

# Невыпуклая оптимизация с положительно-отрицательной оценкой момента и ее применение для нейросетевого распознавания рака кожи

П.А. Ляхов<sup>1,2</sup>, У.А. Ляхова<sup>1,2</sup>, Р.И. Абдулкадиров<sup>2</sup>

<sup>1</sup> Северо-Кавказский федеральный университет, 355017, Россия, г. Ставрополь, ул. Пушкина, д. 1;

<sup>2</sup> Северо-Кавказский центр математических исследований, 355017, Россия, г. Ставрополь, ул. Пушкина, д. 1

## Аннотация

Основная проблема использования стандартных методов оптимизации заключается в необходимости изменять все параметры шагами одинакового размера, независимо от поведения градиента. Более эффективный способ оптимизации нейронной сети состоит в том, чтобы установить адаптивные размеры шага для каждого параметра. Стандартные методы основаны на квадратных корнях экспоненциальных оценок моментов квадратов прошлых градиентов и не используют локальное изменение градиентов. В работе представлены методы адаптивной невыпуклой и доверительной оптимизации с положительно-отрицательной оценкой моментов с соответствующими теоретическими гарантиями сходимости. Данные подходы позволяют более точно сходиться функции потери в области глобального минимума за меньшее количество итераций. Использование преобразований положительно-отрицательной оценки момента и дополнительного параметра, регулирующего размер шага, позволяют обходить локальные экстремумы для достижения более высокой производительности по сравнению с аналогичными методами. Внедрение разработанных алгоритмов в процесс обучения различных архитектур мультимодальных нейросетевых систем анализа гетерогенных данных позволило повысить точность распознавания пигментных новообразований кожи на 2,33 – 5,69 процентных пункта по сравнению с известными методами оптимизации. Мультимодальные нейросетевые системы анализа разнородных дерматологических данных, обученные с применением предложенных алгоритмов оптимизации, могут использоваться в качестве инструмента вспомогательной медицинской диагностики, который позволит сократить потребление финансовых и трудовых ресурсов, задействованных в медицинской отрасли, а также повысить шанс раннего выявления пигментных онкопатологий.

**Ключевые слова:** оптимизация, натуральный градиентный спуск, искусственный интеллект, мультимодальные нейронные сети, разнородные данные, рак кожи, меланома.

**Цитирование:** Ляхов, П.А. Невыпуклая оптимизация с положительно-отрицательной оценкой момента и ее применение для нейросетевого распознавания рака кожи / П.А. Ляхов, У.А. Ляхова, Р.И. Абдулкадиров // Компьютерная оптика. – 2024. – Т. 48, № 2. – С. 260-271. – DOI: 10.18287/2412-6179-CO-1308.

**Citation:** Lyakhov PA, Lyakhova UA, Abdulkadirov RI. Non-convex optimization with using positive-negative moment estimation and its application for skin cancer recognition with a neural network. Computer Optics 2024; 48(2): 260-271. DOI: 10.18287/2412-6179-CO-1308.

## Введение

В настоящее время медицина является одной из наиболее перспективных областей для внедрения искусственного интеллекта [1]. Системы вспомогательной медицинской диагностики по точности и эффективности способны сравняться с возможностью специалистов, при этом минимизируя влияние человеческого фактора [2]. Интеллектуальные системы распознавания пигментных образований кожи на сегодняшний день могут превзойти по точности визуальной диагностики дерматологов-онкологов [3, 4]. Данные системы позволяют анализировать пигментные поражения кожи более быстрым, удобным и доступным способом [5]. Однако подобные системы не могут заменить решающее мнение патологоанатома и дерматолога-онколога в диагностике рака кожи в связи с возможностью ложноотрицательных случаев

прогнозирования [6]. Поэтому в настоящее время актуальной становится разработка высокоточных интеллектуальных систем, которые могут использоваться в качестве вспомогательных диагностических инструментов для выявления злокачественных новообразований на ранних стадиях.

На сегодняшний день основной проблемой в современной теории нейронных сетей является разработка математических методов, позволяющих ускорить процесс обучения и повысить точность интеллектуального распознавания данных. Данную проблему можно решить с помощью адаптивной выборки данных [7] или же введения метаданных [8]. Подобная модификация позволяет улучшить процесс обучения искусственного интеллекта, не затрачивая дополнительных временных ресурсов. Однако архитектура соответствующей модели становится более сложной и процесс минимизации функции потерь затрудняется.

Вследствие чего возникает необходимость в применении современных алгоритмов оптимизации.

Наиболее распространенными алгоритмами оптимизации в современных нейронных сетях являются стохастический градиентный спуск с импульсом и модификацией Нестерова (SGDM Nesterov) [9] и адаптивная оценка моментов (Adam) [10]. Первый метод минимизирует целевую функцию обучения с помощью направления градиентов, которые подкрепляются значениями импульса и демпфированием для избежания появления затухающих и взрывающихся градиентов. Второй метод, помимо направления градиентов, учитывает средние экспоненциальные сдвиги градиента и его квадрата, что позволяет сходиться к точкам экстремума за меньшее число итераций. Данные подходы способны регулировать значения весовых коэффициентов в глубоких нейронных сетях, которые могут быть подкреплены метаданными [11]. Однако в работе [12] было показано, что SGDM Nesterov и Adam не гарантируют сходимость в глобальном минимуме, вследствие чего точность распознавания нейронной сети не увеличивается. Данный недостаток можно наблюдать в процессе минимизации функции Растригина, где приведенные алгоритмы не способны достигать глобального минимума и сходятся в локальном. Для функции Розенброка процесс сходимости требует слишком много итераций. Для решения этой проблемы были предложены более передовые методы оптимизации, такие как DiffGrad [13], Yogi [14] и AdaBelief [15].

Представленные подходы для минимизации функции потерь позволили увеличить точность распознавания в сверточных нейронных сетях благодаря введению параметра трения и преобразования экспоненциального среднего сдвига квадрата градиента соответственно. Алгоритмы оптимизации DiffGrad, Yogi и AdaBelief, в отличие от SGDM Nesterov и Adam, смогли достичь глобального минимума функции Растригина и за меньшее количество итераций минимизировать функцию Розенброка. Сверточные нейронные сети, использующие данные методы оптимизации, смогли достичь большей точности распознавания изображений CIFAR-10 и CIFAR-100 [13–15] за меньшее количество эпох по сравнению с известными аналогами. Помимо DiffGrad, Yogi и AdaBelief, существуют модификации, которые используют положительно-отрицательные средние сдвиги. Данные модификации называются PNM и AdaPNM [16].

В данной работе мы предлагаем модифицированные алгоритмы адаптивной невыпуклой и доверительной оптимизации на основе положительно-отрицательных оценок моментов, которые позволяют более точно сходиться в области глобального минимума. Как и в работе [10], помимо предложенных алгоритмов оптимизации, приводятся оценки их скоростей сходимости, отображающие меры эффективно-

сти систем обучения с подкреплением. Затем, в экспериментальной части, демонстрируются результаты распознавания изображений злокачественных пигментных образований кожи на основе глубоких сверточных нейронных сетей, содержащих предложенные алгоритмы оптимизации.

### 1. Предварительные сведения

Пусть  $f: R^n \rightarrow R$  – бесконечно дифференцируемая на всем множестве определения функция, содержащая один или несколько экстремумов. Задача методов оптимизации состоит в поиске такого  $\theta^*$ , при котором  $f(\theta^*)$  принимает минимально возможное значение.

