neural Network Model [Электронный ресурс] / Zhou H.H. // Contrast Media & Molecular Imaging. – 2022. – Vol. 2022. – Режим доступа: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9277203/>
5. Li Z. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma [Электронный ресурс] / Li Z., Wang Y., Yu J., Guo G. Y., Cao W. // Sci Rep. 2017. – Vol. 7 (1). – Режим доступа: <https://pubmed.ncbi.nlm.nih.gov/28710497/>

