

# Разработка гибридного интерпретируемого программного решения для диагностики диабетической ретинопатии по фундус-снимкам

Е. Н. Волков

Федеральный исследовательский центр  
«Информатика и управление»  
Российской Академии Наук

envolkoff@gmail.com

А. Н. Аверкин

Федеральный исследовательский центр  
«Информатика и управление»  
Российской Академии Наук

averkin2003@inbox.ru

**Аннотация.** Диабетическая ретинопатия является вторым по частоте заболеванием, ведущем к частичной или полной потере зрения. Применение технологий искусственного интеллекта в диагностике диабетической ретинопатии по фундус-снимкам вывело возможности клиницистов на совершенно новый уровень. Однако, из-за непрозрачности систем, основанных на применении искусственных нейронных сетей, доверие к результату их работы может быть снижено. Использование методов визуального объяснения результата работы нейронных сетей как моделей «чёрного ящика» может увеличить доверие как офтальмолога, так и пациента к результатам работы системы. Представлена реализация прототипа программного решения гибридной архитектуры на основе технологий компьютерного зрения, искусственных нейронных сетей, объяснительного искусственного интеллекта, нечёткой логики для диагностики диабетической ретинопатии по фундус-снимкам.

**Ключевые слова:** искусственный интеллект, объяснительный искусственный интеллект, объяснимый искусственный интеллект, искусственный нейронный сети, нейросети, инс, компьютерное зрение, система поддержки принятия врачебных решений, спнвр, сппр, офтальмология, диабетическая ретинопатия

## I. ВВЕДЕНИЕ

Использование систем на основе искусственного интеллекта (ИИ) выводит диагностику на основе анализа медицинских изображений на совершенно новый уровень. Однако, в таких областях как офтальмология технологии ИИ не получили широкого распространения.

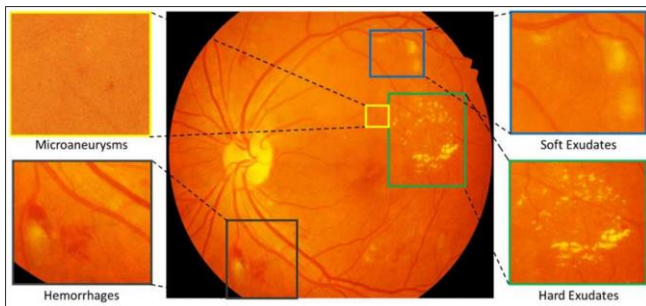


Рис. 1. Повреждения глазного дна при диабетической ретинопатии

Создание гибридных интеллектуальных систем для диагностики диабетической ретинопатии с

использованием объяснительного ИИ (ОИИ) имеет высокую актуальностью, что показано в работах [1, 2].

## II. МАТЕРИАЛЫ И МЕТОДЫ

### A. Предлагаемый подход

Предлагаемый подход универсален и базируется на идее того, что имеются формализованные критерии постановки диагноза на основе анализа медицинского изображения. Другим важным аспектом является возможность извлечь признаки с помощью ИНС (классификация, сегментация, детекция) и обработать их с помощью нечёткого классификатора. В данном случае, работа построенного прототипа может быть представлена в виде следующих шагов:

- предобработка поступившего изображения с помощью алгоритмов компьютерного зрения;
- предварительная классификация для отсева снимков без признаков заболевания;
- семантическая сегментация признаков заболевания;
- вывод масок сегментации и визуального объяснения, значений метрик качества обнаруженных классов;
- использование нечёткого классификатора на основе нечётких правил для определения стадии заболевания.

### B. Наборы данных

Ограниченное использование ИНС в задачах анализа офтальмологических изображений, во многом, обусловлено низкой доступностью данных для обучения моделей. Так, согласно [3], на 2023 год всего представлено 120 наборов изображений, из которых 43 набора по тематике заболевания «диабетическая ретинопатия», но только около трети наборов находятся в открытом доступе.

В качестве набора данных для задачи классификации был выбран APTOS2019 [4], состоящий из 3662 изображений, разделённых на 5 классов (No DR, Mild DR, Moderate DR, Severe DR, Proliferative DR). Исходя из постановки задачи как бинарной классификации, набор данных был переработан (классы: No DR, DR). С помощью методов аугментации набор был расширен

до 4800 изображений, с равными по количеству примеров двумя классами. Примеры изображений из набора данных APTOS2019 представлены на рис. 2.



