

К.В. СОБЯНИН, С.П. КУЛИКОВА  
**UNET-BOOSTED CLASSIFIER – МУЛЬТИЗАДАЧНАЯ  
АРХИТЕКТУРА ДЛЯ МАЛЫХ ВЫБОРОК НА ПРИМЕРЕ  
КЛАССИФИКАЦИИ МРТ СНИМКОВ ГОЛОВНОГО МОЗГА**

*Собянин К.В., Куликова С.П. Unet-boosted classifier – мультизадачная архитектура для малых выборок на примере классификации МРТ снимков головного мозга.*

**Аннотация.** Проблема обучения глубоких нейронных сетей на малых выборках особенно актуальна для медицинских задач. В работе рассматривается влияние попиксельной разметки значимых объектов на изображении, в дополнении к истинной метке класса, на качество решения задачи классификации. Для достижения лучших результатов классификации на малых выборках предлагается мультизадачная архитектура Unet-boosted classifier (UBC), обучаемая одновременно для решения задач классификации и семантической сегментации. В качестве исследуемого набора данных используются МРТ-снимки пациентов с доброкачественной глиомой и глиобластомой, взятые из открытого набора данных BraTS 2019. В качестве входа рассматривается один горизонтальный срез МРТ-изображения, содержащий глиому (всего 380 кадров в обучающей выборке), в качестве выхода – вероятность глиобластомы. В качестве базового решения используется ResNet34, обученный без аугментаций с функцией потерь на основе взаимной энтропии (CrossEntropyLoss). В качестве альтернативного решения используется UBC-ResNet34 – тот же ResNet34 усиленный декодером, построенным по принципу U-Net, и предсказывающим положение глиомы. В качестве дополнительной функции потерь используется сложенный коэффициент Соренсена-Дайса (DiceLoss). Результаты на тестовой выборке: доля правильных ответов (acc@casu) для базовой модели составила 0.71, для альтернативной – 0.81, коэффициент Дайса (Dice score) при этом составил 0.77. Таким образом, глубокую модель можно качественно обучить даже на небольшом наборе данных, используя предложенную архитектуру и добавив в разметку информацию о пораженных тканях в виде семантической маски. Предлагаемый подход потенциально может быть полезен и в любых других задачах классификации изображений с ограниченным набором данных.

**Ключевые слова:** классификация изображений, глубокое обучение, малый набор данных, семантическая сегментация, мультизадачная архитектура, церебральная патология, диагностика опухоли.

**1. Введение.** Разработка методов персонализированной медицины является одним из приоритетных направлений исследований в рамках Стратегии научно-технологического развития Российской Федерации. Это направление предоставляет обширные возможности для применения современных методов анализа данных и машинного обучения. В частности, одним из наиболее эффективных и перспективных методов анализа медицинских сигналов и изображений являются глубокие нейронные сети [1, 2]. Так, например, в обзоре [3] рассмотрено 49 научных исследований, посвященных диагностированию рака молочной железы, 39 из которых использовали глубокие сверточные сети. Обзор [4] описывает результаты 16 исследований, использующих глубокие сверточные сети

для диагностирования болезни Альцгеймера по МРТ-снимкам. Доля правильных ответов в этих исследованиях варьировалась от 0.79 до 0.99 на различных наборах данных. Также предыдущие исследования показывают, что глубокие сверточные сети успешно справляются с задачами классификации и локализации заболеваний органов грудной клетки [5] или выявления рака кожи на уровне компетентности специалистов [6].

Однако для обеспечения высокого качества работы таких моделей требуются значительные по объему массивы обучающих данных. Так, авторы в [6] использовали в своей работе по определению рака кожи модель архитектуры Inception v3 [7], предобученную на 1.41 млн изображений из набора ImageNet [8], которую они дообучили на наборе данных, содержащем 129 450 клинических изображений, размеченных экспертами-дерматологами. Для решения задачи классификации заболеваний органов грудной клетки (легких и сердца) авторы в [5] построили свою модель, объединив 4 другие модели, предобученные на ImageNet: AlexNet [9], GoogLeNet [10], VGGNet-16 [11] и ResNet-50 [12]. В этой работе для обучения классификатора и переходных слоев, объединяющих признаки всех четырех моделей, использовали набор данных «ChestX-ray8», собранный в 1992-2015 гг и содержащий 108 948 рентгеновских снимков грудной клетки 32 717 уникальных пациентов.

В то же время сбор данных в рамках исследования отдельной клинической задачи, как правило, является весьма трудоемким и зачастую продолжительным и дорогостоящим процессом. Стоит также отметить, что клинические данные нередко имеют пропущенные данные из-за невозможности прохождения пациентом тех или иных видов обследования, а сами данные могут варьироваться в зависимости от используемого клинического оборудования и протоколов обследований. Ввиду этих обстоятельств полноценно обследованная выборка даже из ста пациентов уже считается хорошим результатом. В связи с этим ограничением возникает потребность рассмотрения подходов, позволяющих повысить качество обучения модели на малых выборках.

Современные модели машинного обучения, сравнимые по качеству работы с человеком [13] часто содержат миллионы параметров [11] и обучаются на миллионах изображений [8]. Этот факт отсылает нас к проблеме «проклятия размерности», сформулированной в 1961 году американским математиком Ричардом Бэллманом [14], которую можно интерпретировать как «количество данных необходимое для оценки произвольной функции с заданным уровнем точности растет экспоненциально по отношению к числу входных данных» [15]. Таким

образом, число входных данных, то есть сам по себе размер изображения накладывает ограничение на задачу классификации. Для преодоления этой проблемы есть два принципиальных подхода: 1) увеличение количества обучающих данных и 2) регуляризация – ограничение количества степеней свободы модели, то есть фактически понижение размерности пространства входных данных.

В своем недавнем обзоре [16] авторы проанализировали 26 исследований, опубликованных за период с 2015 по 2021 г. и посвященных проблемам обучения моделей на малых наборах данных. В результате анализа авторы выделили 5 семейств подходов, используемых для обучения моделей на малых выборках:

– Модификации архитектуры – включают в себя различные модификации архитектуры на всех масштабах, регулирующие обучение: от дропаут-техник (англ. dropout) [17] и специальных «узких» блоков (англ. bottleneck block) [18] до различных встроенных фильтров, нацеленных на поиск определенных сигналов или геометрических примитивов, например гармонических блоков, которые обладают свойством инвариантности к вращению [19]. Также классическими примерами таких подходов являются дистилляция и ускорение обучения [20, 21]. Так, например, в [22] авторы использовали инвариантные к аффинному масштабированию сверточные блоки, использование которых позволило удвоить долю правильных ответов по сравнению с альтернативными моделями, таким как классические сверточные, гармонические [19] и капсульные нейронные сети [23], на ограниченной выборке (по 5 изображений на класс) из набора данных по распознаванию дорожных знаков Traffic Sign [24]. Еще одним примером использования подходов такого рода служит работа [25], в которой они использовали тангенсальные ядра (англ. Neural Tangent Kernel, NTK), эквивалентные линейным слоям неограниченной ширины, что позволило на 3% превзойти ResNet-34 [12] по результатам обучения на 640 изображениях из набора CIFAR-10 [26].

– Модификации функции потерь – обозначают подходы, ограничивающие обучение модели на уровне функции потерь. Подобным образом эффективно может быть использована косинусная мера близости [27], или в сочетании с обычной функцией потерь может использоваться метафункция потерь, обученная под конкретную задачу [28]. В частности, в [22] авторы использовали регуляризатор, инвариантный к вращению, который добавляется к функции потерь в качестве штрафного слагаемого. Еще одним примером подходов такого рода является работа [29], в которой авторы предложили геометрическую функцию потерь OLE loss (ang. Orthogonal Low-rank Embedding). Такая

функция потерь поощряет геометрическую разделимость между классами на уровне признаков, что позволило в сочетании с обычной функцией потерь достичь кратного прироста в качестве на выборках, содержащих менее 100 изображений из наборов MNIST [30] CIFAR-10, CIFAR-100 [26] и других.