Наиболее распространенным алгоритмом оптимизации функции ошибок в нейронных сетях является Adam [10], который отличается от SGDM [9] наличием экспоненциальных скользящих средних градиента  $m_i$  и квадрата градиента  $v_i$  с гиперпараметрами  $\beta_1, \beta_2 \in [0, 1)$ . Метод оптимизации Adam представлен следующим образом:

$$\begin{aligned} g_i &= \nabla f(\theta_{i-1}) + \lambda \theta_{i-1}, \\ m_i &= \beta_1 m_{i-1} + (1 - \beta_1) g_i, \\ v_i &= \beta_2 v_{i-1} + (1 - \beta_2) g_i^2, \\ \hat{m}_i &= \frac{m_i}{(1 - \beta_1^i)}, \quad \hat{v}_i = \frac{v_i}{(1 - \beta_2^i)}, \\ \theta_i &= \theta_{i-1} - \frac{\gamma \hat{m}_i}{\sqrt{\hat{v}_i} + \epsilon}, \end{aligned} \tag{1}$$

где  $\nabla$  – оператор набла,  $g_i$  – градиент от целевой функции  $f$ ,  $\theta_i$  – весовое значение,  $\lambda$  – распад веса,  $\gamma$  – скорость обучения,  $\epsilon$  – параметр затухания,  $\hat{m}_i$  и  $\hat{v}_i$  – нормализованные экспоненциальные оценки моментов градиента и квадрата градиента соответственно. Однако данный подход не всегда достигает глобального минимума. Это объясняется неспособностью экспоненциальных средних сдвигов анализировать выпуклость целевой функции обучения, что приводит к сходимости в области локальных минимумов. Вследствие чего был разработан модифицированный алгоритм разностного градиента DiffGrad [13].

Подход разностного градиента основан на методе оценки моментов. При этом, вместо преобразования скорости обучения и весов, производится вычисление коэффициента трения  $\xi_i$ , предназначенного для управления скоростью обучения с использованием информации о краткосрочном поведении градиента и выраженного как:

$$\xi_i = \frac{1}{1 + e^{-|\Delta g_i|}}, \tag{2}$$

где  $\Delta g_i$  – разница между предыдущим и текущим градиентами, заданная формулой

$$\Delta g_i = g_i - g_{i-1}. \tag{3}$$

Затем по итеративной формуле

$$\theta_i = \theta_{i-1} - \frac{\gamma \hat{m}_i}{\sqrt{\hat{v}_i + \epsilon}} \quad (4)$$

происходит минимизация функции потерь.

Преимущество DiffGrad заключается в эффективном изменении краткосрочных градиентов. Помимо алгоритма разностного градиента, существуют модификации, преобразовывающие скользящие средние квадрата градиента, такие как Yogi [14] и AdaBelief [15].

Алгоритм оптимизации Yogi основан на уменьшении градиентов на квадратный корень из экспоненциальных скользящих средних из предыдущих квадратов градиентов и контролирует увеличение скорости обучения [10]. Это позволяет решить проблему сбоя сходимости в простых настройках выпуклой оптимизации, с которыми Adam не справляется. Данный подход имеет следующий вид:

$$v_i = \beta_2 v_{i-1} + (1 - \beta_2) \text{sign}(v_{i-1} - g_i^2) g_i^2, \quad (5)$$

$$\theta_i = \theta_{i-1} - \frac{\gamma \hat{m}_i}{\sqrt{\hat{v}_i + \epsilon}}.$$

Алгоритм Yogi показывает относительно лучшие результаты в сверточных нейронных сетях по сравнению с DiffGrad и другими известными методами оптимизации. Однако в работе [15] был предложен метод AdaBelief, который превосходил Yogi по результатам распознавания визуальных данных.

Главной особенностью AdaBelief является адаптация скорости обучения в соответствии с «доверием» в текущее направление градиента. Согласно  $s_i$ , как предсказанию градиента на следующем временном шаге, если наблюдаемый градиент сильно отклоняется от предсказания, то текущему наблюдению не доверяют и делают небольшой шаг; если наблюдаемый градиент близок к предсказанию, значит, ему доверяют и делают большой шаг. Подход в минимизации функции потерь описан следующим образом:

$$s_i = \beta_2 s_{i-1} + (1 - \beta_2)(g_i - m_i)^2, \quad (6)$$

$$\hat{s}_i = \frac{s_i}{1 - \beta_2},$$

$$\theta_i = \theta_{i-1} - \frac{\gamma \hat{m}_i}{\sqrt{\hat{s}_i + \epsilon}}.$$

В работе [16] был предложен метод положительно-отрицательных оценок моментов. Алгоритмы, использующие данный метод, называются обычной (PNM) и адаптивной (AdaPNM) положительно-отрицательной оценкой моментов.

Алгоритм PNM содержит положительно-отрицательную оценку момента градиента  $m_i$  и дополнительный параметр регулирования скорости обучения  $\beta_0$ . Данную модификацию можно описать следующим образом:

$$m_i = \beta_1^2 m_{i-1} + (1 - \beta_1^2) g_i, \quad (7)$$

$$\theta_i = \theta_{i-1} - \frac{\gamma}{\sqrt{(1 + \beta_0^2) + \beta_0^2}} [(1 + \beta_0) m_i - \beta_0 m_{i-1}].$$

Алгоритм AdaPNM является адаптивной версией PNM, которая содержит оценку момента квадрата градиента, имеющую максимальное значение с каждой итерацией. Описывается данный подход следующим образом:

$$m_i = \beta_1^2 m_{i-1} + (1 - \beta_1^2) g_i, \quad (8)$$

$$\hat{m}_i = \frac{(1 + \beta_0) m_i - \beta_0 m_{i-1}}{(1 - \beta_1^2)},$$

$$v_i = \beta_2 v_{i-1} + (1 - \beta_2) g_i^2,$$

$$v_{max} = \max(v_i, v_{max}),$$

$$\hat{v}_i = \frac{v_{max}}{(1 - \beta_2^i)},$$

$$\theta_i = \theta_{i-1} - \frac{\gamma \hat{m}_i}{\sqrt{(1 + \beta_0^2) + \beta_0^2} (\sqrt{\hat{v}_i} + \epsilon)}.$$

Все перечисленные методы являются расширенными версиями стандартных алгоритмов оптимизации SGD и Adam. Наибольшую точность прогнозирования у интеллектуальных систем возможно получить при использовании известных методов оптимизации Yogi и AdaBelief. Поэтому данные подходы были выбраны для дальнейшей модификации при помощи положительно-отрицательной оценки момента.

## 2. Алгоритмы оптимизации, основанные на положительно-отрицательной оценке моментов

Основной идеей для разработки адаптивных алгоритмов невыпуклой и доверительной оптимизаций служит замена экспоненциальных оценок моментов положительно-отрицательными в Yogi и AdaBelief. Это позволит увеличить скорость сходимости в области глобального минимума, что закономерно увеличивает точность распознавания, особенно для мультимодальных нейронных сетей. Предложенный метод адаптивной невыпуклой оптимизации описан в виде псевдокода в алгоритме 1.

**Алгоритм 1.** Невыпуклая адаптивная оптимизация на основе положительно-отрицательных оценок моментов

**Input:**  $\gamma$  (скорость обучения),  $\beta_0, \beta_1, \beta_2$  (коэффициенты для вычисления  $m_i$  и  $v_i$ ),  $\theta_0$  (входные данные),  $f$  (функция потерь),  $\lambda$  (распад веса),  $\epsilon$  (параметр затухания),  $n$  (номер эпохи)

**Output:**  $\theta_n$  (выходные данные)

1.  $m_0 \leftarrow 0$  (первый момент),  $v_0 \leftarrow 0$  (второй момент)
2. **for**  $i$  from 1 to  $n$  **do**

3.  $g_i \leftarrow \nabla f(\theta_{i-1}) + \lambda \theta_{i-1}$
4.  $m_i \leftarrow \beta_1^2 m_{i-1} + (1 - \beta_1^2) \text{sign}(v_{i-1} - g_i^2) g_i$
5.  $\hat{m}_i \leftarrow \frac{(1 + \beta_0) m_i - \beta_0 m_{i-1}}{1 - \beta_1}$  //нормализованный  $m_i$
6.  $v_i \leftarrow \beta_2 v_{i-1} + (1 - \beta_2) \text{sign}(v_{i-1} - g_i^2) g_i^2$
7.  $v_{max} \leftarrow \max(v_i, v_{max})$
8.  $\hat{v}_i \leftarrow \frac{v_{max}}{1 - \beta_2}$  //нормализованный  $v_i$
9.  $\theta_i \leftarrow \theta_{i-1} - \frac{\gamma \hat{m}_i}{\sqrt{(1 + \beta_0^2) + \beta_0^2 (\sqrt{\hat{v}_i} + \epsilon)}}$
10. **end for**

Данный подход содержит те же преимущества, что и метод оптимизации Yogi, только с учетом положительно-отрицательных скользящих средних, способных ускорить процесс сходимости в области глобального минимума, при этом увеличивая точность распознавания.

