Математическая биология и биоинформатика
2020. Т. 15. № 2. С. 180–194. doi: 10.17537/2020.15.180

==========ИНФОРМАЦИОННЫЕ И ВЫЧИСЛИТЕЛЬНЫЕ ==========
==========ТЕХНОЛОГИИ В БИОЛОГИИ И МЕДИЦИНЕ==========

Классификация повреждений кожи по данным дермаскопии с использованием методов глубокого обучения

Щетинин Е.Ю.[1], Севастьянов Л.А.[2], Демидова А.В.[2], Кулябов Д.С.[2]

[1]*Финансовый университет при Правительстве РФ, Москва, Россия*
[2]*Российский университет дружбы народов, Москва, Россия*

Аннотация. В настоящей работе предложен подход к решению задачи обнаружения злокачественных новообразований кожи, а именно, меланомы, на основе анализа дермоскопических изображений с использованием методов глубокого обучения. Для этого разработана архитектура глубокой сверточной нейронной сети, которая была применена к обработке дермоскопических изображений различных повреждений кожи, содержащихся в наборе данных HAM10000. Исследуемые данные предварительно были подвергнуты обработке с целью устранения шума, загрязнений, а также изменения размеров и формата изображений. Кроме того, поскольку классы заболеваний являются несбалансированными, был выполнен ряд преобразований по их балансировке. Полученные таким образом данные были разбиты на два класса Melanoma и Benign. Компьютерные эксперименты по применению построенной глубокой нейронной сети на полученных таким образом данных показали, что предложенный подход обеспечивает точность на тестовой выборке 94 %, что превышает аналогичные результаты, полученные другими алгоритмами глубокого обучения.

Ключевые слова: *меланома, классификация, нейронные сети, интеллектуальные медицинские системы.*

ВВЕДЕНИЕ

Меланома – смертельная форма рака кожи, которая на ранней стадии часто не диагностируется или ошибочно диагностируется как доброкачественное поражение кожи. Ее раннее и точное выявление крайне важно, поскольку от этого зависит жизнь пациентов. В своей практике врачи привыкли полагаться на свой профессиональный опыт и оценивать повреждения каждого пациента на основе личного осмотра. Однако, такая система обнаружения поражений кожи является трудоемкой, поскольку нуждается в увеличении и освещении изображений кожи, чтобы улучшить четкость пигментных пятен [1–3]. Кроме того, процедура ручной дермоскопии более склонна к ошибкам, требует многолетнего опыта работы в сложных ситуациях, огромного количества визуальных исследований сходств и различий между различными поражениями кожи [4, 5].

Клинические исследования позволяют получить точность диагноза меланомы от 65 до 80 % , что до некоторого времени являлось неплохим результатом [6, 7]. Однако современные исследования утверждают, что использование дермоскопических изображений в постановке диагноза значительно повышает точность диагностики кожных поражений. Тем не менее, визуальные различия между меланомой и

доброкачественным поражением кожи могут быть очень незначительными, что делает диагностику сложной даже для врача-эксперта. Последние достижения в применении методов искусственного интеллекта в анализе медицинских изображений позволили считать разработку интеллектуальных медицинских систем диагностики на основе визуализации весьма перспективным направлением, которое помогут врачу в принятии более эффективных решений о здоровье пациентов и постановки диагноза уже на ранней стадии и в неблагоприятных условиях [8].

В настоящей работе исследован подход к решению проблемы классификации кожных заболеваний, а именно, обнаружения меланомы на ранней стадии, основанный на методах глубокого обучения. Для этого разработана архитектура глубокой нейронной сети, которая была применена к обработке дермоскопических изображений различных повреждений кожи, содержащихся в общедоступном наборе HAM10000, подготовленного международной организацией ISIC [9]. Компьютерные эксперименты по обучению и тестированию построенной глубокой нейронной сети на полученных таким образом данных показали, что предложенный подход обеспечивает точность на тестовой выборке 91 %, что превышает аналогичные результаты, полученные другими алгоритмами.

ОБЗОР СОВРЕМЕННЫХ ДОСТИЖЕНИЙ В ОБЛАСТИ КОМПЬЮТЕРНОЙ ОБРАБОТКИ ДЕРМОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ

Большинство классических методов в области классификации меланомы полагаются на ручной отбор признаков, таких как тип поражения (первичная морфология), конфигурация поражения (вторичная морфология), цвет, распределение, форма, текстура и неравномерность границ пигментного пятна. Затем, после извлечения основных характеристик изображений, для решения задачи классификации используются методы машинного обучения, такие как алгоритм *k*-ближайших соседей (*k*-NN), логистическая регрессия, деревья решений и другие [10].

Современные компьютерные исследования по диагностике кожных заболеваний с целью обнаружения меланомы активно внедряют методы глубокого обучения и направлены на совершенствование существующих и разработку новых моделей глубоких нейронных сетей, прежде всего сверточных нейронных сетей (CNN) [11–13]. В работе [14] предложена архитектура CNN для диагностики меланомы, где клинические изображения были предварительно обработаны таким образом, чтобы уменьшить освещенность изображения. Результаты исследований показали, что предложенный метод способен диагностировать случаи поражения меланомой в 70 % случаев. В работе [15] представлена экспертная система MED-NODE, чтобы помочь врачам в обнаружении меланомы. Предложенная система использовала извлеченные области поражения на изображении, затем вычисляла такие признаки, как цвет и текстура, а также визуальные атрибуты, предоставленные экспертами.

В работе [16] показано, что сверточные нейронные сети превосходят традиционные методы машинного обучения. Авторами предложен гибридный автоматический компьютеризированный метод классификации кожных заболеваний с использованием трех предварительно обученных глубоких сетей (VGG16, ResNet-18) для извлечения признаков. Извлеченные таким образом признаки затем используются для обучения машины опорных векторов на 150 изображениях из набора данных ISIC-2017. Emre Celebi с соавторами [17] предложили ансамбль пороговых методов определения границ поражения кожи.

Heckler A. C соавторами [18] применили методы глубокого обучения для классификации гистопатологической диагностики меланомы и сравнили результат с квалифицированными гистопатологами. В работе [20] разработана предварительно обученная нейронная сеть INCEPTIONV3 для классификации девяти классов

заболеваний кожи из набора дерматологических данных, содержащего 3374 дермоскопического изображения. Точность нейронной сети составила 72 %. Сотрудники Harangi B. [20] использовали ансамблевый метод DCNN (Deep Convolutional Neural Network), объединив результаты четырех различных архитектур и повысив точность классификации меланомы на наборе данных ISBI-2017 до 77 %.

Xie F. и Bovik A.C. [21] предложили метод сегментации кожных поражений, в котором модель глубокой нейронной сети U-net комбинируется с генетическим алгоритмом. В работе [22] предложен метод сегментации кожных поражений на дермоскопических изображениях из набора данных ISIC-2017 и их классификации различных типов рака кожи с использованием моделей глубоких нейронных сетей Mask R-CNN и U-net. Предложенный метод состоит в предварительной обработке и сегментации с использованием гибридного алгоритма обучения. Цель первого этапа – удаление шума с помощью метода фильтрации. На втором этапе изображения сегментируются на основе метода кластеризации. В работе [23] для распознавания меланомы было предложено использовать глубокую сверточную сеть ResNet50, точность которой на используемых данных составила 68 %.