– Аугментации данных – к этому семейству относятся все методы, увеличивающие размер набора обучающих данных [31], начиная от стандартных (кадрирование, смещение, поворот или масштабирование изображения и т.д.), до сложных методов генерации данных, например, с помощью генеративно-состязательных сетей (англ. Generative adversarial network, GAN) на основе исходного набора и бинарных масок [32]. Один из подходов такого рода представлен в работе [33], в которой была предложена соревновательная техника аугментации малых наборов данных DADA (англ. deep adversarial data augmentation) с добавлением объектов «фейковых» классов, находящихся в «спорных» точках признакового пространства с точки зрения меры близости объектов. Предложенный подход показал свои преимущества в том числе в задачах классификации электроэнцефалографических (ЭЭГ) сигналов и в определении рака груди на маммографических снимках.

– Аугментация признаков – это подходы, при которых к признакам добавляется шум с целью улучшения устойчивости обучения модели [34]. Одним из примеров такого подхода является предложенная в [34] состязательная аугментация признаков случайно выбранных скрытых слоев, что позволяет стабильно получать прирост в качестве 2-5% на небольших подвыборках MNIST [30] и CIFAR-10 [26].

– «Теплый старт» – это стратегия инициализации весов, при которой отдельные части модели предварительно обучаются на других данных, которые могут улучшить сходимость модели. Это могут быть другие реальные данные [35] или данные сгенерированные, например, с добавлением «фейковых» классов [33]. Так, например, в [36] авторы использовали обучение без учителя, чтобы обучить кодировщик MoCo v2 [37] для последующей дистилляции под задачу классификации.

Стоит отметить, что для малых выборок все ещё популярны классические модели, так как они зачастую содержат меньше параметров, и их довольно просто регуляризовать [38]. Так, например, Кристодулу и соавторы [39] отмечают в своем мета-анализе, что во многих клинических предсказательных задачах результаты работы моделей на основе методов машинного обучения зачастую сопоставимы с результатами моделей на основе обычной логистической регрессии. Тем не менее, авторы отмечают перспективность разработки подходов на основе машинного обучения.

Кроме выше перечисленных подходов, отдельно можно выделить семейство многозадачных архитектур, решающих одновременно несколько задач машинного обучения. Примером применения такого подхода является работа Мойаб и соавторов [40], в которой модель одновременно обучали решать задачу классификации и сегментации. Применение этого подхода позволило достичь прироста доли правильных ответов до 30% по сравнению с базовым решением в виде ResNet101 [12] на малых выборках из открытых баз данных MNIST [30], CIFAR-10 и CIFAR-100 [26], а также продемонстрировало свою работоспособность на клинической выборке, содержащей 45 изображений глазного дна, каждое из которых принадлежало одному из трех классов (без патологии, с признаками диабетической ретинопатии, наличие глаукомы).

Также хорошие результаты обучения на малых выборках для решения задачи семантической сегментации демонстрирует и архитектура U-Net [41], предложенная в 2015 году Роннебергом и соавт. В частности, авторам архитектуры удалось достичь качества сегментации клеток глиобластомы-астроцитомы U373 с коэффициентом Жаккара в 0.92 (англ. Intersection over Union, IoU), обучив модель на наборе данных PhC-U373, содержащем 32 частично размеченных снимка. Такие результаты обеспечиваются использованием меньшего количества параметров модели и большего объема полезной информации, извлекаемой из каждого отдельного пикселя разметки, по сравнению с архитектурами, содержащими полносвязные слои и обучающимися предсказывать только метку класса. Архитектура U-Net состоит из двух основных блоков: кодировщика (англ. encoder), который сжимает исходное изображение до признаковой карты (англ. feature map), и декодера (англ. decoder), который восстанавливает разметку. Поскольку кодировщик извлекает признаки из изображения и, по сути, является главным элементом в любой архитектуре, то, естественным образом, возникает желание переиспользовать его в задаче классификации.

В данной работе мы представляем подход, использующий дополнительную информацию об обучающей выборке в виде попиксельной разметки пораженных тканей, для лучшего обучения глубокой сверточной сети. Ключевой идеей подхода является обучение кодировщика одновременно внутри двух моделей – классической сверточной сети, решающей задачу классификации, и полносверточной архитектуры U-net, решающей задачу семантической сегментации и обладающей свойством обучаться даже на небольших наборах данных.

Качество работы предлагаемого подхода было протестировано на задаче классификации одиночных срезов МРТ-снимков на предмет

выявления злокачественной опухоли головного мозга. В результате обучения модели на снимках 76 пациентов предложенный подход показал прирост доли правильных ответов на 14% по сравнению с базовым решением.

Применение предлагаемого подхода может быть расширено на другие задачи анализа медицинских изображений или в других предметных областях для задач с малыми выборками.

**2. Данные.** В качестве исследуемых данных рассматривается открытый набор данных MICCAI по сегментации опухолей головного мозга (BraTS 2019) [42 – 44].

BraTS 2019 содержит данные 76 пациентов с диагнозами глиома (англ. lower grade glioma, LGG) и 259 пациентов с диагнозом глиобластома (англ. high grade glioma, HGG). Для каждого пациента представлены снимки 4 видов МРТ последовательностей:

- T1-взвешенное изображение (пропорционально времени продольной релаксации тканей);
  - T1-взвешенное изображение с контрастированием гадолинием;
  - T2-взвешенное изображение (пропорционально времени поперечной релаксации тканей);
  - T2 изображение инверсионного восстановления с подавлением сигнала от жидкости (англ. Fluid Attenuated Inversion Recovery, FLAIR).
- Указанные данные были записаны с использованием различных МРТ сканеров и клинических протоколов в 19 клинических центрах, но в дальнейшем были совмещены с единым анатомическим шаблоном [45] и передискретизированы к разрешению  $1 \times 1 \times 1$  мм.

Также для каждого пациента имеются аннотации, указывающие на тип ткани, верифицированные 4 клиническими экспертами [43]:

- Нормальная ткань;
- Перитуморальная отечная/инвазированная ткань;
- Неконтрастирующее ядро;
- Некротическое ядро, наполненное жидкостью;
- Контрастирующее ядро.

В качестве исследуемых данных были взяты снимки 76 пациентов с LGG и случайным образом выбранных 76 пациентов с HGG. Полученная выборка была разделена на две равные части с сохранением пропорции классов: обучающую и тестовую. Для каждого пациента были извлечены по 5 горизонтальных срезов FLAIR последовательности с наибольшей площадью поражения (рисунок 1) и соответствующие им бинарные маски (рисунок 2), являющиеся объединением всех пораженных тканей на срезе. В общей сложности, размер обучающей и тестовой выборки составил по

190 одноканальных изображений каждого класса с разрешением  $224 \times 224$  пикселей.

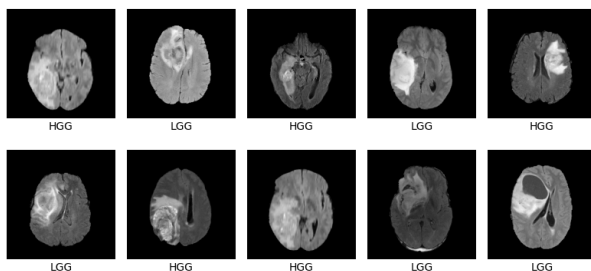


Рис. 1. Пример входных данных. На изображениях представлены МРТ-снимки FLAIR последовательности пациентов с диагнозами глиома (LGG) и глиобластома (HGG)

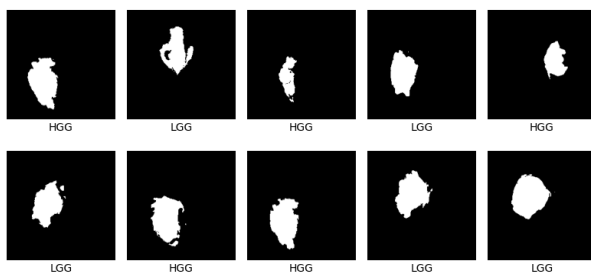


Рис. 2. Примеры бинарных масок новообразований головного мозга соответствующих МРТ-изображениям, представленным на рисунке 1