Рис. 2. Изображения из набора данных APTOS2019 [4]

В отличие от задачи классификации, количество наборов для задачи сегментации значительно ограничено. Для решения поставленной задачи были выбраны наборы данных FGADR [5] (1842 изобр.) IDRID [6] (81 изобр.) поскольку они содержат одинаковые классы сегментации. Объединив два набора, был получен набор данных, состоящий из 1923 изображений, разделённых на 5 классов (soft exudates, hard exudates, hemorrhage, microaneurysms, neovascularization). На рис. 3 представлены примеры масок классов наборов данных FGADR и IDRID.

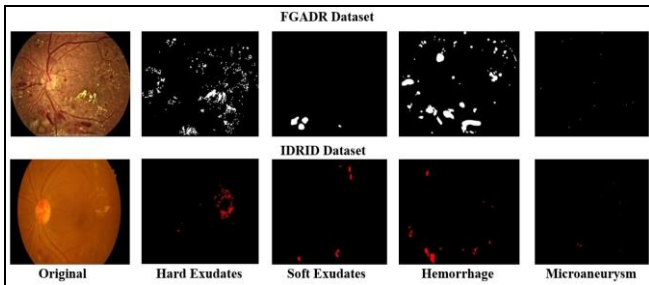


Рис. 3. Изображения из наборов данных FGADR [5] и IDRID [6]

### C. Архитектуры ИНС

EfficientNetV0 является базовой моделью из семейства EfficientNets, разработанного для оптимизации как точности, так и эффективности нейросетей. Данная модель будет использована для решения задачи бинарной классификации. Архитектура этих моделей основана на идее составного масштабирования, которое одновременно масштабирует глубину (количество слоев), ширину (количество каналов) и разрешение входных изображений, обеспечивая улучшенную производительность и эффективность. Эта идея была предложена в исследовательской работе Mingxing Tan и Quoc V. Le в 2019 году [7]. Используемая модель состоит из 132 слоёв и имеет 5,3 миллиона параметров. Архитектура сети представлена на рис. 4.

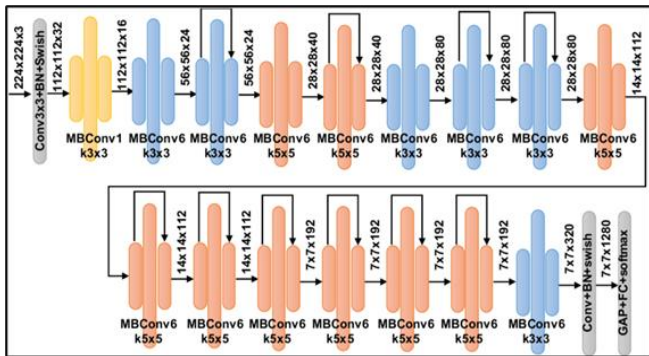


Рис. 4. Архитектура EfficientNetV0 [7]

ResUNet++ была представлена D. Jha в 2019 году и представляет собой усовершенствованную архитектуру ИНС U-Net для задач сегментации с улучшенной способностью к более точному выделению интересующих объектов на сложных изображениях благодаря внедрению механизмов внимания и многоуровневой обработки контекста [8]. Архитектура ResUNet++ представлена на рис. 5.

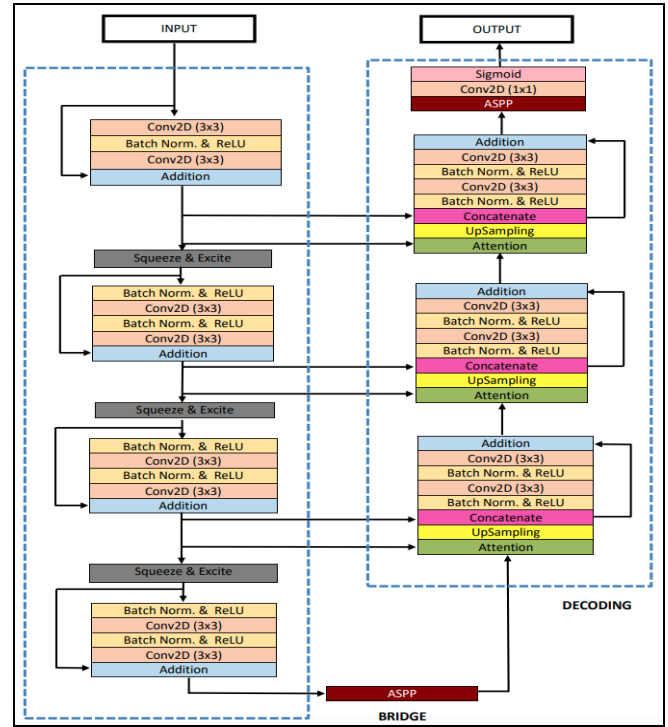


Рис. 5. Архитектура ResUNet++ [8]