ОПИСАНИЕ ДАННЫХ И ИХ ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА

В клинической дерматологии существует относительно немного наборов данных с цифровыми изображениями повреждений кожи. Большинство из этих наборов слишком малы и/или не являются общедоступными, что создает дополнительное препятствие для проведения исследований в данной области. Примерами таких наборов данных изображений, связанных с дерматологией, являются: Библиотека изображений Dermofit [24] – набор данных, содержащий 10 различных классов, включающих 1300 высококачественных изображений повреждений кожи. Dermnet – атлас заболеваний кожи с поддержкой веб-сайта содержит более 23 000 изображений кожи, разделенных на 23 класса [25]. В начале 2016 года состоялся международный симпозиум биомедицинских изображений (ISBI), опубликовавший набор данных для анализа повреждений кожи с целью раннего выявления меланомы.

С целью поддержки обучения клинических врачей-дерматологов и развития новых информационных технологий Международное общество по визуализации кожи (ISIC) разработало международный депозитарий дермоскопических изображений, известный как архив ISIC [9]. Этот набор данных содержит пигментированные поражения кожи, полученные с помощью стандартной дермоскопии. Каждый год ISIC пополняет свой архив новыми изображениями и продвигает задачу по внедрению компьютерных методов обнаружения меланомы и других заболеваний кожи. Созданный в рамках этой организации набор данных HAM10000 служил полигоном для проведения международных соревнований компьютерных алгоритмов ISIC-2018 Challenge. В 2019 году количество образцов в нем насчитывало более 25000 изображений для дермоскопии, доступных для обучения в 7 различных категориях:

- **Melanocytic nevi -** меланоцитарные невусы – доброкачественные новообразования меланоцитов [6705 изображений];
- **Melanoma** – меланома – злокачественное новообразование [1113 изображений];
- **Benign keratosis** – доброкачественный кератоз – общий класс, который включает себорейный кератоз, солнечный лентиго и лишайник-плоскоклеточный, как кератоз [1099 изображений];
- **Basal cell carcinoma** – базально-клеточная карцинома – распространенный вариант эпителиального рака кожи, который редко метастазирует, но при отсутствии лечения разрастается [514 изображений];
- **Actinic keratoses** – актинический кератоз и интраэпителиальная карцинома – распространенные неинвазивные варианты плоскоклеточного рака,

которые можно лечить локально без хирургического вмешательства [327 изображений];

- **Vascular lesions** – сосудистые поражения кожи от ангиом вишни до ангиокератом и пиогенных гранулем [142 изображений];
- **Dermatofibroma** – дерматофиброма – доброкачественное поражение кожи, рассматриваемое как доброкачественная пролиферация или воспалительная реакция на травму [115 изображений].

На рисунке 1 приведены примеры изображений из набора данных HAM10000. Размеры исходных дермоскопических изображений имеют размеры 1022 в ширину и 767 в высоту в пикселях. Перед использованием они предварительно были обработаны следующим образом: сначала была проведена нормализация значений пикселей из диапазона $[0, 255]$ в интервал $[0, 1]$, затем было изменено разрешение исходного изображения до 224×224 в пикселях.

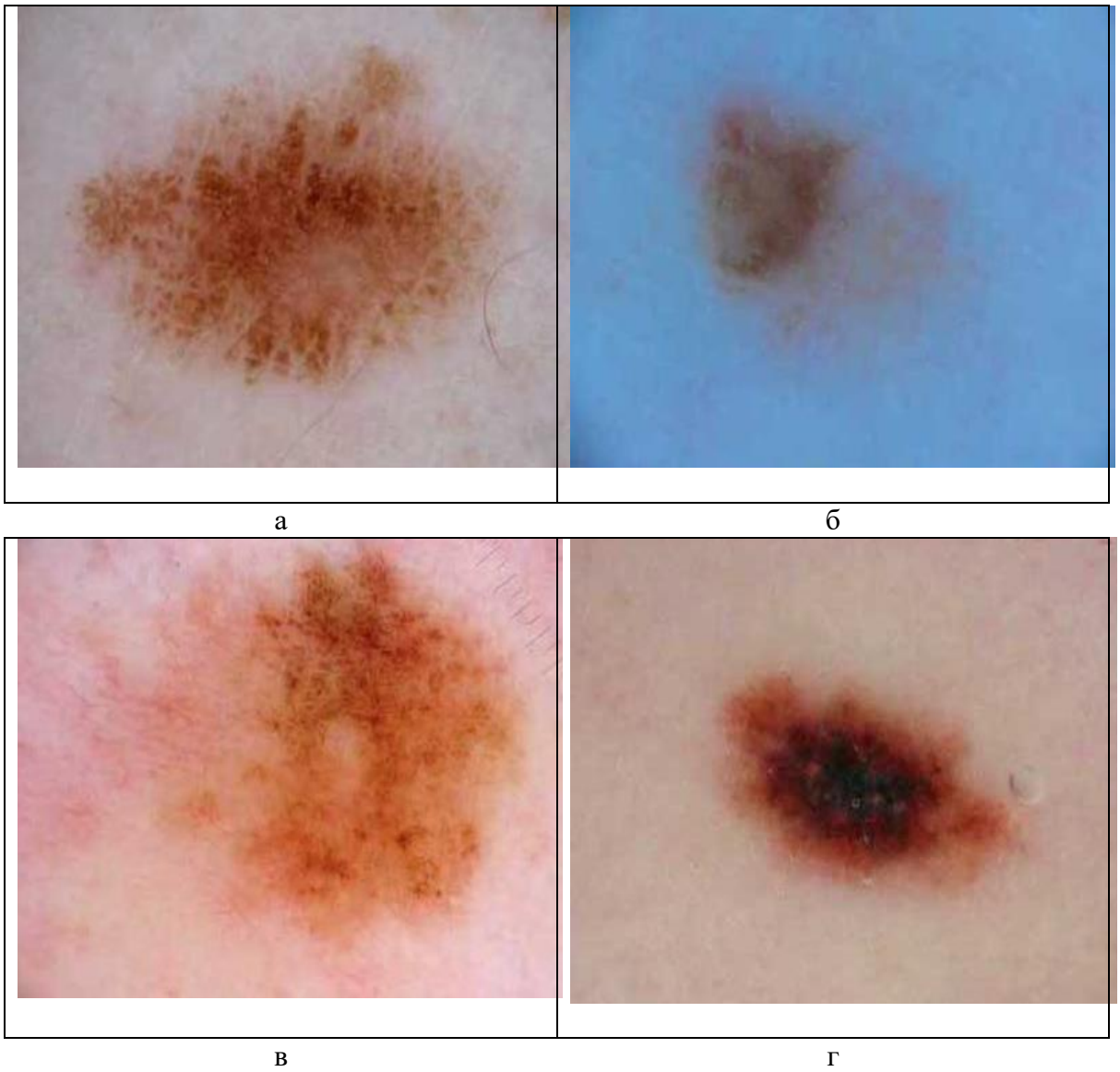

Рис.1. Примеры дермоскопических изображений из базы изображений HAM10000: а) – Melanocytic nevi; б) – Actinic keratoses; в) – Basal cell carcinoma; г) – Melanoma.