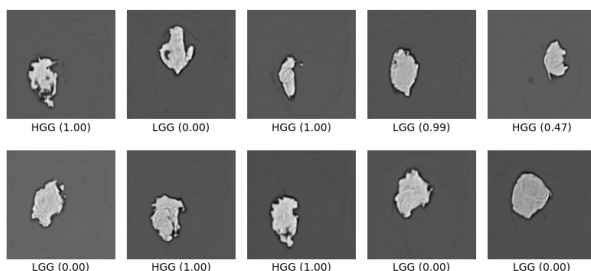


Рис. 3. Ответы UBC-resnet34 на изображения, представленные на рисунке 1. Представлена карта откликов до бинаризации, подпись указывает на истинный класс, число в скобках определяет уверенность модели в положительном классе – глиобластоме (HGG)

**3. Базовая модель.** Глубокие модели машинного обучения для решения задач классификации изображений обычно состоят из двух частей: кодировщика (англ. encoder), извлекающего признаки (англ. feature map) из изображения, и классификатора (англ. classification head), предсказывающего вероятность классов на основе признаков, извлеченных кодировщиком (рисунок 4). Благодаря иерархическому устройству сверточных сетей каждый последующий уровень модели извлекает все более сложные признаки, при этом на выходе из кодировщика объекты разных классов иногда даже становятся линейно разделимы в пространстве признакового описания. Поэтому в большинстве известных архитектур глубоких сверточных нейронных сетей, начиная с LeNet [46], классификатор представляет собой последовательность нескольких линейных полносвязных слоев с нелинейными функциями активации между ними.

Полносвязный слой суммирует все входные значения с некоторыми коэффициентами столько раз, сколько имеет выходов, и фактически решает задачу линейной регрессии для каждого выхода (1).

$$z_i(\mathbf{x}) = \mathbf{w}_i \cdot \mathbf{x} + b_i, \quad (1)$$

где  $z_i$  – значение  $i$ -го выхода полносвязного слоя,  $\mathbf{x}$  – входной вектор,  $\mathbf{w}_i$  – весовые коэффициенты  $i$ -го выхода,  $b_i$  – сдвиг  $i$ -го выхода (англ. bias).

Последовательность из нескольких полносвязных слоев очень быстро «раздувает» модель и во многих архитектурах большая часть параметров содержится именно в полносвязных слоях.

На малых выборках такие полносвязные сети имеют свойство очень быстро переобучаться – запоминать особенности выборки. Переобученная модель может хорошо описывать данные, на которых она обучалась, но её поведение непредсказуемо на новых данных, так как модель запомнила особенности именно тренировочного набора и не обладает обобщающей способностью. Поэтому, чтобы снизить эффект переобучения последних полносвязных слоев, в качестве базового решения была выбрана модель архитектуры ResNet34 [12] с однослойным линейным классификатором.

ResNet – это архитектура, решившая проблему плохого обучения глубоких моделей с помощью добавления специальных соединений (англ. residual connections), идущих в обход следующего блока и позволяющих градиенту во время обучения беспрепятственно распространяться через всю сеть.



ResNet34 состоит из 34 обучаемых слоев, 33 из которых содержатся в кодировщике и 1 в классификаторе. Кодировщик ResNet34 имеет входной блок, состоящий из сверточного слоя (Conv2d) с ядром  $7 \times 7$ , который расслаивает исходное изображение на 64 канала, не изменяя при этом его размер, нормализующего слоя (BatchNorm), функции активации (ReLU) и слоя понижающего дискретизацию признаков с сохранением наиболее сильных активаций (MaxPooling).

Далее устройство кодировщика имеет регулярную структуру, состоящую из 4-х уровней, каждый из которых работает со своим масштабом признаков и состоит из одинаковых базовых блоков (BasicBlock). Базовый блок представляет собой последовательность двух сверточных слоев (Conv2d) с ядром  $3 \times 3$  с последующей нормализацией (BatchNorm) и с активацией (ReLU) по середине. Вход каждого базового блока прибавляется к его выходу, усредняя тем самым исходное внутреннее представление и полученное в результате работы базового блока. Такой подход гарантирует прохождение градиента через всю сеть, даже если он занулится в каком-то конкретном слое. Переход к следующему масштабу обеспечивается дополнительным сверточным слоем в конце блока с ядром  $1 \times 1$  и с шагом 2 (англ. stride). В общей сложности, кодировщик ResNet34 содержит 21.3 млн обучаемых параметров и сжимает изображение разрешения  $224 \times 224$  до признаковой карты с разрешением  $512 \times 7 \times 7$  (таблица 1).

Таблица 1. Уровни кодировщика ResNet34

ResNet34 encoder levels and params					
name	B-blocks	convs	params	input shape	output shape
Input	0	1	3.26 K	$1 \times 224 \times 224$	$64 \times 112 \times 112$
level1	3	6	222 K	$64 \times 112 \times 112$	$64 \times 112 \times 112$
level2	4	8	1.12 M	$64 \times 112 \times 112$	$128 \times 56 \times 56$
level3	6	12	6.82 M	$128 \times 56 \times 56$	$256 \times 28 \times 28$
level4	3	6	13.1 M	$256 \times 28 \times 28$	$512 \times 7 \times 7$
total	16	33	21.3 M	$1 \times 224 \times 224$	$512 \times 7 \times 7$

Классификатор ResNet34 содержит 1026 параметров и состоит из AveragePooling слоя, который усредняет значения каналов, и одного линейного полносвязного слоя с одним выходом – вероятностью положительного класса.

**4. Unet-Boosted Classifier.** Предлагаемая в данной работе модель представляет собой гибридную мультизадачную архитектуру, одновременно решающую задачи классификации и семантической

сегментации (рисунок 5). Требование на уровне архитектуры кодирования и восстановления потенциально важных для классификации областей изображения позволяет обучить более глубокую сеть на тех же данных и повысить качество классификации. Поскольку такой подход нацелен на улучшение решения задачи классификации, мы назвали предложенную архитектуру Unet-Boosted Classifier (UBC), а модель, построенную по этому принципу на базе ResNet34 – UBC-ResNet34.

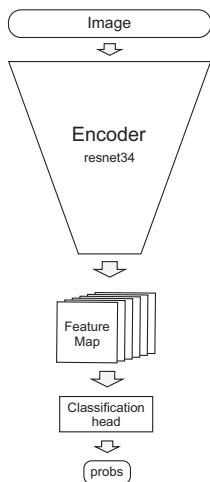


Рис. 4. Базовая модель:  
ResNet34

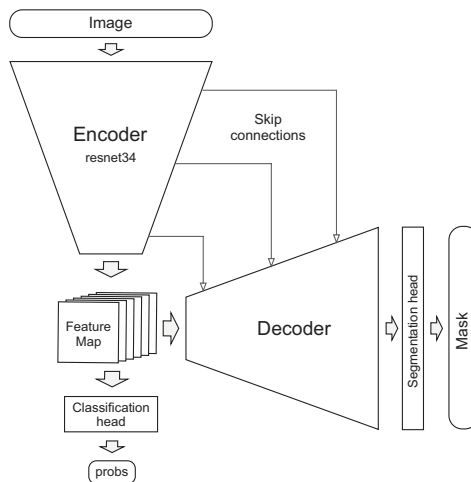


Рис. 5. Предлагаемая модель:  
UBC-ResNet34

UBC-ResNet34 состоит из тех же блоков, что и ResNet34: кодировщика и классификатора, но также содержит декодер (англ. decoder), восстанавливающий маску из признаков. Декодер UBC-ResNet34 построен по принципу U-Net и соединён с выходом кодировщика и его промежуточными результатами с помощью специальных связей (англ. skip connections), обеспечивающих лучшее распространение градиента и пропускание признаков от кодировщика к декодеру на разных масштабах. Декодер принимает на вход признаки всех масштабов и состоит из пяти уровней. Каждый уровень декодера смешивает признаки двух масштабов и отбирает наиболее значимые. Фактически представляет собой один повышающий дискретизацию слой (англ. up-convolution или ConvTranspose2d) для признаков меньшего масштаба с последующей конкатенацией с признаками следующего масштаба и последовательность двух сверточных слоев с ядром 3x3 (Conv2D). После каждого сверточного

слоя применяется нормализация (BatchNorm) для лучшей устойчивости обучения, а так же активация (ReLU). В общей сложности декодер содержит 3.15 млн. параметров и растягивает выход кодировщика в 32 раза до исходного размера изображения (таблица 2).