### D. Нечёткий классификатор и формирование нечётких правил

Использование нечёткого классификатора невозможно без формирования базы (совокупности) нечётких правил, которая может быть сформирована на основе формализации имеющихся экспертных знаний [9]. От качества формирования совокупности правил напрямую зависит точность работы самого классификатора. Например, при использовании для задач диагностики заболеваний, основой для формирования базы правил могут выступать утверждённые на официальном уровне стандарты диагностики – клинические рекомендации.

В представленной работе, база нечётких правил была сформирована на основе критериев разметки набора данных Messidor [10], созданной врачами и состоящей из чётких правил. Авторами набора данных были введены следующие обозначения:  $\mu A$  – количество микроаневризм,  $N$  – количество ретинальных кровоизлияний,  $NV$  – наличие неоваскуляризации. Всего было получено 108 нечётких правил для трёх стадий заболевания и варианта нормы. Примеры соотношения чётких и нечётких правил для первой стадии диабетической ретинопатии представлены в табл. I.

ТАБЛИЦА I. ПРИМЕР СФОРМИРОВАННЫХ ПРАВИЛ

Тип правила	Формализованный вид
Детерминированное	$(0 < \mu A \leq 5) AND (H = 0)$
Нечёткое	

Такой шаг был предпринят в связи с отсутствием возможностей полной классификации патологий, входящих в перечень диагностики для постановки диагноза «диабетическая ретинопатия» в соответствии с клиническими рекомендациями Минздрава РФ.

#### E. Метод ОИИ

Применительно к поставленной задаче был выбран метод ОИИ Score-CAM, являющийся методом локального объяснения результатов работы непрозрачной модели – ИНС. Score-Weighted Visual Explanations for Convolutional Neural Networks (Score-CAM) был предложен Haofan Wang в 2020 году как обобщённая и усовершенствованная версия появившегося ранее метода CAM (Class Activation Mapping) [11] с модификациями, благодаря которым не требуется внесение изменений в архитектуру сети, что достигается избавлением от линейности в финальной части интерпретируемой модели. В отличие от предыдущих методов, основанных на картировании активации классов, Score-CAM избавляется от зависимости от градиентов, получая вес каждой карты активации через её прямое прохождение по целевому классу, окончательный результат получается путём линейной комбинации весов и карт активации [12].

#### F. Метрики оценки качества

Метрики оценки качества работы ИНС были выбраны исходя из решаемых в работе задач, а именно, бинарной классификации и семантической сегментации. Другим важным критерием при выборе метрик стал соотношение классов в наборах данных.

Для оценки работы ИНС в задаче бинарной классификации, учитывая сбалансированность набора данных, были выбраны метрики accuracy и specificity.

Accuracy – это доля объектов, для которых классификатор правильно определил класс. Вычисляется на основе значений матрицы ошибок: положительный-предсказано (TP), отрицательный-предсказано (FP), отрицательный-факт (FN), отрицательный-предсказано (FP). Значения данной метрики хорошо интерпретируемы поскольку характеризуют качество модели, агрегированное по всем классам.

$$Accuracy(ACC) = \frac{TP + TN}{TP + TN + FP + FN}$$

Specificity – это доля истинно отрицательных результатов. Специфичность отражает то, как часто классификатор точно спрогнозировала ложный результат, в том случае, когда результат является действительно ложным.

$$Specificity(SP) = \frac{TN}{TN + FP}$$

Для оценки качества работы ИНС в задачах сегментации используются метрики, оценивающие совпадение истинного и предсказанного на изображении региона пикселей. Оценка качества сегментации производилась на основе метрик: m\_IoU и Dice.