ОСНОВНЫЕ АРХИТЕКТУРЫ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ В АНАЛИЗЕ ДЕРМОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ

В настоящей работе для анализа дермоскопических изображений использовались сверточные нейронные сети (CNN) различными архитектурами, как наиболее эффективные модели глубоких нейронных сетей, применяемых в обработке изображений. Нами определены 7 основных шагов по созданию модели классификатора на основе глубокой сверточной сети:

Шаг 1 – Сверточный слой (Convolutional Layer). Создает карту объектов, которая содержит аргументы, записанные фильтром для входного изображения.

Шаг 2 – Слой объединения (Pooling Layer) – это слой, который использует функции с картой признаков в качестве входных данных и обрабатывает их с помощью различных статистических операций. Слой объединения в модели CNN обычно вставляется после нескольких слоев свертки. Слой объединения в архитектуре модели CNN позволяет постепенно уменьшать размер выходного объема в карте объектов, чтобы сократить число используемых параметров и вычислений в сети. В большинстве моделей сверточных сетей обычно применяется слой максимального пулинга (Max Pooling). Он делит выходные данные слоя свертки на несколько небольших сеток, а затем берет максимальное значение из каждой сетки для составления уменьшенной матрицы изображения.

Шаг 3 – Слой выравнивания (Flatten Layer). Этот слой служит связующим звеном между данными, полученными сетью, и выходным вектором, преобразуя многомерный выход предшествующего слоя в одномерный (вектор). В качестве входных признаков этого слоя используется объединенная карта объектов, а слой Flatten преобразует ее в столбец.

Шаг 4 – Полностью связанный слой (Fully Connected Layer) – это слой, в котором все активированные нейроны из предыдущего слоя соединены со всеми нейронами следующего слоя, и предназначен для преобразования размеров данных так, чтобы данные можно было линейно классифицировать.

Шаг 5 – Компиляция модели нейронной сети.

Шаг 6 –Построение оценок с использованием построенной архитектуры (Fitting the CNN).

Шаг 7. Оценка метрик точности, построение матрицы ошибок, вычисление функции потерь.

РАЗРАБОТКА МОДЕЛЕЙ НЕЙРОННЫХ СЕТЕЙ И КОМПЬЮТЕРНЫЕ ЭКСПЕРИМЕНТЫ

Для разработки моделей нейронных сетей нами была использована библиотека глубокого обучения Keras на языке Python [26]. Далее вся используемая терминология будет относиться к стандартным понятиям и характеристикам нейронных сетей, подробное объяснение которым можно найти на сайте Keras [26] или учебных пособиях, посвященных обучению нейронных сетей [27]. Модель CNN инициализирована в виде последовательности слоев методом Sequential, что является наиболее популярным подходом при построении моделей нейронных сетей. Далее, был добавлен сверточный слой Conv2D (шаг 1), с входными параметрами карты признаков изображения (32, 32, 3), где 32 – размер пространственных признаков входной карты, 3 – количество цветовых каналов (в данном случае – это цветность изображения в формате RGB). Свертка определяется следующими параметрами: размер шаблонов, извлекаемых из входных данных – 3×3; глубина выходной карты признаков – количество фильтров, вычисляемых сверткой. В данной модели первый сверточный слой выводит карту признаков размером (30, 30, 32) и вычисляет 32 фильтра по входным данным. Каждый из этих 32 выходных каналов содержит сетку 30×30 значений – карту ответов фильтра

Математическая биология и биоинформатика. 2020. Т. 15. № 2. doi: 10.17537/2020.15.180

на входных данных, определяющую ответ этого шаблона фильтра для разных участков входных данных. Последний параметр сверточного слоя – функция активации, которую мы используем для активации нейронов в нейронной сети, в нашем случае, это нелинейная функция ReLU.

На втором шаге был использован слой объединения (MaxPooling2D) с картой (3 × 3). Основной целью применения этого слоя является уменьшение числа коэффициентов в карте признаков для обработки, а также внедрения иерархий пространственных фильтров путем создания последовательных слоев свертки для просмотра все более крупных окон. Затем, на третьем шаге создаем вектор для полносвязного слоя, включив для этого в модель слой выравнивания. Далее строится полностью связанный слой (шаг 4). Слой Dense имеет 2 параметра – количество узлов для выходного слоя (128) и функция активации, которая также будет ReLU. Выходной слой имеет один узел, где используется функция активации ‘sigmoid’. Далее модель нужно скомпилировать и оптимизировать весовые коэффициенты и функцию потерь для оценки модели (шаг 5, шаг 6). В качестве оптимизатора нашей модели выбран алгоритм Adam [27]. Наконец, на шаге 7 производится оценка потерь, вычисление основных метрик точности модели. В качестве функции потерь выбрана функция кросс-энтропии. Для классификации двух классов в среде программирования Keras используется бинарная кросс-энтропия, а в случае более двух классов применяется категориальная кросс-энтропия [27]. В общем виде выражение для функции потерь класса кросс-энтропия имеет следующий вид:

$$L(p) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(p(s_i)\right) + \left(1 - y_i\right)\log\left(1 - p(s_i)\right)\right], \quad (1)$$

где y_i – значения меток классов {0, 1, ..}, $s_i = s(y_i, \omega)$ –прогноз i-го класса, полученный нейронной сетью, ω – параметры нейронной сети, $p(s_i) = \frac{1}{1+e^{-s_i}}$ – вероятность класса y_i, N – количество классов. Поскольку мы имеем два класса (доброкачественная или злокачественная опухоли), то $N = 2$. На рисунке 2 приведен граф архитектуры построенной нейронной сети.

РЕЗУЛЬТАТЫ КОМПЬЮТЕРНЫХ ЭКСПЕРИМЕНТОВ

Используемый в работе набор данных ISIC-2018, известный также как HAM10000, содержит 10015 изображений, предварительно разделенных на 8000 обучающих и 2015 тестовых изображений [9]. Компьютерные эксперименты проводились в два этапа. На первом этапе нами был создан корпус изображений из 1255 единиц, изображения были размечены как содержащие доброкачественные (класс 0 – Positive) и злокачественные повреждения кожи (класс 1 – Negative). Далее все изображения были разделены на обучающую и тестовую выборки. Обучающая выборка содержит 878 изображения. Тестовая выборка содержит 377 изображений. В качестве оптимизационной функции использовался класс ADAM и бинарная перекрестная энтропия как функция потерь (1), для оценивания качества классификации использована метрика *Accuracy*.