Аналогично алгоритму 1, предлагается способ адаптивной доверительной оптимизации при помощи положительно-отрицательных оценок моментов. Предложенный подход к оптимизации описан в алгоритме 2.

**Алгоритм 2.** Доверительная адаптивная оптимизация на основе положительно-отрицательных оценок моментов

**Input:**  $\gamma$  (скорость обучения),  $\beta_0, \beta_1, \beta_2$  (коэффициенты для вычисления  $m_i$  и  $s_i$ ),  $\theta_0$  (входные данные),  $f$  (функция потерь),  $\lambda$  (распад веса),  $\epsilon$  (параметр затухания),  $n$  (номер эпохи)

**Output:**  $\theta_n$  (выходные данные)

1.  $m_0 \leftarrow 0$  (первый момент),  $s_0 \leftarrow 0$  (второй момент)
2. **for**  $i$  from 1 to  $n$  **do**
3.  $g_i \leftarrow \nabla f(\theta_{i-1}) + \lambda \theta_{i-1}$
4.  $m_i \leftarrow \beta_1^2 m_{i-1} + (1 - \beta_1^2) g_i$
5.  $\hat{m}_i \leftarrow \frac{(1 + \beta_0) m_i - \beta_0 m_{i-1}}{1 - \beta_1}$  //нормализованный  $m_i$
6.  $s_i \leftarrow \beta_2 s_{i-1} + (1 - \beta_2) (g_i - m_i)^2 g_i^2$
7.  $s_{max} \leftarrow \max(s_i, s_{max})$
8.  $\hat{s}_i \leftarrow \frac{s_{max}}{1 - \beta_2}$  //нормализованный  $s_i$
9.  $\theta_i \leftarrow \theta_{i-1} - \frac{\gamma \hat{m}_i}{\sqrt{(1 + \beta_0^2) + \beta_0^2 (\sqrt{\hat{s}_i} + \epsilon)}}$
10. **end for**

Доверительное регулирование скорости обучения, подкрепленное дополнительным параметром  $\beta_0$  и положительно-отрицательной оценкой моментов, явля-

ется основным преимуществом алгоритма 2 над известными аналогами, особенно над AdaBelief.

Теоретический анализ предложенных алгоритмов строится на оценивании скорости сходимости  $R(T)$ , определенной следующим образом:

$$R(T) = \sum_{t=1}^T [f_t(\theta_t) - f_t(\theta^*)], \quad (9)$$

где  $\theta_t$  и  $\theta^*$  – полученные и верные промежуточные данные соответственно.

Скорости сходимости предложенных алгоритмов адаптивной невыпуклой и доверительной оптимизации оценены в теоремах 1 и 2.

**Теорема 1.** Пусть  $f_t$  – функция потерь нейронной сети,  $g_{t,\theta}$  – градиент, удовлетворяющий условиям для различных положительных чисел  $G, G_\infty, D, D_\infty$ :

$$\|g_{t,\theta}\|_2 = \sqrt{\int_{-\infty}^{+\infty} (g_{t,\theta})^2 d\theta_t} \leq G,$$

$$\|g_{t,\theta}\|_\infty = \sup_{\theta_t \in (-\infty, \infty)} g_{t,\theta} \leq G_\infty,$$

$$\|\theta_n - \theta_m\|_2 \leq D \text{ и } \|\theta_n - \theta_m\|_\infty \leq D_\infty$$

для всех  $m, n \in \{1, \dots, T\}$  и для любых  $\theta \in R^d$ . Пусть  $\eta = \beta_1^2 / \sqrt{\beta_2}$  и  $\beta_0, \beta_1, \beta_2 \in [0, 1)$  удовлетворяют условиям  $\beta_1^2 / \sqrt{\beta_2} < 1$ ,  $\alpha_t = \alpha / \sqrt{t}$ ,  $\beta_1^2 + \beta_1^2 \beta_0 - \beta_0 = \beta'$  и  $\beta_{1,t} = \beta_1 \lambda^{t-1}$ ,  $\lambda \in (0, 1)$ . Тогда для любых  $T \geq 1$  предложенный алгоритм 1 гарантирует следующую оценку скорости сходимости:

$$R(T) \leq \frac{D^2}{2\alpha(1-\beta_0)(1-\beta_1)^2} \sum_{i=1}^d \sqrt{T v_{T,i}} + \frac{\alpha(1+\beta') G_\infty^3}{(1-\beta_0)(1-\beta_1)^2 \sqrt{1-\beta_2} (1-\eta)^2 G^2} \sum_{t=1}^d \|g_{1:t,i}\|_2 + \sum_{t=1}^d \frac{D_\infty^2 G_\infty \sqrt{1-\beta_2}}{2\alpha(1-\beta_0)(1-\beta_1^2)(1-\lambda)^2}. \quad (9)$$

**Доказательство:** Используя Лемму 10.2 из [10] для функции  $f_t$ , можно утверждать, что

$$f_t(\theta_t) - f_t(\theta^*) \leq g_t^T(\theta_t - \theta^*) = \sum_{i=1}^d g_{t,i}(\theta_{t,i} - \theta_{t,i}^*). \quad (10)$$

В неравенстве (10) необходимо найти  $g_{t,i}(\theta_{t,i} - \theta_{t,i}^*)$  для  $i = 1, \dots, d$ . Обновление весовых коэффициентов из алгоритма 1 имеет следующий вид:

$$\begin{aligned} \theta_{t+1} &= \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t}} = \\ &= \theta_t - \frac{\alpha}{1-\beta_1^i} \left( \frac{\beta_{1,t}^2 + \beta_{1,t}^2 \beta_{0,t} - \beta_{0,t}}{\sqrt{\hat{v}_t}} m_{t-1} + \right. \\ &\quad \left. + \frac{(1-\beta_{0,t})(1-\beta_{1,t}^2)}{\sqrt{\hat{v}_t}} g_t \right). \end{aligned} \quad (11)$$

Пусть

$$\beta_{1,t}^2 + \beta_{1,t}^2 \beta_{0,t} - \beta_{0,t} = \beta_t' \text{ и } \alpha = \gamma / \sqrt{(1 + \beta_0^2) + \beta_0^2}.$$

В обеих частях равенства (11) вычитаем  $\theta_i^*$  и | получаем:

$$(\theta_{t+1,i} - \theta_i^*)^2 = (\theta_{t,i} - \theta_i^*)^2 + \alpha^2 \frac{(\widehat{m}_{t,i})^2}{\sqrt{\widehat{v}_{t,i}}} - \frac{2\alpha}{1-\beta_1'} \left( \frac{\beta_t'}{\sqrt{\widehat{v}_t}} m_{t-1,i} + \frac{(1-\beta_{0,t})(1-\beta_{1,t}^2)}{\sqrt{\widehat{v}_t}} g_{t,i} \right) (\theta_{t,i} - \theta_i^*). \tag{12}$$