Intersection over Union (IoU) и Dice Similarity Coefficient (DSC) – это метрики, основанные на сравнении отношения истинных пикселей в рамках маски объекта к спрогнозированным пикселям маски сегментации. Различие этих метрик состоит в том, что IoU наказывает за недостаточную и чрезмерную сегментацию больше, чем DSC поэтому они часто используются в паре.

$$IoU = \frac{TP}{TP + FN + FP}$$

$$m\_IoU = \frac{1}{C} \sum_c IoU_c$$

$$DSC = \frac{2 \times TP}{(TP + FP) \times (TP + FN)}$$

### III. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Обучение, валидация и тестирование используемых моделей ИНС проводилось, исходя из задачи, на развитых на обучающую, валидационную, тестовую выборки в соотношении 70 %, 20 %, 10 % наборах данных APTOS2019 и FGARD+IDRID. Для улучшения результатов на предобученных версиях сетей была применена техника трансферного обучения. Дополнительно был произведен подбор гиперпараметров. Полученные результаты сравнивались с результатами базовых моделей, для которых процессы были такими же. Численные результаты представлены для тестовых выборок.

Результаты, представленные в табл. II свидетельствуют о превосходстве EfficientNetB0 над базовой моделью по всем показателям: accuracy – 0.97, specificity – 0.94. Во-многом, это объясняется преимуществами основной идеи архитектуры EfficientNet, которая заключается в подходе к масштабированию модели, когда не просто увеличивается глубина, но также возрастает количество каналов на каждом слое и разрешение изображений в сбалансированном соотношении, основываясь на фиксированных множителях, что в совокупности, даже при меньшем количестве параметров сети, позволяет достичь лучшего результата, а также получить значительный выигрыш в производительности.

ТАБЛИЦА II. РЕЗУЛЬТАТЫ БИНАРНОЙ КЛАССИФИКАЦИИ

Модель ИНС	ACC (%)	SP (%)
Base (VGG-16)	0.93	0.91
EfficientNetB0	<b>0.97</b>	<b>0.94</b>

Результаты, представленные в табл. III отражают превосходство модификации ResUNet++, по сравнению с изначальной архитектурой U-Net. Однако, достигнутые результаты m\_IoU – 0.86 и DSC – 0.88, показывают необходимость дальнейшего подбора

гиперпараметров или применения техник улучшения работы ИНС. На рис. 6 представлены результаты сегментации патологий моделью ResUNet++.

ТАБЛИЦА III. РЕЗУЛЬТАТЫ СЕМАТИЧЕСКОЙ СЕГМЕНТАЦИИ

Модель ИНС	m_IoU (%)	DSC (%)
Base (U-Net)	0.67	0.71
ResUNet++	<b>0.86</b>	<b>0.88</b>

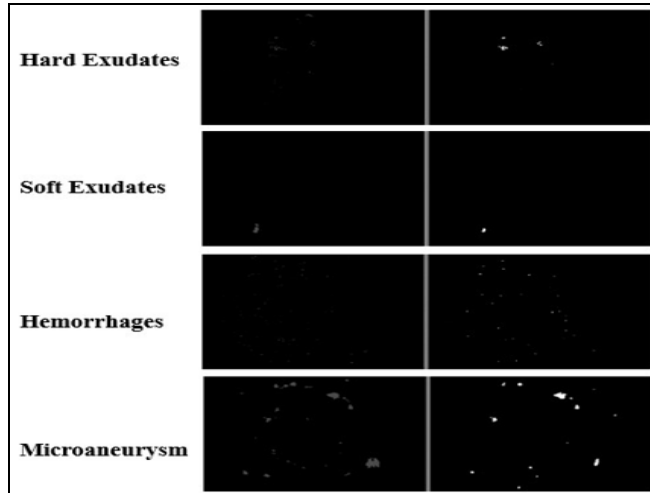


Рис. 6. Результаты сегментации с помощью ResUNet++

На рис. 7 представлено визуальное объяснение классов повреждений глазного дна (справа – микроаневризмы, слева – твёрдые экссудаты), полученное с помощью метода Score-CAM. По интенсивности цвета тепловой карты можно увидеть области с наибольшим количеством релевантных признаков, которые оказали ключевое влияние на процесс принятия решения моделью ИНС.