Также еще одним показателем производительности модели нейронной сети, используемым нами в статье, является матрица неточностей (confusion matrix) [31]. В таблице 1 приведена матрица неточностей, полученная в результате работы нейронной сети DCNN на тестовой выборке. В ней величины *TP* (True Positive) обозначают число верно распознанных нейронной сетью доброкачественных повреждений кожи; *TN* (True Negative) – число верно распознанных нейронной сетью злокачественных повреждений кожи; *FP* (False Positive) – число ложно распознанных нейронной сетью повреждений кожи как доброкачественных, которые в действительности принадлежат к классу злокачественных повреждений; *FN* (False Negative) – число ложно распознанных

нейронной сетью повреждений кожи как злокачественных, в действительности являющихся доброкачественными. В таблице 1 представлены характеристики матрицы неточностей в случае бинарной классификации и приведены ее значения, полученные в результате работы нейронной сети DCNN на описанных выше данных. На рисунке 3 также нами приведены результаты построения матрицы неточностей и ее нормализованного вида с использованием разработанной авторами программы plot_confusion_matrix на языке программирования Python [32].

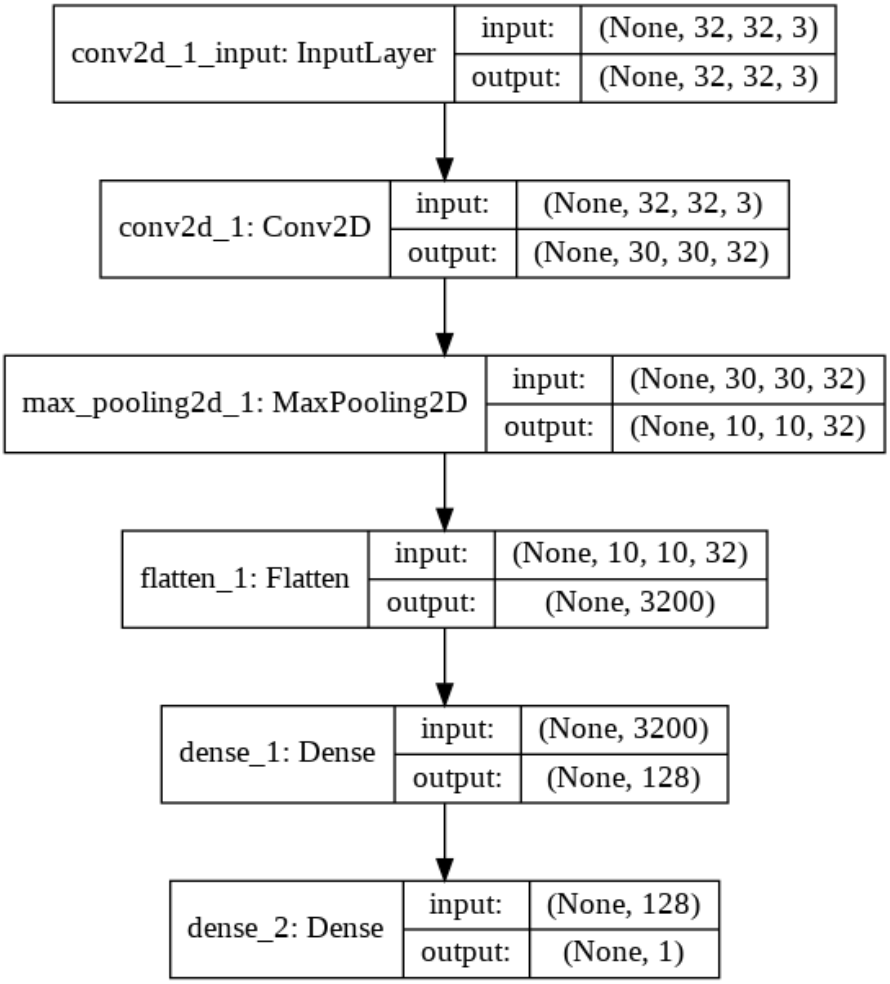

Рис. 2. Архитектура глубокой сверточной сети DCNN для классификации дермоскопических изображений.

Таблица 1. Матрица неточностей, полученная на обучающей выборке для бинарной классификации повреждений кожи на данных HAM10000

Матрица неточностей (confusion matrix)		Прогнозируемый класс	
		Доброкачественное повреждение	Злокачественное повреждение
Истинный класс	Доброкачественное повреждение	*TP* = 186	*FN* = 11
	Злокачественное повреждение	*FP* =11	*TN* = 169

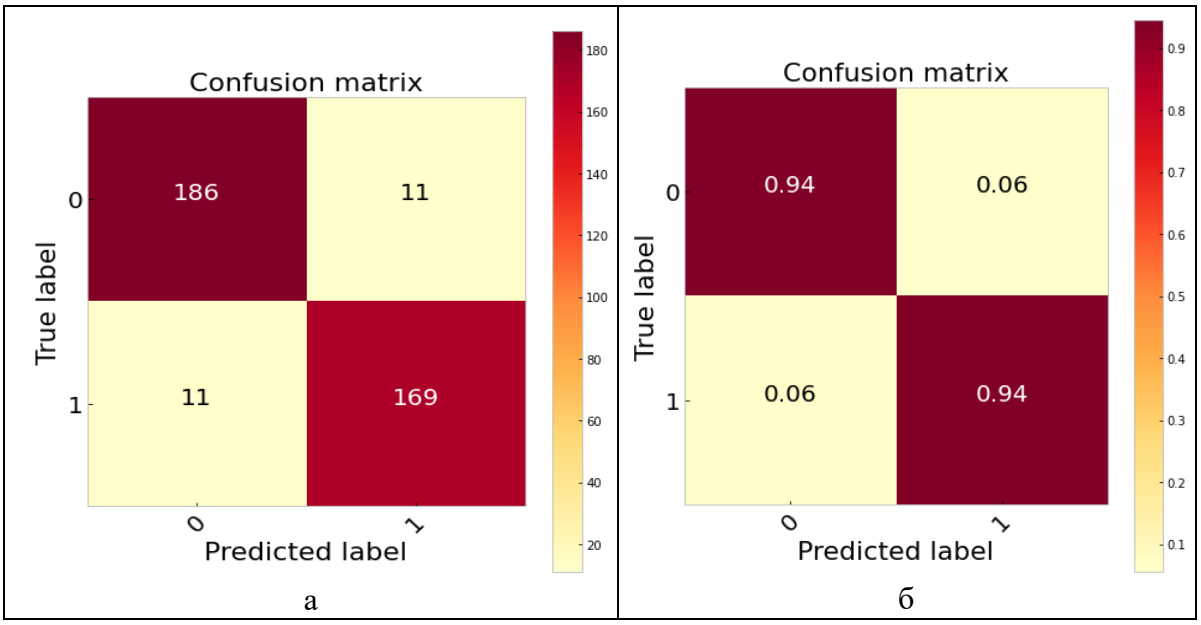

Рис. 3. Результаты построения матрицы неточностей с использованием программы plot_confusion_matrix: а) – матрица неточностей б) – нормализованная матрица неточностей. 0 – класс доброкачественных повреждений (Positive); 1 – класс злокачественных повреждений (Negative).