Из данного равенства находим  $g_{t,i}(\theta_{t,i} - \theta_i^*)$ :

$$g_{t,i}(\theta_{t,i} - \theta_i^*) = \frac{(1-\beta_1')\sqrt{\widehat{v}_{t,i}}((\theta_{t,i} - \theta_i^*)^2 - (\theta_{t+1,i} - \theta_i^*)^2)}{2\alpha_t(1-\beta_{0,t})(1-\beta_{1,t}^2)} + \frac{\beta_t'}{(1-\beta_{0,t})(1-\beta_{1,t}^2)} \sqrt{\frac{\widehat{v}_{t-1,i}}{\alpha_{t-1}}} \frac{(\theta_{t,i} - \theta_i^*) m_{t-1,i} \sqrt{\alpha_{t-1}}}{\sqrt[4]{\widehat{v}_{t-1,i}}} + \frac{\alpha_t(1-\beta_1')}{2(1-\beta_{0,t})(1-\beta_{1,t}^2)} \frac{(\widehat{m}_{t,i})^2}{\sqrt{\widehat{v}_{t,i}}} + \frac{((\theta_{t,i} - \theta_i^*)^2 - (\theta_{t+1,i} - \theta_i^*)^2)\sqrt{\widehat{v}_{t,i}}}{2\alpha_t(1-\beta_{0,t})(1-\beta_{1,t}^2)} + \frac{\beta_t'(\theta_{t,i} - \theta_i^*)^2\sqrt{\widehat{v}_{t-1,i}}}{(1-\beta_{0,t})(1-\beta_{1,t}^2)} + \frac{\beta_t'}{(1-\beta_{0,t})(1-\beta_{1,t}^2)} \frac{(\theta_{t,i} - \theta_i^*) m_{t-1,i}^2 \alpha_{t-1}}{\sqrt{\widehat{v}_{t-1,i}}} + \frac{\alpha_t}{2(1-\beta_{0,t})(1-\beta_{1,t}^2)} \frac{(\widehat{m}_{t,i})^2}{\sqrt{\widehat{v}_{t,i}}}. \tag{13}$$

Подставляя (13) в (10), получаем следующее неравенство:

$$R(T) \leq \sum_{i=1}^d \frac{(\theta_{1,i} - \theta_i^*)^2 \sqrt{\widehat{v}_{1,i}}}{2\alpha_1(1-\beta_0)(1-\beta_1^2)} + \sum_{i=1}^d \sum_{t=2}^T \frac{(\theta_{t,i} - \theta_i^*)^2}{2(1-\beta_0)(1-\beta_1^2)} \left( \frac{\sqrt{\widehat{v}_{t,i}}}{\alpha_t} - \frac{\sqrt{\widehat{v}_{t-1,i}}}{\alpha_{t-1}} \right) + \frac{\alpha_t G_\infty (1+\beta_1')}{(1-\beta_0)(1-\beta_1^2)\sqrt{1-\beta_2}(1-\eta)^2} \sum_{i=1}^d g_{1:T,i,2} + \sum_{i=1}^d \sum_{t=1}^T \frac{\beta_t'(\theta_{t,i} - \theta_i^*)^2 \sqrt{\widehat{v}_{t,i}}}{2\alpha_t(1-\beta_0)(1-\beta_1^2)}. \tag{14}$$

Используя  $\alpha_t = \alpha / \sqrt{t}$ ,  $\|\theta_n - \theta_m\|_2 \leq D$  и  $\|\theta_n - \theta_m\|_\infty \leq D_\infty$ , выводим неравенство (14), как

$$R(T) \leq \frac{D^2 \sqrt{T}}{2\alpha(1-\beta_0)(1-\beta_1^2)} \sum_{i=1}^d \sqrt{\widehat{v}_{T,i}} + \frac{\alpha(1+\beta_1')G_\infty^3}{(1-\beta_0)(1-\beta_1^2)\sqrt{1-\beta_2}(1-\eta)^2 G^2} \sum_{i=1}^d \|g_{1:T,i}\|_2 + \sum_{i=1}^d \sum_{t=1}^T \frac{\beta_t D_\infty^2 G_\infty \sqrt{1-\beta_2}}{2\alpha(1-\beta_0)(1-\beta_1^2)} \sqrt{t}. \tag{15}$$

Пользуясь неравенством  $\sum_{t=1}^T \frac{\beta_1}{(1-\beta_0)(1-\beta_1^2)} \sqrt{t} \leq \frac{1}{(1-\beta_0)(1-\beta_1^2)(1-\lambda)^2}$  из [10] в (15), границы сходимости оцениваются следующим образом:

$$R(T) \leq \frac{D^2 \sqrt{T}}{2\alpha(1-\beta_0)(1-\beta_1^2)} \sum_{i=1}^d \sqrt{\widehat{v}_{T,i}} + \frac{\alpha(1+\beta_1')G_\infty^3}{(1-\beta_0)(1-\beta_1^2)\sqrt{1-\beta_2}(1-\eta)^2 G^2} \sum_{i=1}^d \|g_{1:T,i}\|_2 + \sum_{i=1}^d \sum_{t=1}^T \frac{D_\infty^2 \beta_1}{2\alpha(1-\beta_0)(1-\beta_1^2)} \sqrt{t v_{t,i}} \leq \frac{D^2 \sqrt{T}}{2\alpha(1-\beta_0)(1-\beta_1^2)} \sum_{i=1}^d \sqrt{\widehat{v}_{T,i}} + \frac{\alpha(1+\beta_1')G_\infty^3}{(1-\beta_0)(1-\beta_1^2)\sqrt{1-\beta_2}(1-\eta)^2 G^2} \sum_{i=1}^d \|g_{1:T,i}\|_2 + \sum_{i=1}^d \sum_{t=1}^T \frac{\beta_1}{(1-\beta_0)(1-\beta_1^2)} \sqrt{t} \leq \frac{1}{(1-\beta_0)(1-\beta_1^2)(1-\lambda)^2}.$$

Теорема доказана.

**Теорема 2.** При таких же условиях, как и в теореме 1, для предложенного алгоритма 2 справедлива следующая оценка скорости сходимости:

$$R(T) \leq \frac{D^2 \sqrt{T}}{2\alpha(1-\beta_0)(1-\beta_1^2)} \sum_{i=1}^d \sqrt{\widehat{s}_{T,i}} + \frac{\alpha(1+\beta_1')\sqrt{1+\log T}}{2\sqrt{c}(1-\beta_0)^3(1-\beta_1^2)^3} \sum_{i=1}^d g_{1:T,i,2} + \sum_{i=1}^d \frac{D_\infty^2 G_\infty \sqrt{1-\beta_2}}{2\alpha(1-\beta_0)(1-\beta_1^2)(1-\lambda)^2}. \tag{16}$$

**Доказательство:** Подставляя  $\widehat{s}_{t,i}$  вместо  $\widehat{v}_{t,i}$  в неравенстве (14), получаем

$$\begin{aligned}
 R(T) \leq & \sum_{i=1}^d \frac{(\theta_{1,i} - \theta_i^*)^2 \sqrt{\widehat{s}_{1,i}}}{2\alpha_1(1-\beta_0)(1-\beta_1^2)} + \\
 & + \sum_{i=1}^d \sum_{t=2}^T \frac{(\theta_{t,i} - \theta_i^*)^2}{2(1-\beta_0)(1-\beta_1^2)} \left( \frac{\sqrt{\widehat{s}_{t,i}}}{\alpha_t} - \frac{\sqrt{\widehat{s}_{t-1,i}}}{\alpha_{t-1}} \right) + \\
 & + \frac{\alpha G_\infty (1+\beta'_t)}{(1-\beta_0)(1-\beta_1^2)\sqrt{1-\beta_2}(1-\eta)^2} \sum_{i=1}^d g_{1,T,i2} + \\
 & + \sum_{i=1}^d \sum_{t=1}^T \frac{\beta'(\theta_{t,i} - \theta_i^*)^2 \sqrt{\widehat{s}_{t,i}}}{2\alpha_t(1-\beta_0)(1-\beta_1^2)}.
 \end{aligned} \tag{17}$$