Таблица 2. U-Net декодер

U-Net decoder levels				
name	convs	params	input shape	output shape
level1	2	2.4 M	256×28×28 + 512×7×7	256×28×28
level2	2	590 K	128×56×56 + 256×28×28	128×56×56
level3	2	148 K	64×112×112 + 128×56×56	64×112×112
level4	2	46.2 K	64×112×112 + 64×112×112	32×112×112
level5	2	6.98 K	32×112×112	16×224×224
total	10	3.15 M	-	16×224×224

Далее результат работы декодера поступает на вход однослойной сегментирующей сети (англ. segmentation head), содержащей 145 параметров, которая восстанавливает маску изображения.

Таким образом, на входе модель также принимает одноканальное изображение, но имеет два выхода: выход классификатора – вероятность положительного класса, и выход сегментатора – маску, указывающую на поврежденные ткани.

**5. Обучение.** Для реализация моделей и их обучения использовались библиотеки pytorch и pytorch lightning. Реализация ResNet34 была взята из модуля torchvision.models, декодера Unet – из библиотеки segmentation models pytorch. Обучение моделей выполнялось методом обратного распространения ошибки (англ. backpropagation), с помощью стохастического градиентного спуска (англ. mini-batch gradient descent). Размер выборки итерации (англ. batch size) для обучения обеих моделей был одинаковый и составлял 50 обучающих примеров. Оптимизатор использовался адаптивный с инерцией Adam [47] с начальным шагом  $2 \cdot 10^5$  (англ. learning rate) и множителем угасания 0.9 (англ. learning rate annealing), применяющимся каждые 5 шагов.

Для оценки вклада предложенной архитектуры по сравнению с базовым решением инициализация параметров моделей происходила случайно, аугментации не использовались, а для вычисления ошибки классификации использовалась стандартная функция потерь на базе взаимной энтропии (CrossEntropyLoss) (2).

$$CE(\hat{p}, p) = \frac{1}{N} \sum_i^N \sum_{c=1}^C p_{ic} \log \hat{p}_{ic}, \quad (2)$$

где  $p$  – вектор разметки для выборки,  $\hat{p}$  – вектор вероятностей, полученный от классификатора,  $N$  – размер выборки,  $C$  – количество классов, в нашем случае 2.

Для вычисления ошибки сегментации использовалась функция потерь на основе сглаженного коэффициента Соренсена-Дайса (DiceLoss) (3).

$$SD(\hat{m}, m) = 1 - 2 \frac{\sum (\hat{m} \circ m) + \epsilon}{\sum (\hat{m} + m) + \epsilon}, \quad (3)$$

где  $m$  – бинарная маска, фактически матрица со значениями 0 и 1;  $\hat{m}$  – предсказанная моделью маска;  $\circ$  – операция поэлементного умножения, фактически обнуляет все ложные срабатывания, оставляя только пересечение  $m$  и  $\hat{m}$ , после чего происходит суммирование одновременно по всем осям;  $\epsilon$  – сглаживающий параметр.

Для обучения мультизадачной модели ошибка определялась как сумма ошибок классификации и сегментации (4).

$$L(\hat{p}, \hat{m}, p, m) = \alpha \cdot CE(\hat{p}, p) + (1 - \alpha) \cdot SD(\hat{m}, m), \quad (4)$$

где  $\alpha$  – балансовый гиперпараметр, который в исследовании был принят равным 0.5.

Для более комплексного сравнения каждая из моделей была обучена на малых выборках разного объема: от 1 до 190 изображений на класс. Выборки размером от 1 до 38 изображений на класс представляют собой подвыборки обучающего набора данных, содержащие по 1 случайному изображению на пациента. Выборки из 76 и 190 изображений на класс содержали соответственно по 2 и 5 изображений от каждого пациента из обучающего набора данных.

Модели обучались на протяжении 200 эпох, значения метрик для сравнения выбирались при достижении плато и усреднялись по 5 значениям, вычисленным на тестовой выборке. Тестирование моделей, обученных на подвыборках разных объемов, выполнялось на одном и том же тестовом наборе данных, содержащем 380 изображений от 76 пациентов.

**6. Результаты и дискуссия.** Большая часть моделей, обучавшихся на выборках содержащих менее 50 изображений, достигла локального минимума функции потерь менее, чем за 100 шагов обучения, то есть, при batch size равным 50, менее чем за 100 эпох. ResNet34, обучавшийся на

полном наборе данных (380 изображений) – за 20 эпох, UBC-ResNet34 – за 60 эпох. Результаты обучения на выборках разной величины представлены на рисунке 6. Взаимосвязь качества классификации и сегментации моделей UBC-ResNet34 представлена на рисунке 7.

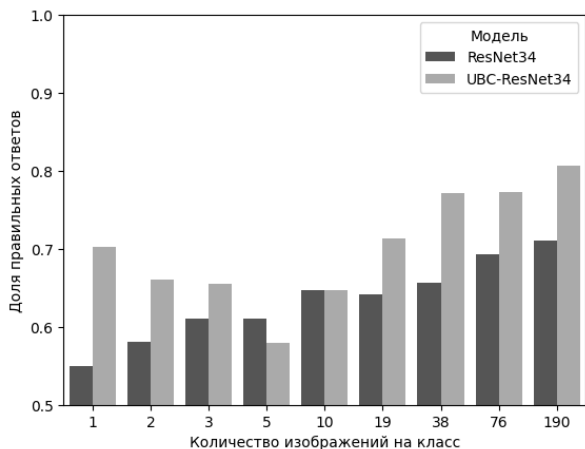


Рис. 6. Результаты обучения моделей на малых выборках

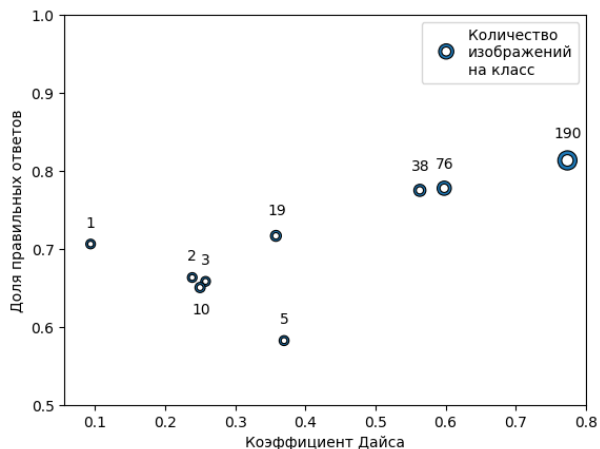


Рис. 7. Взаимосвязь доли правильных ответов с коэффициентом Дайса

Высокое значение доли правильных ответов модели UBC-ResNet34 обученной на выборке из двух изображений можно интерпретировать как выброс, по скольку согласно точечной диаграмме на рисунке 7,

наблюдается положительная зависимость между качеством сегментации и классификации. Мы полагаем, что два данных изображения, попавших в обучающую выборку, наиболее характерны для всего представленного набора данных, поскольку добавление последующих восьми изображений снижает качество классификации на каждом этапе. Так, интерпретируя глубокую просадку качества модели UBC-ResNet34 и снижение темпа роста качества базовой модели – ResNet34, можно прийти к выводу что выборка, состоящая из 10 изображений (по 5 на класс), содержит нехарактерные изображения для своего класса.