Далее с применением процедур accuracy_score, precision_score, recall_score и f1_score из библиотеки sklearn.metrics были получены оценки метрик точности: *Accuracy* = 0.94, *Precision* = 0.945, *Recall* = 0.939, *F1* = 0.914. Аналогичные компьютерные эксперименты были проведены для архитектуры глубоких сверточных сетей U-net [22] и LeNet [14]. Все результаты вычислений приведены в таблице 2.

Таблица 2. Значения метрик точности, полученные различными моделями нейронных сетей на тестовой выборке

Модель нейронной сети	*Accuracy*, %	*Precision*,%	*Recall*,%	*F1*,%
DCCN (рис. 5)	94	94.5	93.9	94.19
U-net [30]	74	69	64	66.4
LeNet [14]	78.2	72	68	70

На втором этапе исследований были проведены компьютерные эксперименты по многоклассовой классификации описанных выше данных HAM10000. Для решения поставленной задачи в работе использовались модели глубоких нейронных сетей, такие как ResNet50, MobileNet, InceptionV3, а также VGG16. Выбор данных моделей обусловлен их популярностью у исследователей и эффективностью в применении к обработке сложных изображений. Также в настоящей работе предложена стратегия обучения и тонкой настройки параметров модели глубокой нейронной сети ResNet34, предварительно обученной на данных ImageNet [28]. Для ее реализации использовалась библиотека глубокого обучения FastAI [29], одна из эффективных для обучения нейронных сетей. Библиотека FastAI предоставляет стратегии повышения эффективности обучения глубоких нейронных сетей, основанные на варьировании скорости обучения, устанавливая ее различные значения для различных частей сети. Ее особенностью и преимуществом является применение метода заморозки слоев сети, когда по умолчанию сначала обучаются только последние полностью соединенные слои (fully connected layers). Для этого использованы процедуры *fit fit_one_cycle* из библиотеки FastAI [29]. Процедура *fit* – это программная реализация метода обучения нейронной

сети с постоянной скоростью обучения LR, в то время как метод *fit_one_cycle* использует принцип одного цикла, когда скорость обучения меняется с течением времени для достижения лучшей точности. После их последовательного применения, когда полностью соединенные слои обучены, можно разморозить другие слои и обучить всю сеть.

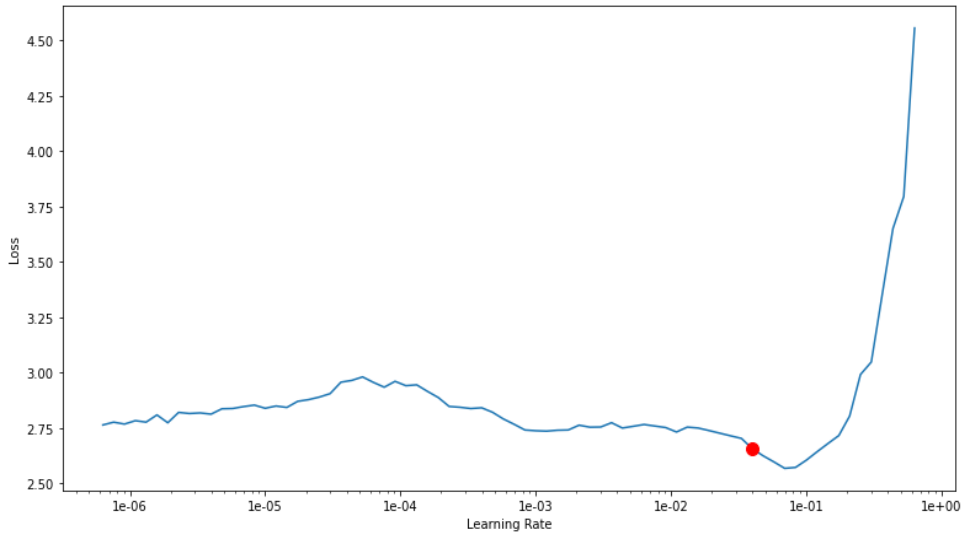

Рис. 4. График зависимости потерь от скорости обучения с использованием процедуры lr_find().

Скорость обучения является наиболее важным гиперпараметром для обучения нейронной сети. Чтобы найти ее оптимальные значения, в процессе обучения используются процедуры *lr_find и learner* из библиотеки FastAI [29]. Процедура *lr_find* запускает сначала пробное обучение, начиная с низкой скорости обучения и увеличивая ее экспоненциально с каждой партией (batch). Из графика зависимости потерь от скорости обучения мы можем выбрать самый высокий уровень обучения, где наши потери все еще значительно снижаются. На рисунке 5 представлен график зависимости потерь от значения скорости обучения LR в процессе обучения и тонкой настройки построенной нами нейронной сети ResNet34. Затем, после того, как оптимальное значение LR найдено, выполняется процедура *learner.fit_one_cycle* обучения сети с полученным значением LR. Его результаты передаются на следующий этап, который состоит в разморозке верхних слоев сети и тонкой настройке всей сети с использованием процедур *learner.unfreeze(), learner.lr_find(),* а также *learner.fit_one_cycle* из все той же библиотеки FastAI.

На данном этапе исследований для оценивания качества классификации изображений нами использовался метрический показатель ROC_*AUC* [30]. Кривая рабочей характеристики приемника (Receiver Operating Characteristic, ROC-кривая) часто используется для оценки качества классификации. Напомним, что график ROC-кривой представляет собой зависимость показателя *TPR* (True Positive Rate) от показателя *FPR* (False Positive Rate). Доля верных положительных исходов *TPR* показывает, какую долю из объектов класса Positive (в нашем случае – это Melanoma) алгоритм предсказал верно:

$$TPR = \frac{TP}{TP + FN}$$

Доля ложных положительных исходов (*FPR*) показывает, какую долю из объектов класса Negative (все остальные классы повреждений кожи) алгоритм предсказал неверно:

$$FPR = \frac{FP}{FP + TN}$$

ROC-кривая сравнивает наличие истинно положительных и ложных положительных исходов на каждом вероятностном пороге (т.е. вероятности, при которой наблюдение предсказывается как класс). Построив график кривой ROC, можно увидеть, насколько хорошо работает классификатор. Чем он лучше, тем резче поднимается график наверх и сильнее отличается от случайного классификатора. По умолчанию в библиотеке *scikit-learn* языка программирования Python принято, что исход принадлежит положительному классу, если вероятностный порог больше 0.5. Однако вместо этого требуется сместить порог в ту или иную сторону по ряду специфических обстоятельств. Например, в том случае, когда ложно положительный исход дорого обходится, и предпочтительнее использовать модель с более высоким вероятностным порогом (аналогично можно оперировать с ложно отрицательными предсказаниями). Пусть нам не удастся предсказать некоторые истинно положительные исходы, но, когда исход предсказан как положительный, мы будем более уверены в том, что этот прогноз является правильным, чем в случае стандартной модели. В нашем случае, поскольку положительным классом являются изображения с доброкачественными повреждениями кожи, то разумно повысить требования к предсказаниям из этого класса (за счет снижения величины *TPR* и повышения порога), при этом уменьшить долю отрицательных исходов (злокачественные повреждения), предсказуемых как положительные (за счет снижения величины *FPR*). Этот компромисс и реализуется с помощью ROC-кривой, которая также может использоваться как общий метрический показатель качества классификации модели. Чем лучше модель, тем выше ее ROC-кривая, тем больше площадь под ней.