Из [17] применим к (17) следующее неравенство:

$$\begin{aligned}
 \sum_{i=1}^d \sum_{t=1}^T \frac{\alpha_t (\widehat{m}_{t,i})^2}{\sqrt{\widehat{s}_{t,i}}} & \leq \\
 & \leq \frac{\alpha \sqrt{1+\log T}}{2\sqrt{c}(1-\beta_0)^3(1-\beta_1^2)^3} \sum_{i=1}^d \|g_{1:T,i}\|_2.
 \end{aligned} \tag{18}$$

Таким образом, получаем оценку скорости сходимости Алгоритма 2:

$$\begin{aligned}
 R(T) \leq & \frac{D^2 \sqrt{T}}{2\alpha(1-\beta_0)(1-\beta_1^2)} \sum_{i=1}^d \sqrt{\widehat{s}_{T,i}} + \\
 & + \frac{\alpha(1+\beta')\sqrt{1+\log T}}{2\sqrt{c}(1-\beta_0)^3(1-\beta_1^2)^3} \sum_{i=1}^d \|g_{1:T,i}\|_2 + \\
 & + \sum_{i=1}^d \frac{D_\infty^2 G_\infty \sqrt{1-\beta_2}}{2\alpha(1-\beta_0)(1-\beta_1^2)(1-\lambda)^2}.
 \end{aligned} \tag{19}$$

Теорема доказана.

Из формул (9) и (16) видно, что предложенные алгоритмы 1 и 2 являются  $O(\sqrt{T})$ -ограниченными, то есть их скорости сходимости являются эффективными для обучения с подкреплением.

Предложенная модификация для алгоритмов оптимизации Yogi и AdaBelief является более эффективной по сравнению с известными аналогами. Это объясняется наличием положительно-отрицательных оценок моментов градиента и квадрата градиента, что позволяет достигать глобального минимума за меньшее число итераций, увеличивая при этом точность распознавания. С помощью дополнительного нулевого момента  $\beta_0$  предложенные модификации регулируют скорость обучения в процессе минимизации функции потерь нейронной сети.

### 3. Мультимодальная нейросетевая система анализа гетерогенных дерматологических данных

Наиболее распространенными типами данных в дерматологических базах являются визуальные мно-

гомерные данные и статистические данные о пациентах. К визуальным многомерным данным относятся изображения пигментных поражений кожи. К статистическим данным относятся пол, возраст, а также локализация пигментного новообразования на теле пациента. Визуальные клинические данные являются основной формой диагностирования дерматологических онкопатологий [17]. Однако статистические параметры пациентов также могут указывать на риск развития злокачественных форм пигментных новообразований. Поэтому возникает необходимость во всестороннем анализе разнородных данных для более точной диагностики [18].

В настоящее время мультимодальное машинное обучение является перспективной областью исследований, в которой разрабатываются нейросетевые модели для анализа информации разных модальностей [19]. При слиянии разнородной информации учитывается представление признаков различных модальностей для более точного и полного анализа данных [20]. Неочевидная взаимосвязь между обрабатываемыми данными и результатами диагностики извлекается за счет дополнительного исследования информации между модальностями. Таким образом нейронные сети способны использовать дополнительные данные путем интеграции нескольких модальностей в общую структуру [21]. Предлагаемая мультимодальная нейросетевая система состоит из СНС для анализа визуальных данных и линейного многослойного перцептрона для анализа статистических данных. На вход предлагаемой мультимодальной нейросетевой системы поступают предварительно обработанные дерматологические изображения  $I_{RGB}$  и вектор закодированных статистических признаков  $\vec{M}$ . В многослойном перцептроне нейроны выполняют суммирование полученного вектора входных данных  $\vec{M}$  и коэффициента смещения  $b$ , формируя выходной сигнал следующим образом:

$$v_M = f\left(\sum_i^n M_i w_i + b\right), \tag{20}$$

где  $w_i$  – веса нейронов;  $b$  – это коэффициент смещения.

Получение карт признаков дерматологических изображений  $P_{RGB}$  после прохождения сверточных слоев СНС производится параллельно следующим образом:

$$P_f(x, y) = b + \sum_{i=-\frac{w-1}{2}}^{\frac{w-1}{2}} \sum_{j=-\frac{w-1}{2}}^{\frac{w-1}{2}} \sum_{k=0}^{I-1} w_{ijk}^{(l)} P(x+i, y+j, k), \tag{21}$$

где  $P_f$  – это карта признаков дерматологического изображения;  $w_{ijk}^{(l)}$  – это коэффициент фильтра размера  $p \times p$ .

Объединение карты признаков дерматологического изображения  $P_f$  и выходного сигнала многослойно-

го персептрона  $v_M$  производится на слое конкатенации следующим образом:

$$C_q = \sum_i \sum_j \sum_k P_{ijk} w_{ijkq}^{(2)} + \sum_{i=1}^n v_M w_{iq}^{(2)}, \quad (22)$$

где  $w_{ijkq}^{(2)}$  – это весовые коэффициенты для обработки карты признаков дерматологического изображения;  $w_{iq}^{(2)}$  – это весовые коэффициенты для обработки выходного сигнала многослойного персептрона.

Активация последнего слоя мультимодальной нейросетевой системы производится следующим образом:

$$C'(y|x, \theta) = \text{softmax}(x; \theta) = \frac{\exp(w_q^n x_q + g_q^n)}{\sum_{q=1}^Q \exp(w_q^n x_q + g_q^n)}, \quad (23)$$

где  $w_q^n$  – это весовой вектор выходного сигнала, связанный с классом  $q$ . Регулирование значений весовых

коэффициентов линейного многослойного персептрона и сверточных слоев производится при помощи выбранного метода оптимизации для более точной сходимости функции потерь в область глобального минимума.

На рис. 1 представлена предлагаемая мультимодальная нейросетевая система распознавания пигментных образований кожи на основе сверточной архитектуры AlexNet. В результате предварительной обработки метаданных формируется вектор входных признаков, состоящий из 13 переменных. После завершения процесса обработки визуальных данных при помощи сверточных слоев нейросетевой архитектуры формируется выходной слой, состоящий из 10 нейронов, соответствующих количеству распознаваемых классов. Аналогичным образом формируется выходной слой из линейной нейросетевой архитектуры, состоящий также из 10 нейронов, соответствующих количеству распознаваемых классов.

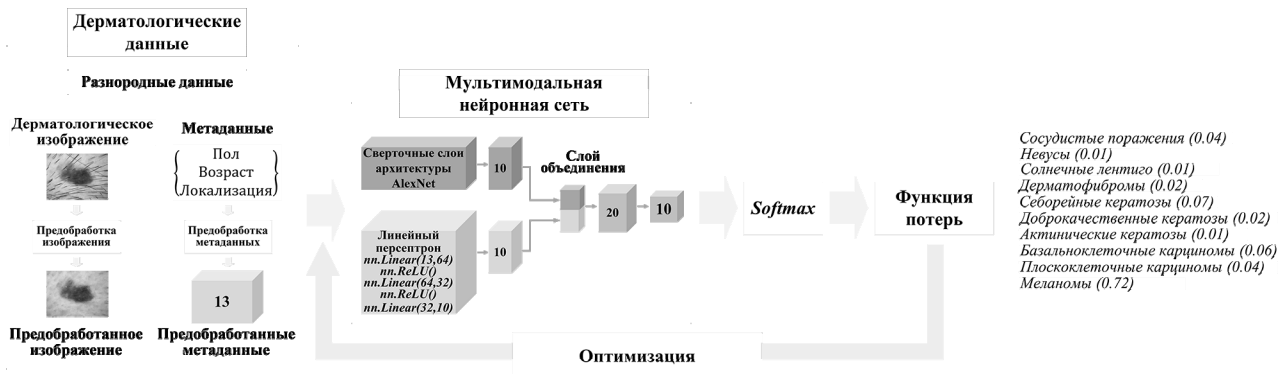


Рис. 1. Схема обучения мультимодальной нейросетевой системы распознавания пигментных образований кожи на основе сверточной архитектуры AlexNet