Результаты обучения на полных данных представлены в таблице 3. На тестовых данных из 380 изображений, ResNet34 достиг доли правильных ответов ( $accuracy = (TP + TN)/n$ ) в  $0.71 \pm 0.05$ , что значительно отличается от тривиального решения (англ. dummy classifier)  $\hat{y}(x) = 1$  с результатом в  $0.50 \pm 0.05$ . В случае использования предложенной модели UBC-ResNet34 доля правильных ответов составила  $0.81 \pm 0.04$ , коэффициент Дайса (Dice score) при этом составил 0.77. Интервальные оценки получены методом Уилсона для 95% уровня доверия. Примеры ответов модели представлены на рисунке 3.

Таблица 3. Результаты обучения ResNet34 и UBC-ResNet34 на 380 изображениях

Результаты на тестовой выборке из 380 изображений				
Модель	Параметры	Метки	Доля правильных ответов, 95% ДИ	Коэффициент Дайса
dummy	-	-	$0.50 \pm 0.05$	-
ResNet34	21.3 М	190	$0.71 \pm 0.05$	-
UBC-ResNet34	24.4 М	0.90 М	$0.81 \pm 0.04$	0.77

Полученные результаты можно объяснить одновременным влиянием как минимум двух факторов. Во-первых, количество обучающей информации увеличилось соизмеримо с количеством пикселей в изображении, так как в среднем по обучающей выборке, площадь поражения составляет 5% от площади снимка. Таким образом, количество обучаемых параметров выросло всего на 15% (столбец «Параметры» в таблице 3), а количество положительных меток в обучающей выборке выросло более чем на три порядка (столбец «Метки» в таблице 3), что в свою очередь повлияло на качество обучения кодировщика. Во-вторых, есть основания полагать, что требование описывать положение пораженных тканей на уровне признаков позволило кодировщику научиться извлекать важные признаки для детектирования злокачественной опухоли.

Из второго фактора следует важное потенциальное преимущество предложенной архитектуры – возможность продолжать обучать кодировщик на данных пациентов с неустановленным диагнозом. Такие обучающие примеры можно использовать в одном цикле с полностью размеченными данными и обновлять веса кодировщика, используя градиент, полученный только от декодера.

Так же второй фактор представляется наиболее интересным с точки зрения стратегии разметки изображений. Развивая мысль о взаимосвязи качества классификации с качеством семантической маски, мы полагаем, что потенциальный прирост может дать обучение с многоклассовой семантической маской, показывающей, например, принадлежность объекта на изображении к определенной анатомической области. Наличие данной информации может также потенциально улучшить возможность клинической интерпретации результатов работы модели.

В то же время добавление в аннотацию информации об анатомической принадлежности того или иного участка изображения само по себе является непростой задачей. Золотым стандартом для анатомического аннотирования изображений по-прежнему является ручная разметка, выполняемая клиническим специалистом (радиологом, неврологом, патологоанатомом), которая требует значительных временных затрат, и качество которой зависит от опыта специалиста.

Наиболее распространенные автоматические подходы для анатомической сегментации МРТ-изображений головного мозга, реализованные в таких программных продуктах как FreeSurfer, BrainVisa, volBrain, FSL, SPM, BrainSuite, Cat12 и т.д., основываются на использовании существующих, вручную созданных, анатомических атласов головного мозга и позволяют определить принадлежность пикселей изображений к одной из предопределенных в атласе области интереса. Однако качество работы таких алгоритмов сильно зависит от качества совмещения МРТ-снимка с шаблоном атласа, что в свою очередь, также зависит от качества самого МРТ-снимка. При этом стоит отметить значительную индивидуальную вариабельность анатомического строения головного мозга, что также усложняет задачу аннотирования.

Кроме этого, неизбежно возникают сложности, связанные с проблемой соотнесения области пораженной ткани с соответствующей её анатомической областью атласа, так как атласы создаются на основе МРТ-изображений здоровых взрослых людей, а области поражения отличаются изменениями интенсивности изображения в одной или нескольких модальностях. В результате этих обстоятельств, ни один

из выше указанных подходов не обладают анатомической точностью сопоставимой с результатами аннотирования «вручную». Существует надежда, что современные методы машинного обучения смогут справиться и с этой задачей [51 – 55]. Однако здесь возникает некий «порочный круг», так как обучение этих методов само по себе требует больших объемов анатомически аннотированных данных, а использование в качестве обучающих данных результатов наивных автоматизированных систем сегментации, часто содержащих неточные нейроанатомические описания, очень вероятно приведут к переносу этих нейроанатомических неточностей и в модели машинного обучения.

Поскольку подобные аннотации создаются вручную, стоит так же упомянуть проблему меж-экспертной вариабельности при разметке медицинских изображения (англ. *inter-expert variations*), в том числе МРТ-снимков, что потенциально может негативно сказаться на качестве результатов классификации. Однако эта проблема может быть частично решена с использованием полуавтоматических подходов, например, таких, когда эксперт сначала выделяет область возможного новообразования, а затем алгоритм сравнивает области здорового и пораженного мозга и уточняет границу опухоли. Такой подход позволяет достичь среднего коэффициента Дайса в 90%, что соизмеримо с экспертной оценкой [48]. Кроме того, существуют подходы, позволяющие уменьшить меж-экспертные различия уже существующей разметки с помощью вариационных автоэнкодеров [49].

Также не исключено, что качество разметки пораженных тканей может не сильно влиять на качество итоговой классификации, так как классификатор применяет к признаковой карте поканальное усреднение (*Average Pooling*), т.е. фактически никак не учитывает локализацию поражения. Возможно, для качества модели гораздо важнее зафиксировать сам факт наличия признаков, описывающих пораженные структуры, и для достижения максимального эффекта достаточно предсказывать только ограничивающую рамку (англ. *bounding box*).

Развивая идею важности наличия признаков, можно предположить, что часть полезных признаков при описанном в работе способе обучения может скрываться на других масштабах кодировщика. Поэтому дополнительный прирост качества классификации может быть обеспечен требованием восстановления семантической маски на всех масштабах декодера по аналогии с подходом «пирамида признаков» (англ. *feature pyramid*) [50].

Ещё одним важным направлением будущих исследований предложенной архитектуры является изучение вопроса устойчивости её



работы к наличию шумов и артефактов на изображениях. В настоящей работе, как и в большинстве существующих, посвященных вопросам сегментации МРТ изображений, и обучение модели, и её тестирование проводилось на относительно качественных, специально отобранных и предобработанных, данных с хорошим пространственным разрешением из базы данных BraTS. Однако, в реальной клинической практике качество изображений сильно зависит от многих факторов таких, как используемый сканер (в том числе качество магнитной катушки, её сила), параметры протокола сканирования (в том числе пространственное разрешение), интенсивность движений пациента в ходе сканирования (вызывают разнообразные артефакты движения) и применяемых методов для пост-коррекции артефактов [56]. Поэтому в дальнейших исследованиях для проверки возможностей практического применения предлагаемой архитектуры в реальных задачах клинической практики, необходимо провести тестирование устойчивости работы разработанного подхода в зависимости от всех выше перечисленных параметров.

**7. Заключение.** В работе рассмотрено сравнение моделей ResNet34 и UBC-ResNet34 – мультизадачной архитектуры Unet-Boosted Classifier (UBC) на базе ResNet34 на малой выборке клинических МРТ-изображений головного мозга. Использование предложенной архитектуры позволяет достичь существенно лучших результатов на тех же данных с использованием расширенных аннотаций.

Как видно из результатов исследования, предложенный подход позволяет эффективно обучить даже такую объемную модель, как ResNet34, содержащую 21.3 млн параметров, всего лишь на 380 снимках, без использования каких либо других приемов. В сочетании с правильно подобранными аугментациями, использованием предобученного компактного кодировщика и другими популярными приемами этот подход может стать серьезным инструментом в руках исследователей на пути к достижению лучших метрик на малых обучающих выборках и представляет собой перспективное направление для дальнейших исследований.