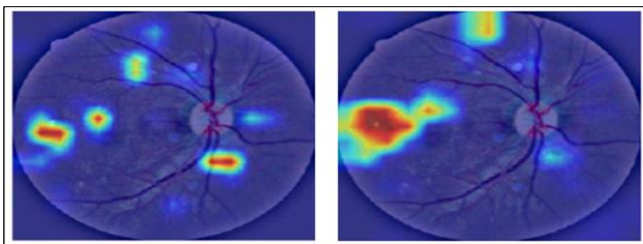


Рис. 7. Визуальное объяснение с помощью Score-CAM

#### IV. ЗАКЛЮЧЕНИЕ

Таким образом, в данном исследовании обосновано и подробно описано создание прототипа гибридного интерпретируемого программного решения для диагностики диабетической ретинопатии по фундус-снимкам. Проведён анализ источников литературы, показавший преимущества гибридных систем перед системами, использующими лишь одну технологию. Совместное использование технологий компьютерного зрения, ИНС и объяснительного ИИ, нечёткой логики позволяет создать комплексное решение путём разбивки основной задачи на несколько подзадач с применением разных технологических стеков для их решения. Предложенный подход к созданию гибридных систем является универсальным в области медицинской

диагностики, основанной на анализе изображений специализированных доменов. Реализованное на языке Python программное решение показало высокие результаты работы составляющих его части ИНС: для бинарной классификации (EfficientNetB0) accuracy – 0.97, specificity – 0.94; для семантической сегментации (ResUNet++) mIoU – 0.86, DSC – 0.88. В дальнейшем запланирована реализация системы с учётом формирования нечётких правил на основе российских критериев диагностики, содержащихся в клинических рекомендациях Минздрава РФ, что потребует разработки дополнительных механизмов предобработки изображений, а также создания нового набора данных. Кастомизация используемых архитектур ИНС также может увеличить точность работы на реальных данных. Однако, стоит отметить, что только проведение испытаний в условиях реальной клинической практики способны показать действительно слабые и сильные стороны предложенного решения.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Averkin A.N., Volkov E.N., Yarushev S.A. Explainable Artificial Intelligence in Deep Learning Neural Nets-Based Digital Images Analysis // Journal of Computer and Systems Sciences International. 2024. V. 65. № 1. P. 1-29. DOI: 10.1134/S1064230724700138.
- [2] Volkov E.N., Averkin A.N. Possibilities of explainable artificial intelligence for glaucoma detection using the lime method as an example // 2023 XXVI International Conference on Soft Computing and Measurements (SCM). IEEE: Saint-Petersburg, 2023. P. 130-133. DOI: 10.1109/SCM58628.2023.10159038.
- [3] Krzywicki T., Brona P., Zbrzezny A. B. et al. A global review of publicly available datasets containing fundus images: characteristics, barriers to access, usability, and generalizability // Journal of Clinical Medicine. 2023. V. 12. № 10. P. 3587. DOI: 10.3390/jcm12103587.
- [4] APTOS 2019 Blindness Detection // Kaggle URL: <https://www.kaggle.com/competitions/aptos2019-blindness-detection> (дата обращения: 09.03.2024).
- [5] Zhou Y., Wang B., Huang L. et al. A benchmark for studying diabetic retinopathy: segmentation, grading, and transferability //IEEE Transactions on Medical Imaging. 2020. Vol. 40. №. 3. P. 818-828. DOI: 10.1109/TMI.2020.3037771.
- [6] Porwal P., Pachade S., Kamble R. et al. Indian diabetic retinopathy image dataset (IDRID): a database for diabetic retinopathy screening research // Data. 2018. Vol. 3. №. 3. P. 25. DOI: 10.21227/H25W98.
- [7] Tan M., Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks //International conference on machine learning. PMLR, 2019. P. 6105-6114.
- [8] Jha D., Smedsrud P.H., Riegler M.A. et al. Resunet++: An advanced architecture for medical image segmentation // 2019 IEEE international symposium on multimedia (ISM). IEEE, 2019. P. 225-2255. DOI: 10.1109/ISM46123.2019.00049.
- [9] Рыжов А.П. Модели поиска информации в нечеткой среде. Москва, Издательство Центра прикладных исследований при механико-математическом факультете МГУ, 2004, 96 с.
- [10] Decencière E., Zhang X., Cazuguel G. et al. Feedback on a publicly distributed image database: the Messidor database //Image Analysis and Stereology. 2014. Vol. 33. №. 3. P. 231-234. DOI: 10.5566/ias.1155.
- [11] Zhou B., Khosla A., Lapedriza A. et al. Learning deep features for discriminative localization // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. P. 2921-2929.
- [12] Wang H., Wang Z., Du M. et al. Score-CAM: Score-weighted visual explanations for convolutional neural networks //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020. P. 24-25.