#### 4. Моделирование мультимодальной нейросетевой системы с использованием различных методов оптимизации

Для проведения моделирования мультимодальной нейросетевой системы была отобрана база дерматологических данных из открытого архива International Skin Imaging Collaboration (ISIC). Набор данных включает в себя 41725 дерматологических изображений различного размера и качества. Каждое изображение связано с набором статистических факторов пациента и установленным диагнозом. Все данные распределены по 10 диагностически значимым категориям. Набор статистических факторов для каждого изображения включает в себя информацию о половой принадлежности пациента, возрастной группе с шагом в пять лет и информацию по локализации пигментного образования на теле пациента. На этапе предварительной обработки статистических данных фактор «Возраст» был разделен на четыре группы в соответствии с возрастной классификацией, принятой Всемирной организацией здравоохранения (ВОЗ). Таким образом, вариабельность параметра «Возраст» была снижена с 18 до 4 возможных значений. Моде-

лирование проводилось с использованием языка программирования высокого уровня Python 3.11.0. Все расчеты проводились на ПК с процессором Intel(R) Core(TM) i5-8500 с частотой 3,00 ГГц с 16 ГБ оперативной памяти и 64-битной операционной системой Windows 10. Обучение мультимодальных нейросетевых систем проводилось с использованием графического процессора (GPU) на базе видеочипсета NVIDIA GeForce GTX 1050TI. Для моделирования нейросетевых систем применялся фреймворк машинного обучения Pytorch. Для разработки предложенных методов оптимизации использовались библиотеки Math и Pytorch. Для моделирования мультимодальной нейросетевой системы распознавания пигментных новообразований кожи были выбраны нейросетевые архитектуры AlexNet, GoogLeNet и EfficientNet\_B0, которые были предварительно обучены на наборе естественных изображений ImageNet.

На первом этапе моделирования была произведена предварительная обработка отобранных дерматологических данных. Предварительная обработка статистических данных заключалась в создании входного вектора при помощи метода one hot encoding. Предварительная обработка визуальных данных заключа-

лась в применении метода удаления волосных структур [22]. Второй шаг предварительной обработки визуальных данных заключался в преобразовании размера входных данных до  $227 \times 227$  пикселей для архитектуры AlexNet и до  $224 \times 224$  пикселей для архитектур GoogLeNet и EfficientNet\_B0. Для дальнейшего моделирования база дерматологических данных была разделена на данные для обучения и тестирования в процентном соотношении 80 к 20. К обучающему набору визуальных данных были применены аффинные преобразования для аугментации. Для процесса обучения предварительно обработанные дерматологические изображения пигментных поражений кожи подавались на вход выбранных архитектур СНС. Вектор предварительно обработанных статистических данных из обучающей выборки подавался на вход многослойной нейросетевой архитектуры, состоящей из трех линейных слоев и слоев активации ReLu. После прохождения разномодальных сигналов через СНС и линейный персептрон выходные векторы признаков объединялись на слое конкатенации. Выходной сигнал подавался на слой *softmax* для определения вероятностного соотношения прогнозируемых меток по 10 диагностически значимым категориям. Полученные вероятности сравнивались с истинными метками к обучающим данным, и производилось вычисление значения ошибки при помощи функции кросс-энтропийных потерь. После чего производилась оптимизация весовых коэффициентов нейросетевой архитектуры. В результате происходило постепенное запоминание истинных векторов и минимизация потерь при обучении. В табл. 1–3 представлены результаты оценки при тестировании мультимодальных нейросетевых систем, обученных с использованием различных методов оптимизации. Для тестовой статистической оценки моделей были выбраны такие количественные методы, как специфичность (*Sp*), чувствительность (*Se*), коэффициент корреляции Мэтьюса (*MCC* – Matthews correlation co-

efficient), частота ложноотрицательных результатов (*FNR* – false negative rate), частота ложноположительных результатов (*FPR* – false positive rate). Для оценки скорости сходимости разработанных алгоритмов 1 и 2 в EfficientNet\_B0 при решении задачи распознавания пигментных новообразований кожи использовались три вида функции регрессии  $\alpha x^3 + \beta x^2 + \gamma x + \delta$ ,  $\exp(\alpha + \beta x)$ ,  $\beta + \alpha \sqrt{x}$  в [23, 24]. В табл. 4 приведены оценки, описывающие скорость достижения минимума функции потерь по эмпирическим данным для всех алгоритмов оптимизации за 7 эпох обучения рассматриваемой нейронной сети. В табл. 5 собраны средние относительные ошибки для каждой регрессии, которые показывают соответствие эмпирическим данным функции потерь.

Наибольший показатель точности распознавания дерматологических данных составил 85,76 % и был получен при тестировании мультимодальной нейросетевой системы на основе архитектуры AlexNet, которая была обучена с применением предложенного алгоритма 1. Наилучшие показатели специфичности (*Sp*), чувствительности (*Se*) и *MCC* были получены при тестировании мультимодальной нейросетевой системы на основе архитектуры AlexNet с использованием предложенного алгоритма 1 и составили 0,9841, 0,8576 и 0,6692 соответственно. По всем рассмотренным методам статистической тестовой оценки у систем, обученных с применением предложенного алгоритма 1, результат выше, чем у мультимодальных систем распознавания пигментных новообразований кожи, обученных с использованием известных методов оптимизации. При тестировании всех мультимодальных нейросетевых систем, обученных с применением предложенных алгоритмов 1 и 2, значения всех статистических метрик были лучше по сравнению со значениями мультимодальных нейросетевых систем, обученных с использованием известных методов оптимизации Yogi и AdaBelief.

Табл. 1. Результаты статистической оценки при тестировании мультимодальной нейросетевой системы на основе архитектуры AlexNet, обученной с использованием различных методов оптимизации

Метод	Оптимизатор	Точность, %	Значение функции ошибки	<i>MCC</i>	<i>Sp</i>	<i>Se</i>	<i>FNR</i>	<i>FPR</i>
Предложенный	Алгоритм 1	85,76	0,4230	0,7252	0,9841	0,8576	0,1423	0,0158
	Алгоритм 2	85,67	0,4205	0,7217	0,9840	0,8566	0,1433	0,0159
Известный	DiffGrad	85,57	0,4129	0,7190	0,9839	0,8557	0,1442	0,0160
	Yogi	85,56	0,4144	0,7219	0,9839	0,8556	0,1443	0,0161
	AdaPNM	85,28	0,4396	0,7125	0,9836	0,8528	0,1471	0,0163
	AdaBound	85,10	0,4065	0,7121	0,9834	0,8510	0,1489	0,0165
	SGD	84,25	0,3993	0,7002	0,9825	0,8425	0,1574	0,0174
	Adam	84,12	0,4321	0,6864	0,9823	0,8412	0,1587	0,0176
	PNM	83,48	0,4476	0,6804	0,9816	0,8347	0,1652	0,0183
	AdaBelief	83,43	0,4301	0,6788	0,9815	0,8342	0,1657	0,0184

По сравнению с известными аналогами предложенные алгоритмы 1 и 2 за 7 эпох обучения модели нейронной сети достигают области минимума функции потерь с наибольшей скоростью и наименьшими

средними относительными ошибками регрессий. В случае кубической регрессии оценки разработанных алгоритмов 1 и 2 содержат наименьшие старшие коэффициенты 0,0024 и 0,0041 соответственно. При



этом средние относительные ошибки составляют 3,5915 для алгоритма 1 и 3,5834 для алгоритма 2. Из оценки алгоритмов оптимизации с помощью экспоненциальной регрессии следует, что наибольшей скоростью сходимости обладают предложенные алгоритмы 1 и 2, так как соответствующие коэффициенты  $\alpha$  принимают наименьшие значения, равные  $-0,2704$  и  $-0,2701$ . Средняя относительная ошибка для пред-

ложенных оптимизаторов также является наименьшей, так как средние относительные ошибки принимают наименьшие значения 4,3977 и 4,3380 соответственно. Сравнение регрессий  $\beta + \alpha\sqrt{x}$  показывает, что наименьшие значения коэффициента  $\alpha$  составили  $-0,7199$  для алгоритма 1 и  $-0,7012$  для алгоритма 2. Соответствующие средние относительные ошибки принимают наименьшие значения 4,5408 и 4,7658.