В частности, с практической точки зрения, описанный подход может быть применен во многих задачах классификации медицинских изображений, в том числе к задаче определения этиологии криптогенного ишемического инсульта [57], установление точной причины которого является трудоемкой и социально значимой задачей.

Однако для обучения модели архитектуры UBC требуется дополнительная аннотация изображений в виде семантической маски, отмечающей важные элементы изображения. Поскольку предлагаемая

архитектура нацелена на решение медицинских задач с ограниченными данными, разметка имеющихся изображений может оказаться существенно более доступным способом повышения качества модели, чем сбор большего количества клинических случаев.

## Литература

1. Li Q., Cai W., Wang X., Zhou Y., Feng D.D., Chen M. Medical image classification with convolutional neural network // 13th international conference on control automation robotics vision (ICARCV). IEEE. 2014. pp. 844–848.
2. Cai L., Gao J., Zhao D. A review of the application of deep learning in medical image classification and segmentation // *Annals of translational medicine*. 2020. vol. 8. no. 11.
3. Murtaza G., Shuib L., Abdul Wahab A.W., Mujtaba G., Mujtaba G., Nweke H.F., Al-garadi M.A., Zulfiqar F., Raza G., Azmi N.A. Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges // *Artificial Intelligence Review*. 2020. vol. 53. pp. 1655–1720.
4. Yamanakkanavar N., Choi J.Y., Lee B. MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer’s disease: a survey // *Sensors*. 2020. vol. 20. no. 11. pp. 3243.
5. Wang X., Peng Y., Lu L., Lu Z., Bagheri M., Summers R.M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases // *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. pp. 2097–2106.
6. Esteva A., Kuprel B., Novoa R.A., Ko J., Swetter S.M., Blau H.M., Thrun S. Dermatologist-level classification of skin cancer with deep neural networks // *nature*. 2017. vol. 542. no. 7639. pp. 115–118.
7. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the inception architecture for computer vision // *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. pp. 2818–2826.
8. Ridnik T., Ben-Baruch E., Noy A., Zelnik-Manor L. Imagenet-21k pretraining for the masses // *arXiv preprint arXiv:2104.10972*. 2021.
9. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks // *Advances in neural information processing systems*. 2012. vol. 25.
10. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions // *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. pp. 1–9.
11. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // *arXiv preprint arXiv:1409.1556*. 2014.
12. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. pp. 770–778.
13. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A., Fei-Fei L. Imagenet large scale visual recognition challenge // *International journal of computer vision*. 2015. vol. 115. pp. 211–252.
14. Hammer P. Adaptive control processes: a guided tour (R. Bellman). 1962. 163 p.
15. Chen L. Curse of Dimensionality // *Encyclopedia of Database Systems*. Boston, MA: Springer US, 2009. pp. 545–546.
16. Brigato L., Barz B., Iocchi L., Denzler J. Image classification with small datasets: Overview and benchmark // *IEEE Access*. 2022. vol. 10. pp. 49233–49250.

17. Hinton G.E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R.R. Improving neural networks by preventing co-adaptation of feature detectors // arXiv preprint arXiv:1207.0580. 2012.
18. Howard A., Sandler M., Chu G., Chen L.-C., Chen B., Tan M., Wang W., Zhu Y., Pang R., Vasudevan V., Le Q., Adam H. Searching for mobilenetv3 // Proceedings of the IEEE/CVF international conference on computer vision. 2019. pp. 1314–1324.
19. Kim J., Jung W., Kim H., Lee J. CyCNN: A rotation invariant CNN using polar mapping and cylindrical convolution layers // arXiv preprint arXiv:2007.10588. 2020.
20. Tan M., Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks // International conference on machine learning. PMLR. 2019. pp. 6105–6114.
21. Tan M., Le Q. Efficientnetv2: Smaller models and faster training // International conference on machine learning. PMLR. 2021. pp. 10096–10106.
22. Xu W., Wang G., Sullivan A., Zhang Z. Towards learning affine-invariant representations via data-efficient cnns // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020. pp. 904–913.
23. Sabour S., Frosst N., Hinton G.E. Dynamic routing between capsules // Advances in neural information processing systems. 2017. vol. 30.
24. Stallkamp J., Schlipsing M., Salmen J., Igel C. The German traffic sign recognition benchmark: a multi-class classification competition // 2011 international joint conference on neural networks. IEEE. 2011. pp. 1453–1460.
25. Arora S., Du S.S., Li Z., Salakhutdinov R., Wang R., Yu D. Harnessing the power of infinitely wide deep nets on small-data tasks // arXiv preprint arXiv:1910.01663. 2019.
26. Krizhevsky A., Hinton G. et al. Learning multiple layers of features from tiny images. 2009. 60 p.
27. Barz B., Denzler J. Deep learning on small datasets without pre-training using cosine loss // Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020. pp. 1371–1380.
28. Hai Z., Liu X. Evolving parametrized Loss for Image Classification Learning on Small Datasets // arXiv preprint arXiv:2103.08249. 2021.
29. Lezama J., Qiu Q., Mus e P., Sapiro G. Ole: Orthogonal low-rank embedding-a plug and play geometric loss for deep learning // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 8109–8118.
30. LeCun Y., Cortes C., Burges C. MNIST handwritten digit database. 2010. vol. 2. 18 p.
31. Shorten C., Khoshgoftaar T.M. A survey on image data augmentation for deep learning // Journal of big data. 2019. vol. 6. no. 1. pp. 1–48.
32. Reinders C., Schubert F., Rosenhahn B. Chimeramix: Image classification on small datasets via masked feature mixing // arXiv preprint arXiv:2202.11616. 2022.
33. Zhang X., Wang Z., Liu D., Lin Q., Ling Q. Deep adversarial data augmentation for extremely low data regimes // IEEE Transactions on Circuits and Systems for Video Technology. 2020. vol. 31. no. 1. pp. 15–28.
34. Ishii M., Sato A. Training deep neural networks with adversarially augmented features for small-scale training datasets // 2019 International Joint Conference on Neural Networks (IJCNN). IEEE. 2019. pp. 1–8.
35. Agarwal N., Sondhi A., Chopra K., Singh G. Transfer learning: Survey and classification // Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS 2020. 2021. pp. 145–155.
36. Zhao B., Wen X. Distilling visual priors from self-supervised learning // Computer Vision – ECCV 2020 Workshops. Springer. 2020. pp. 422–429.

37. Chen X., Fan H., Girshick R., He K. Improved baselines with momentum contrastive learning // arXiv preprint arXiv:2003.04297. 2020.
38. Miranda G., Rubio-Manzano C. Image Classification Using Deep and Classical Machine Learning on Small Datasets: A Complete Comparative. 2022. DOI: 10.20944/preprints202201.0367.v1.
39. Christodoulou E., Ma J., Collins G.S., Steyerberg E.W., Verbakel J.Y., Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models // Journal of clinical epidemiology. 2019. vol. 110. pp. 12–22.
40. Mojab N., Yu P.S., Hallak J.A., Yi D. Cvs: Classification via segmentation for small datasets // arXiv preprint arXiv:2111.00042. 2021.
41. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th international conference. Springer International Publishing, 2015. pp. 234–241.
42. Baid U., Ghodasara S., Mohan S., Bilello M., Calabrese E., Colak E., Farahani K., Kalpathy-Cramer J., Kitamura F.C., Pati S. et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification // arXiv preprint arXiv:2107.02314. 2021.
43. Menze B.H., Jakab A., Bauer S., Kalpathy-Cramer J., Farahani K., Kirby J., Burren Y., Porz N., Slotboom J., Wiest R., et al. The multimodal brain tumor image segmentation benchmark (BRATS) // IEEE transactions on medical imaging. 2014. vol. 34. no. 10. pp. 1993–2024.
44. Bakas S., Akbari H., Sotiras A., Bilello M., Rozycki M., Kirby J.S., Freymann J.B., Farahani K., Davatzikos C. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features // Scientific data. 2017. vol. 4. no. 1. pp. 1–13.
45. Rohlfing T., Zahr N.M., Sullivan E.V., Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure // Human brain mapping. 2010. vol. 31. no. 5. pp. 798–819.
46. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition // Proceedings of the IEEE. 1998. vol. 86. no. 11. pp. 2278–2324.
47. Kingma D.P., Ba J. Adam A method for stochastic optimization // arXiv preprint arXiv:1412.6980. 2014.
48. Akkus Z., Sedlar J., Coufalova L., Korfiatis P., Kline T.L., Warner J.D., Agrawal J., Erickson B.J. Semi-automated segmentation of pre-operative low grade gliomas in magnetic resonance imaging // Cancer Imaging. 2015. vol. 15. pp. 1–10.
49. Liu X., Xing F., Marin T., El Fakhri G., Woo J. Variational inference for quantifying inter-observer variability in segmentation of anatomical structures // Medical Imaging 2022: Image Processing. 2022. vol. 12032. pp. 438–443.
50. Lin T.-Y., Dollar P., Girshick R., He K., Hariharan B., Belongie S. Feature pyramid networks for object detection // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 2117–2125.
51. Huo Y., Xu Z., Xiong Y., Aboud K., Parvathaneni P., Bao S., Bermudez C., Resnick S., Cutting L., Landman B. 3D whole brain segmentation using spatially localized atlas network tiles // NeuroImage. 2019. vol. 194. pp. 105–119.
52. Paschali M., Gasperini S., Roy A.G., Fang M.Y.S., Navab N. 3DQ: Compact quantized neural networks for volumetric whole brain segmentation // Medical Image Computing and Computer Assisted Intervention: 22nd International Conference. Springer International Publishing, 2019. pp. 438–446.