Площадь под ROC-кривой *AUC* (Area Under Curve) также является эффективной характеристикой качества классификации. Данный показатель часто используется для сравнительного анализа нескольких моделей классификации [30]. Кроме того, критерий ROC_*AUC* более устойчив к несбалансированным классам и может быть интерпретирован как вероятность того, что случайно выбранный образец из положительного класса будет проранжирован классификатором выше (будет иметь более высокую вероятность), чем случайно выбранный объект из отрицательного класса.

На рисунке 5 представлена матрица неточностей для модели ResNet34, из которой было получено значение показателя точности *Accuracy* на тестовой выборке равное 0.8875. На рисунках 6 и 7 построены графики ROC-кривых, а также приведены *AUC*-показатели для всех классов повреждений кожи из HAM1000, полученные с использованием моделей ResNet34 и MobileNet соответственно. В таблице 3 приведены результаты применения различных моделей нейронных сетей для классификации изображений повреждений кожи из HAM10000. Из нее следует, что модель ResNet-34 обладает значимо большей точностью, чем остальные исследованные нами модели нейронных сетей.

Таблица 3. Значения метрики *Accuracy*, полученные различными моделями нейронных сетей на тестовой выборке

Модель нейронной сети	*Accuracy*,%
ResNet-34_FastAI	88.75
Inception_v3 [21]	72
MobileNet	78.2
VGG16 [18]	84

Все эксперименты проводились с использованием компьютера, оснащенного процессором Intel Core i5, 8 ГБ DDRAM и графической картой NVIDIA GeForce 920M. Расчетное время, затраченное на получение достигнутой точности, первом случае составило от нескольких минут до получаса в зависимости от количества эпох (от 25 до 100). Во втором случае на достижение указанной точности потребовалось использование

среды Colab.Research.Google.com и графических процессоров K80c 11 ГБ или Тесла T4c 14 ГБ. Время вычислений составило около 11 часов.

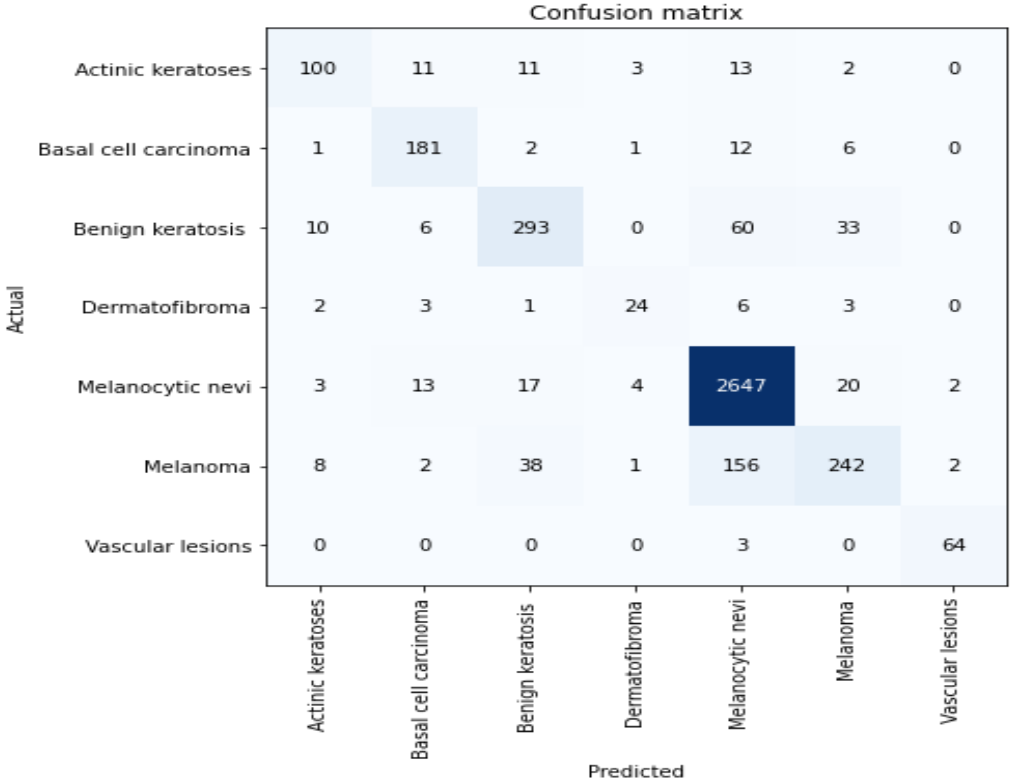

Рис. 5. Матрица неточностей модели нейронной сети ResNet34.

Необходимо отметить, что в целом практически все применявшиеся нейронные сети обладают достаточно высокой точностью при классификации заболеваний, однако, как можно заметить, показатели *AUC* для класса меланома (Melanoma) различаются существенно. Так, для ResNet34 показатель *AUC* = 79.3 %, для MobileNet *AUC* = 71.1 %, а для остальных моделей он оказался еще ниже. Таким образом, можно утверждать, что при общем высоком качестве работы исследованных нейронных сетей, построенная нами сеть ResNet34 оказалась более точной при распознавании меланомы. Сравнивая два подхода к классификации и распознаванию заболеваний кожи, можно сказать, что в первом случае бинарной классификации решается задача обнаружения злокачественного новообразования. Построенная нейронная сеть с поставленной проблемой справляется с небольшими затратами вычислительных ресурсов и с достаточной высокой точностью. Во втором случае, когда необходимо поставить диагноз меланомы, т.е. решить задачу многоклассовой классификации на несбалансированных классах [32], требуется более сложная модель нейронной сети, ее аккуратная тонкая настройка и дообучение. Естественно, что для этого требуются значительные затраты вычислительных ресурсов.

Одних только дермоскопических изображений недостаточно для дерматологического исследования или для надежной дистанционной диагностики пациента. Дермоскопические изображения не имеют контекста. Для их включения в клинические исследования необходимо создать протокол получения изображения, который включает в себя панорамные изображения всего тела пациента, а также аппроксимативные изображения каждого поражения кожи, которые взяты с помощью линейки или другой системы отсчета, видимой на изображении, чтобы предоставить контекстную информацию о размере поражения. Полученные таким образом изображения также важны для пациента, уже находящегося на лечении, чтобы позволить сопровождающему врачу следить за развитием поражения. Как панорамные, так и цифровые изображения должны выполняться в соответствии с протоколом, который

гарантирует, что изображения находятся в фокусе, сняты с правильного расстояния и с правильной освещенностью. В любом случае, когда соответствующие повреждения кожи не могут быть надежно диагностированы с помощью стандартной техники дермоскопии, применяемой в настоящее время, требуется биопсия для установления или подтверждения клинического диагноза.