Табл. 2. Результаты статистической оценки при тестировании мультимодальной нейросетевой системы на основе архитектуры GoogLeNet, обученной с использованием различных методов оптимизации

Метод	Оптимизатор	Точность, %	Значение функции ошибки	MCC	Sp	Se	FNR	FPR
Предложенный	Алгоритм 1	82,65	0,4458	0,6692	0,9807	0,8265	0,1735	0,0192
	Алгоритм 2	82,54	0,4505	0,6619	0,9806	0,8254	0,1745	0,0194
Известный	AdaPNM	82,64	0,4525	0,6575	0,9807	0,8264	0,1736	0,0193
	DiffGrad	82,62	0,4659	0,6624	0,9806	0,8262	0,1737	0,0193
	Yogi	82,43	0,4683	0,6559	0,9804	0,8243	0,1756	0,0195
	AdaBelief	81,87	0,4772	0,6402	0,9798	0,8187	0,1813	0,0201
	PNM	81,26	0,5320	0,6301	0,9792	0,8126	0,1874	0,0208
	SGD	81,22	0,5062	0,6366	0,9791	0,8122	0,1877	0,0208
	Adam	81,15	0,5066	0,6297	0,9790	0,8115	0,1884	0,0209
	AdaBound	78,45	1,0355	0,5890	0,9760	0,7845	0,2154	0,0239

Табл. 3. Результаты статистической оценки при тестировании мультимодальной нейросетевой системы на основе архитектуры EfficientNet\_B0, обученной с использованием различных методов оптимизации

Метод	Оптимизатор	Точность, %	Значение функции ошибки	MCC	Sp	Se	FNR	FPR
Предложенный	Алгоритм 1	83,03	0,4640	0,6704	0,9813	0,8303	0,1692	0,0187
	Алгоритм 2	82,66	0,5139	0,6619	0,9807	0,8266	0,1733	0,0192
Известный	DiffGrad	82,79	0,4675	0,6651	0,9809	0,8279	0,1721	0,0191
	Yogi	82,48	0,4858	0,6645	0,9805	0,8248	0,1751	0,0194
	AdaBelief	82,06	0,4520	0,6546	0,9801	0,8206	0,1794	0,0199
	Adam	81,32	0,5339	0,6429	0,9792	0,8132	0,1868	0,0207
	AdaPNM	81,29	0,5859	0,6418	0,9792	0,8129	0,1870	0,0207
	SGD	80,79	0,5842	0,6341	0,9787	0,8079	0,1921	0,0213
	AdaBound	79,87	0,5770	0,6231	0,9776	0,7987	0,2013	0,0223
	PNM	77,34	1,0624	0,5826	0,9748	0,7734	0,2266	0,0251

Табл. 4. Оценки скорости сходимости функции потерь для архитектуры EfficientNet\_B0 при помощи регрессий

Метод	Оптимизатор	Вид функции регрессии		
		$\alpha x^3 + \beta x^2 + \gamma x + \delta$	$\exp(\alpha + \beta x)$	$\beta + \alpha\sqrt{x}$
Предложенный	Алгоритм 1	$0,0024x^3 + 0,0373x^2 - 0,6558x + 2,3714$	$\exp(0,7581 - 0,2704x)$	$2,3804 - 0,7199\sqrt{x}$
	Алгоритм 2	$0,0041x^3 + 0,0153x^2 - 0,5858x + 2,3714$	$\exp(0,7891 - 0,2701x)$	$2,3784 - 0,7012\sqrt{x}$
Известный	DiffGrad	$0,0049x^3 + 0,0128x^2 - 0,6052x + 2,3714$	$\exp(0,8195 - 0,2681x)$	$2,3840 - 0,7010\sqrt{x}$
	Yogi	$0,0067x^3 + 0,0125x^2 - 0,5190x + 2,3714$	$\exp(0,8131 - 0,2694x)$	$2,3795 - 0,7002\sqrt{x}$
	AdaBelief	$0,0075x^3 + 0,0229x^2 - 0,4839x + 2,3714$	$\exp(0,8322 - 0,2768x)$	$2,3744 - 0,6984\sqrt{x}$
	Adam	$0,0097x^3 + 0,0547x^2 - 0,3631x + 2,3714$	$\exp(0,8907 - 0,2580x)$	$2,3850 - 0,6638\sqrt{x}$
	AdaPNM	$0,0072x^3 + 0,0206x^2 - 0,4874x + 2,3714$	$\exp(0,8344 - 0,2770x)$	$2,3817 - 0,6954\sqrt{x}$
	SGD	$0,0096x^3 + 0,0561x^2 - 0,3484x + 2,3714$	$\exp(0,8950 - 0,2585x)$	$2,3843 - 0,6759\sqrt{x}$
	AdaBound	$0,0113x^3 + 0,0568x^2 - 0,3473x + 2,3714$	$\exp(0,7523 - 0,1503x)$	$2,4915 - 0,6163\sqrt{x}$
	PNM	$0,0103x^3 + 0,0553x^2 - 0,3481x + 2,3714$	$\exp(0,7627 - 0,2047x)$	$2,4997 - 0,6084\sqrt{x}$

Полученные оценки в табл. 4 и 5 для разработанных алгоритмов указывают, что данные оптимизаторы имеют высокую скорость сходимости к минимуму функции потерь с наименьшими значениями средней

относительной ошибки к соответствующим эмпирическим данным. Использование алгоритмов оптимизации с высокой скоростью сходимости для задач распознавания пигментных новообразований кожи при помощи

систем искусственного интеллекта позволяет достичь наибольшего значения точности распознавания за меньшее количество эпох обучения. Для мультимодальной нейросетевой системы на основе архитектуры AlexNet внедрение предложенных методов оптимизации в процессе обучения позволило повысить точность на 2,33 процентных пункта для алгоритма 1 и на 2,24 процентных пункта для алгоритма 2 по сравнению с наименее эффективным по точности известным методом AdaBelief. Применение разработанных методов оптимизации при обучении мультимодальной

нейросетевой системы на основе архитектуры GoogLeNet позволило повысить точность на 4,20 процентных пункта для алгоритма 1 и на 4,19 процентных пункта для алгоритма 2 по сравнению с наименее эффективным по точности известным методом AdaBound. Для системы на основе архитектуры EfficientNet\_B0 использование предложенных методов позволило повысить точность на 5,69 процентных пункта для алгоритма 1 и на 5,32 процентных пункта для алгоритма 2 по сравнению с наименее эффективным по точности известным методом PNM.