53. Roy A.G., Conjeti S., Navab N., Wachinger C. Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation for structure-wise quality control // *NeuroImage*. 2019. vol. 195. pp. 11–22.
54. Henschel L, Conjeti S., Estrada S., Diers K., Fischl B., Reuter, M. Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline // *NeuroImage*. 2020. vol. 219. pp. 117012.
55. Coupe P, Mansencal, B., Clement, M., Giraud, R., de Senneville B.D., Ta V.T., Lepetit V., Manjon J.V. AssemblyNet: A large ensemble of CNNs for 3D whole brain MRI segmentation // *NeuroImage*. 2020. vol. 219(5). pp. 117026.
56. Dubois J., Alison M., Counsell S.J., Hertz-Pannier L., Huppi P.S., Benders M.J. MRI of the neonatal brain: a review of methodological challenges and neuroscientific advances // *Journal of Magnetic Resonance Imaging*. 2021. vol. 53. no. 5. pp. 1318–1343.
57. Ntaios G. Embolic stroke of undetermined source: JACC review topic of the week // *Journal of the American College of Cardiology*. 2020. vol. 75. no. 3. pp. 333–340.

**Собынин Кирилл Валентинович** — младший научный сотрудник, центр когнитивных нейронаук, Национальный исследовательский университет «Высшая школа экономики». Область научных интересов: машинное обучение, применение методов математики и информатики в неврологии и нейронауках. Число научных публикаций — 9. k.sobyenin@gmail.com; бульвар Гагарина, 37а, 614060, Пермь, Россия; р.т.: +7(342)200-9555.

**Куликова Софья Петровна** — Ph.D., доцент, старший научный сотрудник, центр когнитивных нейронаук, Национальный исследовательский университет «Высшая школа экономики». Область научных интересов: моделирование когнитивных процессов, анализ МРТ снимков, развитие головного мозга. Число научных публикаций — 19. spkulikova@hse.ru; бульвар Гагарина, 37а, 614060, Пермь, Россия; р.т.: +7(342)200-9555.

**Поддержка исследований.** Публикация подготовлена в ходе проведения исследования № 23-00-026 «Разработка автоматических подходов для определения этиологии криптогенного инсульта с целью профилактики вторичных острых нарушений мозгового кровообращения» в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)».

K. SOBYANIN, S. KULIKOVA  
**UNET-BOOSTED CLASSIFIER – MULTI-TASK ARCHITECTURE  
FOR SMALL DATASETS APPLIED TO BRAIN MRI  
CLASSIFICATION**

*Sobyanin K., Kulokova S. Unet-boosted Classifier – Multi-Task Architecture for Small Datasets Applied to Brain MRI Classification.*

**Abstract.** The problem of training deep neural networks on small samples is especially relevant for medical issues. The paper examines the impact of pixel-wise marking of significant objects in the image, over the true class label, on the quality of the classification. To achieve better classification results on small samples, we propose a multitasking architecture – Unet-boosted classifier (UBC), that is trained simultaneously to solve classification and semantic segmentation problems. As the exploratory dataset, MRI images of patients with benign glioma and glioblastoma taken from the BRaTS 2019 data set are used. One horizontal slice of the MRI image containing a glioma is considered as the input (a total of 380 frames in the training set), and the probability of glioblastoma – as the output. Resnet34 was chosen as the baseline, trained without augmentations with a loss function based on cross-entropy. As an alternative solution UBC-resnet34 is used – the same resnet34, boosted by a decoder built on the U-Net principle and predicting the pixels with glioma. The smoothed Sorensen-Dice coefficient (DiceLoss) is used as a decoder loss function. Results on the test sample: accuracy for the baseline reached 0.71, for the proposed model – 0.81, and the Dice score – 0.77. Thus, a deep model can be well trained even on a small data set, using the proposed architecture, provided that marking of the affected tissues in the form of a semantic mask is provided.

**Keywords:** image classification, deep learning, small dataset, semantic segmentation, multi-task architecture, cerebral pathology, tumor diagnosis.

## References

1. Li Q., Cai W., Wang X., Zhou Y., Feng D.D., Chen M. Medical image classification with convolutional neural network. 13th international conference on control automation robotics vision (ICARCV). IEEE. 2014. pp. 844–848.
2. Cai L., Gao J., Zhao D. A review of the application of deep learning in medical image classification and segmentation. *Annals of translational medicine*. 2020. vol. 8. no. 11.
3. Murtaza G., Shuib L., Abdul Wahab A.W., Mujtaba G., Mujtaba G., Nweke H.F., Al-garadi M.A., Zulfiqar F., Raza G., Azmi N.A. Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. *Artificial Intelligence Review*. 2020. vol. 53. pp. 1655–1720.
4. Yamanakkanavar N., Choi J.Y., Lee B. MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: a survey. *Sensors*. 2020. vol. 20. no. 11. pp. 3243.
5. Wang X., Peng Y., Lu L., Lu Z., Bagheri M., Summers R.M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. pp. 2097–2106.
6. Esteva A., Kuprel B., Novoa R.A., Ko J., Swetter S.M., Blau H.M., Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *nature*. 2017. vol. 542. no. 7639. pp. 115–118.

7. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 2818–2826.
8. Ridnik T., Ben-Baruch E., Noy A., Zelnik-Manor L. Imagenet-21k pretraining for the masses. arXiv preprint arXiv:2104.10972. 2021.
9. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 2012. vol. 25.
10. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. pp. 1–9.
11. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.
12. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 770–778.
13. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A., Fei-Fei L. Imagenet large scale visual recognition challenge. International journal of computer vision. 2015. vol. 115. pp. 211–252.
14. Hammer P. Adaptive control processes: a guided tour (R. Bellman). 1962. 163 p.
15. Chen L. Curse of Dimensionality. Encyclopedia of Database Systems. Boston, MA: Springer US, 2009. pp. 545–546.
16. Brigato L., Barz B., Iocchi L., Denzler J. Image classification with small datasets: Overview and benchmark. IEEE Access. 2022. vol. 10. pp. 49233–49250.
17. Hinton G.E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R.R. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580. 2012.
18. Howard A., Sandler M., Chu G., Chen L.-C., Chen B., Tan M., Wang W., Zhu Y., Pang R., Vasudevan V., Le Q., Adam H. Searching for mobilenetv3. Proceedings of the IEEE/CVF international conference on computer vision. 2019. pp. 1314–1324.
19. Kim J., Jung W., Kim H., Lee J. CyCNN: A rotation invariant CNN using polar mapping and cylindrical convolution layers. arXiv preprint arXiv:2007.10588. 2020.
20. Tan M., Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning. PMLR. 2019. pp. 6105–6114.
21. Tan M., Le Q. Efficientnetv2: Smaller models and faster training. International conference on machine learning. PMLR. 2021. pp. 10096–10106.
22. Xu W., Wang G., Sullivan A., Zhang Z. Towards learning affine-invariant representations via data-efficient cnns. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020. pp. 904–913.
23. Sabour S., Frosst N., Hinton G.E. Dynamic routing between capsules. Advances in neural information processing systems. 2017. vol. 30.
24. Stallkamp J., Schlipsing M., Salmen J., Igel C. The German traffic sign recognition benchmark: a multi-class classification competition. 2011 international joint conference on neural networks. IEEE. 2011. pp. 1453–1460.
25. Arora S., Du S.S., Li Z., Salakhutdinov R., Wang R., Yu D. Harnessing the power of infinitely wide deep nets on small-data tasks. arXiv preprint arXiv:1910.01663. 2019.
26. Krizhevsky A., Hinton G. et al. Learning multiple layers of features from tiny images. 2009. 60 p.
27. Barz B., Denzler J. Deep learning on small datasets without pre-training using cosine loss. Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020. pp. 1371–1380.

28. Hai Z., Liu X. Evolving parametrized Loss for Image Classification Learning on Small Datasets. arXiv preprint arXiv:2103.08249. 2021.
29. Lezama J., Qiu Q., Mus e P., Sapiro G. Ole: Orthogonal low-rank embedding-a plug and play geometric loss for deep learning. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 8109–8118.
30. LeCun Y., Cortes C., Burges C. MNIST handwritten digit database. 2010. vol. 2. 18 p.
31. Shorten C., Khoshgoftaar T.M. A survey on image data augmentation for deep learning. Journal of big data. 2019. vol. 6. no. 1. pp. 1–48.
32. Reinders C., Schubert F., Rosenhahn B. ChimeraMix: Image classification on small datasets via masked feature mixing. arXiv preprint arXiv:2202.11616. 2022.
33. Zhang X., Wang Z., Liu D., Lin Q., Ling Q. Deep adversarial data augmentation for extremely low data regimes. IEEE Transactions on Circuits and Systems for Video Technology. 2020. vol. 31. no. 1. pp. 15–28.
34. Ishii M., Sato A. Training deep neural networks with adversarially augmented features for small-scale training datasets. 2019 International Joint Conference on Neural Networks (IJCNN). IEEE. 2019. pp. 1–8.
35. Agarwal N., Sondhi A., Chopra K., Singh G. Transfer learning: Survey and classification. Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS 2020. 2021. pp. 145–155.
36. Zhao B., Wen X. Distilling visual priors from self-supervised learning. Computer Vision – ECCV 2020 Workshops. Springer. 2020. pp. 422–429.
37. Chen X., Fan H., Girshick R., He K. Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297. 2020.
38. Miranda G., Rubio-Manzano C. Image Classification Using Deep and Classical Machine Learning on Small Datasets: A Complete Comparative. 2022. DOI: 10.20944/preprints202201.0367.v1.
39. Christodoulou E., Ma J., Collins G.S., Steyerberg E.W., Verbakel J.Y., Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of clinical epidemiology. 2019. vol. 110. pp. 12–22.
40. Mojab N., Yu P.S., Hallak J.A., Yi D. Cvs: Classification via segmentation for small datasets. arXiv preprint arXiv:2111.00042. 2021.
41. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th international conference. Springer International Publishing, 2015. pp. 234–241.
42. Baid U., Ghodasara S., Mohan S., Bilello M., Calabrese E., Colak E., Farhani K., Kalpathy-Cramer J., Kitamura F.C., Pati S. et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314. 2021.
43. Menze B.H., Jakab A., Bauer S., Kalpathy-Cramer J., Farhani K., Kirby J., Burren Y., Porz N., Slotboom J., Wiest R., et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging. 2014. vol. 34. no. 10. pp. 1993–2024.
44. Bakas S., Akbari H., Sotiras A., Bilello M., Rozycki M., Kirby J.S., Freymann J.B., Farhani K., Davatzikos C. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific data. 2017. vol. 4. no. 1. pp. 1–13. Rohlfing T., Zahr N.M., Sullivan E.V., Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure. Human brain mapping. 2010. vol. 31. no. 5. pp. 798–819.



45. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 1998. vol. 86. no. 11. pp. 2278–2324.
46. Kingma D.P., Ba J. Adam A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. 2014.
47. Akkus Z., Sedlar J., Coufalova L., Korfiatis P., Kline T.L., Warner J.D., Agrawal J., Erickson B.J. Semi-automated segmentation of pre-operative low grade gliomas in magnetic resonance imaging. *Cancer Imaging*. 2015. vol. 15. pp. 1–10.
48. Liu X., Xing F., Marin T., El Fakhri G., Woo J. Variational inference for quantifying inter-observer variability in segmentation of anatomical structures. *Medical Imaging 2022: Image Processing*. 2022. vol. 12032. pp. 438–443.
49. Lin T.-Y., Dollar P., Girshick R., He K., Hariharan B., Belongie S. Feature pyramid networks for object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. pp. 2117–2125.
50. Huo Y., Xu Z., Xiong Y., Aboud K., Parvathaneni P., Bao S., Bermudez C., Resnick S., Cutting L., Landman B. 3D whole brain segmentation using spatially localized atlas network tiles. *NeuroImage*. 2019. vol. 194. pp. 105–119.
51. Paschali M., Gasperini S., Roy A.G., Fang M.Y.S., Navab N. 3DQ: Compact quantized neural networks for volumetric whole brain segmentation. *Medical Image Computing and Computer Assisted Intervention: 22nd International Conference*. Springer International Publishing, 2019. pp. 438–446.
52. Roy A.G., Conjeti S., Navab N., Wachinger C. Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation for structure-wise quality control. *NeuroImage*. 2019. vol. 195. pp. 11–22.
53. Henschel L., Conjeti S., Estrada S., Diers K., Fischl B., Reuter, M. Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline. *NeuroImage*. 2020. vol. 219. pp. 117012.
54. Coupe P., Mansencal, B., Clement, M., Giraud, R., de Senneville B.D., Ta V.T., Lepetit V., Manjon J.V. AssemblyNet: A large ensemble of CNNs for 3D whole brain MRI segmentation. *NeuroImage*. 2020. vol. 219(5). pp. 117026.
55. Dubois J., Alison M., Counsell S.J., Hertz-Pannier L., Huppi P.S., Benders M.J. MRI of the neonatal brain: a review of methodological challenges and neuroscientific advances. *Journal of Magnetic Resonance Imaging*. 2021. vol. 53. no. 5. pp. 1318–1343.
56. Ntaios G. Embolic stroke of undetermined source: JACC review topic of the week. *Journal of the American College of Cardiology*. 2020. vol. 75. no. 3. pp. 333–340.

**Sobyanin Kirill** — Junior researcher, Center for cognitive neuroscience, National Research University Higher School of Economics. Research interests: machine learning, application of mathematics and computer science in neurology and neuroscience. The number of publications — 9. k.sobyanin@gmail.com; 37a, Gagarin Blvd, 614060, Perm, Russia; office phone: +7(342)200-9555.

**Kulikova Sofya** — Ph.D., Associate Professor, Senior researcher, Center for cognitive neuroscience, National Research University Higher School of Economics. Research interests: modeling cognitive processes, MRI analysis, brain development. The number of publications — 19. spkulikova@hse.ru; 37a, Gagarin Blvd., 614060, Perm, Russia; office phone: +7(342)200-9555.

**Acknowledgements.** The publication was prepared within the framework of the Academic Fund Program at HSE University (grant № 23-00-026 "Development of automatic approaches to determine the etiology of cryptogenic stroke for the purpose of preventing secondary acute cerebrovascular accidents").