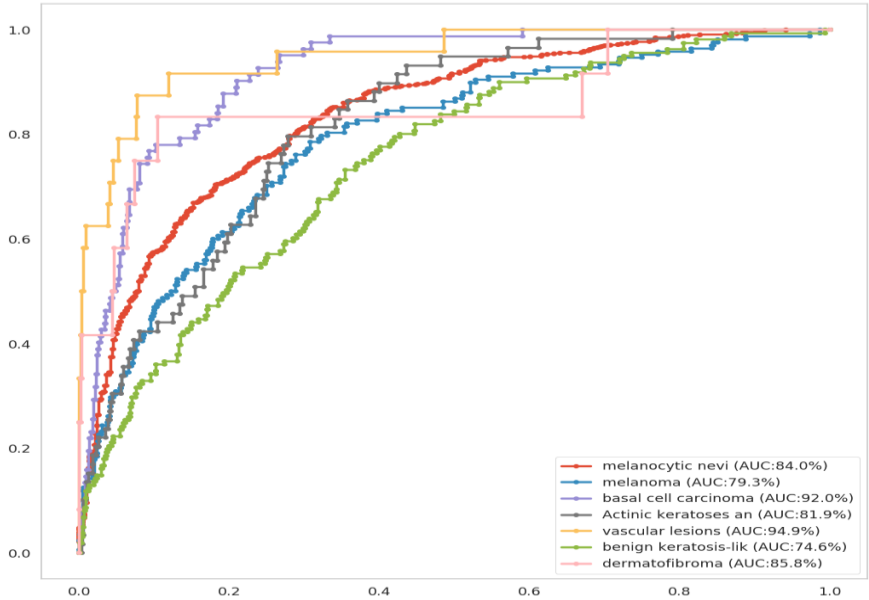

Рис. 6. ROC-кривые – показатели *AUC* – для классов данных HAM1000, полученные с использованием модели сети ResNet34.

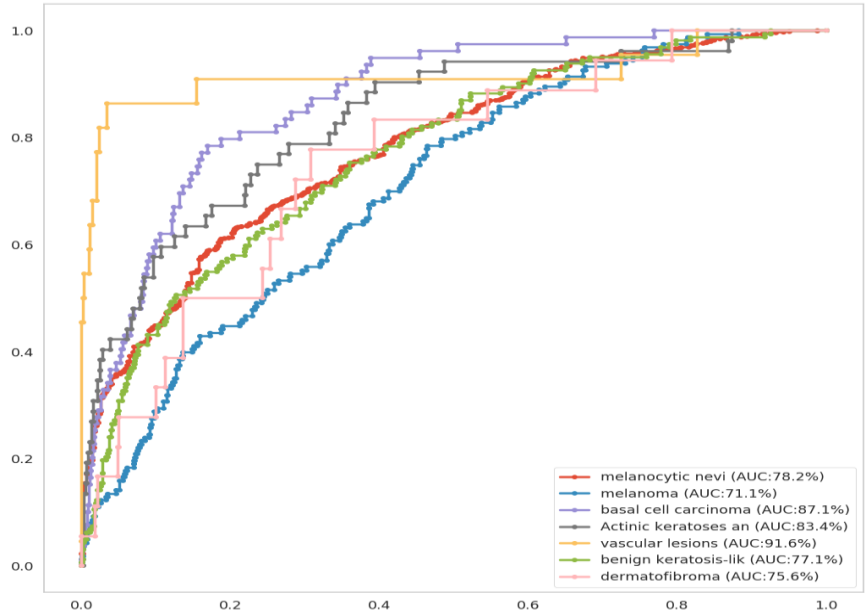

Рис. 7. ROC-кривые для классов данных HAM1000, полученные с использованием MobileNet.

ЗАКЛЮЧЕНИЕ

В работе исследована задача классификации заболеваний кожи и распознавания меланомы среди остальных диагностированных поражений кожи. Для ее решения реализована глубокая сверточная нейронная сеть, которая далее была применена для решения поставленной задачи на примере сформированного набора цифровых

изображений, извлеченных из HAM10000, предоставленных международным сообществом ISIC (International Skin Imaging Collaboration). Разработанная нейронная сеть включает в себя различные методы предварительной обработки изображений с целью получения более информативной и сбалансированной обучающей и контрольной выборок. Точность модели на тестовой выборке составила 94 %. Также были проведены компьютерные эксперименты по многоклассовой классификации заболеваний кожи HAM10000 с использованием группы моделей глубоких нейронных сетей, предварительно обученных на ImageNet, и настроенных на данные HAM10000. Точность настроенной нами модели ResNet34 составила 88 %, а точность распознавания ею меланомы составила 79 %.

Необходимо подчеркнуть, что первый этап компьютерных экспериментов направлен на обнаружение в исследуемых данных злокачественных новообразований, которое, в том числе, включает и меланому. На втором этапе задача состояла в обнаружении именно меланомы, как заболевания кожи, что потребовало разработки и применения более сложных нейронных сетей, их достаточно длительного обучения, и затрат значительных вычислительных ресурсов. Поэтому первый подход может быть использован, например, в телемедицине, качестве экспресс-метода постановки предварительного диагноза, как сигнала пациенту к обращению к специалистам для дальнейшего уточнения диагноза. Второй подход обладает рядом преимуществ и недостатков. Прежде всего, несмотря на направленность на выявление именно меланомы, он ограничен вычислительными затратами. С другой стороны, в отсутствие возможности личной консультации профессионального дерматолога, получение диагноза с высокой точностью на основе предложенного подхода остается чуть ли не единственным выходом в решении поставленной задачи.

Нами показано, что применение глубоких нейронных сетей является перспективным подходом к решению задачи автоматизированного распознавания различных дерматологических заболеваний, в том числе таких опасных как меланома. Наши дальнейшие исследования будут направлены на улучшение архитектуры глубоких нейронных сетей для повышения точности, исследование возможности применения других моделей глубоких нейронных сетей. Кроме того, будут предприняты усилия, направленные на то, чтобы сделать эту модель доступной и пригодной для использования в виде приложения для мобильных устройств в автоматизированных диагностических системах телемедицины.

СПИСОК ЛИТЕРАТУРЫ

1. Siegel R.L., Miller K.D., Jemal A. Cancer statistics 2019. *CA: a cancer journal for clinicians*. 2019. V. 69. № 1. P. 7–34. doi: 10.3322/caac.21551
2. *American Cancer Society: Cancer facts and figures 2018.* URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf. (дата обращения: 09.07.2020).
3. Robin Marks M.B.B.S. An overview of skin cancers. *Cancer* (*Supplement*). 1995. V. 75. № 2. P. 607–612. doi: 10.1002/1097-0142(19950115)75:2+<607::AID-CNCR2820751402>3.0.CO;2-8
4. Rogers H.W., Weinstock M.A., Feldman S.R., Coldiron B.M. Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population. 2012. *JAMA Dermatology*. 2015. V. 151. № 10. P. 1081–1086. doi: 10.1001/jamadermatol.2015.1187
5. Gandhi S.A., Kampp J. Skin cancer epidemiology, detection, and management. *Medical Clinics*. 2015. V. 99. № 6. P. 1323–1335. doi: 10.1016/j.mcna.2015.06.002

6. Kittler H., Pehamberger H., Wolff K., Binder M. Diagnostic accuracy of dermoscopy. *Lancet Oncology*. 2002. V. 3. № 3. P. 159–165. doi: 10.1016/s1470-2045(02)00679-4
7. Salerni G., Teran T., Puig S., Malvehy J., Zalaudek I., Argenziano G., Kittler H. Meta-analysis of digital dermoscopy follow-up of melanocytic skin lesions: a study on behalf of the International Dermoscopy Society. *Journal of the European Academy of Dermatology and Venereology*. 2013. V. 27. № 7. P. 805–814. doi: 10.1111/jdv.12032
8. *The International Symposium on Biomedical Imaging* (*ISBI2016*). URL: https://biomedicalimaging.org/2016/ (дата обращения: 09.07.2020).
9. *International Skin Imaging Collaboration*: Melanoma Project Website. URL: https://isic-archive.com/ (дата обращения: 09.07.2020).
10. Dreiseitl S., Ohno-Machado L., Kittler H., Vinterbo S., Billhardt H., Binder M. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *Journal of Biomedical Informatics.* 2001. V. 34. № 1. P. 28–36.
11. Esteva A., Kuprel B., Thrun S. *Deep networks for early stage skin disease and skin cancer classification*: Project Report. Stanford University, 2015.
12. Harangi B. Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of biomedical informatics*. 2018. V. 86. P. 25–32. doi: 10.1016/j.jbi.2018.08.006
13. Yuan Y., Chao M., Lo Y.-C. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE Transactions on Medical Imaging*. 2017. V. 36. № 9. P. 1876–1886. doi: 10.1109/TMI.2017.2695227
14. Nasr-Esfahani E., Samavi S., Karimi N., Soroushmehr S.M.R., Jafari M.H., Ward K., Najarian K. Melanoma detection by analysis of clinical images using convolutional neural network. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Orlando, FL, 2016. P. 1373–1376. doi: 10.1109/EMBC.2016.7590963
15. Giotis I., Molders N., Land S., Biehl M., Jonkman M.F., Petkov N. MED-NODE: a computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Systems with Applications*. 2015. V. 42. № 19. P. 6578–6585. doi: 10.1016/j.eswa.2015.04.034
16. Mahbod A., Schaefer G., Wang C., Ecker R., Ellinge I. Skin lesion classification using hybrid deep neural networks. In: *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019. P. 1229–1233.
17. Celebi M.E., Wen Q., Hwang S. Lesion Border Detection in Dermoscopy Images Using Ensembles of Thresholding Methods. *Ski. Res. Technol.* 2013. V. 19. № 1. P. e252–e258. doi: 10.1111/j.1600-0846.2012.00636.x
18. Hekler A., Utikal J.S., Enk A.H., Berking C., Klode J., Schadendorf D., Jansen P., Franklin C., Holland-Letz T., Krahl D., et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. *European Journal of Cancer*. 2019. V. 115. P. 79–83. doi: 10.1016/j.ejca.2019.04.021
19. 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. V. 542. № 7639. P. 115–118. doi: 10.1038/nature22985
20. Harangi B. Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of Biomedical Informatics*. 2018. V. 86. P. 25–32. doi: 10.1016/j.jbi.2018.08.006.
21. Xie F., Bovik A.C. Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm. *Pattern Recognition*. 2013. V. 46. № 3. P. 1012–1019.

22. Codella N.C., Nguyen Q.B., Pankanti S., Gutman D.A., Helba B., Halpern A.C. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development.* 2017. V. 61. № 4. P. 5–11.
23. Yu L., Chen H., Dou Q., Qin J., Heng P.A. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Transactions on Medical Imaging*. 2016. V. 36. № 4. P. 994–1004. doi: 10.1109/TMI.2016.2642839
24. Ballerini, L., Fisher, R. B., Aldridge, B., Rees, J. A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. *In Color Medical Image Analysis*. 2013. P. 63–86. Springer Netherlands.
25. *Dermnet - Skin Disease Atlas*. URL: http:// https://www.linksmedicus.com/main-menu/medical-images/dermnet-skin-disease-atlas/ (дата обращения: 09.07.2020).
26. Библиотека глубокого обучения Keras. URL: https://www.keras.io/. (дата обращения: 09.07.2020).
27. Джулли А., Пал С. *Библиотека Keras – инструмент глубокого обучения*. ДМК Пресс, 2017. 294 с. ISBN 978-5-97060-573-8
28. *База данных ImageNet*. URL: http://www.image-net.org/ (дата обращения: 09.07.2020).
29. *GitHub*. URL: https://github.com/fastai/fastai/ (дата обращения: 09.07.2020).
30. Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006. V.27. № 8. P. 861–874. doi: 10.1016/j.patrec.2005.10.010
31. Библиотека машинного обучения на языке программирования Питон. URL: https://www.sklearn.org/auto_examples/model_selection/plot_confusion_matrix.html. (дата обращения: 09.07.2020).
32. Севастьянов Л.А., Щетинин Е.Ю. О методах повышения точности многоклассовой классификации на несбалансированных данных. *Информатика и ее применения.* 2020. Т. 14. вып. 1. С. 67–74.

Рукопись поступила в редакцию 19.05.2020, переработанный вариант поступил 09.10.2020.
Дата опубликования 17.10.2020.

Skin Lesion Classification Using Deep Learning Methods

Shchetinin E.Yu.[1], Demidova A.V.[2], Kulyabov D.S.[2], Sevastyanov L.A.[2]

[1]*Financial University under the Government of Russian Federation, Moscow, Russia*
[2]*Russian Peoples Friendship University*

Abstract. In this paper, we propose an approach to solving the problem of recognizing skin lesions, namely melanoma, based on the analysis of dermoscopic images using deep learning methods. For this purpose, the architecture of a deep convolutional neural network was developed, which was applied to the processing of dermoscopic images of various skin lesions contained in the HAM10000 data set. The data under study were preprocessed to eliminate noise, contamination, and change the size and format of images. In addition, since the disease classes are unbalanced, a number of transformations were performed to balance them. The data obtained in this way were divided into two classes: Melanoma and Benign. Computer experiments using the built deep neural network based on the data obtained in this way have shown that the proposed approach provides 94% accuracy on the test sample, which exceeds similar results obtained by other algorithms.

Key words: *skin lesion, melanoma, classification, deep learning, HAM1000.*

Математическая биология и биоинформатика. 2020. Т. 15. № 2. doi: 10.17537/2020.15.180