Табл. 5. Средняя относительная ошибка регрессий для архитектуры EfficientNet\_B0

Метод	Оптимизатор	Вид функции регрессии		
		$\alpha x^3 + \beta x^2 + \gamma x + \delta$	$\exp(\alpha + \beta x)$	$\beta + \alpha \sqrt{x}$
Предложенный	Алгоритм 1	3,5915	4,3977	4,5408
	Алгоритм 2	3,5834	4,3380	4,7658
Известный	DiffGrad	3,7194	4,4012	4,8573
	Yogi	3,7203	4,3466	4,8122
	AdaBelief	3,8830	4,6075	4,6214
	Adam	4,2846	4,9101	5,1225
	AdaPNM	3,8024	4,5703	4,5977
	SGD	4,2883	5,0262	5,2713
	AdaBound	5,7163	6,0826	6,1988
	PNM	5,3603	5,9523	6,1110

### Заключение

Использование предложенного алгоритма невыпуклой адаптивной оптимизации на основе положительно-отрицательных оценок моментов позволяет повысить точность интеллектуального распознавания пигментных образований кожи по сравнению с основными известными методами оптимизации. Мультимодальная система анализа гетерогенных дерматологических данных, обученная с применением предложенного алгоритма 1, позволяет на 0,20–0,55 процентных пункта точнее классифицировать пигментные образования кожи по сравнению с нейросетевой системой, обученной с использованием известного метода оптимизации Yogi. Предложенный алгоритм доверительной адаптивной оптимизации 2 на основе положительно-отрицательных оценок моментов позволяет повысить точность интеллектуального распознавания пигментных образований кожи по сравнению с известным методом оптимизации AdaBelief на 0,60–2,24 процентных пункта. Внедрение предложенных алгоритмов оптимизации позволяет повысить точность распознавания пигментных новообразований кожи на 2,33–5,69 процентных пункта по сравнению с наименее эффективными по точности известными методами оптимизации. Применение предложенных методов оптимизации в различных архитектурах интеллектуальных систем позволяет значительно повысить точность прогнозирования и снизить риск возникновения ложноположительных ошибок,

что является ключевым фактором при разработке систем вспомогательной медицинской диагностики.

Основным ограничением для внедрения разработанной мультимодальной интеллектуальной системы, обученной с использованием предложенных алгоритмов оптимизации 1 и 2, является то, что система не может применяться в качестве самостоятельного диагностического инструмента из-за возможного риска возникновения ложноотрицательных ошибок. Медики и специалисты могут использовать предложенную систему только как высокоточный вспомогательный диагностический инструмент для помощи при принятии врачебного решения. Внедрение подобных высокоточных систем автоматизированного анализа пигментных новообразований кожи позволит сократить потребление финансовых и трудовых ресурсов, задействованных в медицинской отрасли, а также повысит шанс раннего выявления пигментных онкопатологий.

В дальнейших исследованиях планируется внедрение предложенных методов оптимизации для разработки нейросетевых систем сегментации пигментных образований кожи, а также использование систем сегментации в качестве этапа предварительной обработки визуальных данных.

### Благодарности

Авторы выражают благодарность СКФУ за помощь в рамках проекта поддержки малых научных групп и отдельных ученых. Исследование в парагра-

фе 2 проведено в Северо-Кавказском центре математических исследований в рамках соглашения с Министерством науки и высшего образования Российской Федерации (соглашение № 075-02-2023-938). Исследование в параграфе 3 проведено при поддержке Российского научного фонда (проект № 23-71-10013).

### References

- [1] Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc Mosby* 2020; 92(4): 807-812. DOI: 10.1016/J.GIE.2020.06.040.
- [2] Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism* 2017; 69: 36-40. DOI: 10.1016/J.METABOL.2017.01.011.
- [3] Brinker TJ, Hekler A, Enk AH, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Cancer* 2019; 113: 47-54. DOI: 10.1016/J.EJCA.2019.04.001.
- [4] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter, SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542: 115-118. DOI: 10.1038/nature21056.
- [5] Haggemüller S, Maron RC, Hekler A, et al. Skin cancer classification via convolutional neural networks: Systematic review of studies involving human experts. *Eur J Cancer* 2021; 156: 202-216. DOI: 10.1016/J.EJCA.2021.06.049.
- [6] Wiens J, Saria S, Sendak M, et al. Author correction: Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine* 2019; 25(9): 1337-1340. DOI: 10.1038/S41591-019-0548-6.
- [7] Hwang J, Bose N, Fan S. AUV adaptive sampling methods: A review. *Appl Sci* 2019; 9: 3145-2019. DOI: 10.3390/APP9153145.
- [8] Hospedales T, Antoniou A, Micaelli P, Storkey A. Meta-learning in neural networks: A survey. *IEEE Trans Pattern Anal Mach Intell* 2022; 44: 5149-5169. DOI: 10.1109/TPAMI.2021.3079209.
- [9] Gao S, Pei Z, Zhang Y, Li T. Bearing fault diagnosis based on adaptive convolutional neural network with Nesterov momentum. *IEEE Sens J* 2021; 21: 9268-9276.
- [10] Kingma DP, Ba JL. Adam: A method for stochastic optimization. 3<sup>rd</sup> Int Conf on Learning Representations (ICLR 2015) 2015: 1-13. DOI: 10.48550/arxiv.1412.6980.
- [11] Wang S, Yang Y, Sun J, Xu Z. Variational HyperAdam: A meta-learning approach to network training. *IEEE Trans Pattern Anal Mach Intell* 2022; 44: 4469-4484. DOI: 10.1109/TPAMI.2021.3061581.
- [12] Abdulkadrirov RI, Lyakhov PA, Nagornov NN. Accelerating extreme search of multidimensional functions based on natural gradient descent with Dirichlet distributions. *Mathematics* 2022; 10: 3556-2022. DOI: 10.3390/MATH10193556.
- [13] Dubey SR, Chakraborty S, Roy SK, Mukherjee S, Singh SK, Chaudhuri BB. DiffGrad: An optimization method for convolutional neural networks. *IEEE Trans Neural Netw Learn Syst* 2020; 31: 4500-4511. DOI: 10.1109/TNNLS.2019.2955777.
- [14] Zaheer M, Reddi SJ, Sachan D, Kale S, Research G, Kumar S. Adaptive methods for nonconvex optimization. *Proc 32<sup>nd</sup> Int Conf on Neural Information Processing Systems (NIPS'18)* 2018: 9815-9825.
- [15] Zhuang J, Tang T, Ding Y, Tatikonda SC, Dvornek N, Papademetris X, Duncan JS. AdaBelief optimizer: Adapting stepsizes by the belief in observed gradients. *Adv Neural Inf Process Syst* 2020; 33: 18795-18806.
- [16] Xie Z, Yuan L, Zhu Z, Sugiyama M. Positive-negative momentum: Manipulating stochastic gradient noise to improve generalization. *Thirty-eighth Int Conf on Machine Learning (ICML 2021)* 2021: 11448-11458.
- [17] Kurtansky NR, Dusza SW, Halpern AC, Hartman RI, Geller AC, Marghoob AA, Rotemberg VM, Marchetti MA. An epidemiologic analysis of melanoma overdiagnosis in the United States, 1975-2017. *J Invest Dermatol* 2022; 142: 1804-1811. DOI: 10.1016/J.JID.2021.12.003.
- [18] Turkay C, Lundervold A, Lundervold AJ, Hauser H. Hypothesis generation by interactive visual exploration of heterogeneous medical data. In Book: Holzinger A, Pasi G, eds. *Human-computer interaction and knowledge discovery in complex, unstructured, big data. Third international workshop (HCI-KDD 2013)*. Berlin, Heidelberg: Springer-Verlag; 2013: 1-12. DOI: 10.1007/978-3-642-39146-0\_1.
- [19] Wang S, Yin Y, Wang D, Wang Y, Jin Y. Interpretability-based multimodal convolutional neural networks for skin lesion diagnosis. *IEEE Trans Cybern* 2022; 52(12): 12623-12637.
- [20] Goh G, Carter S, Petrov M, Schubert L, Radford A, Olah, C. Multimodal neurons in artificial neural networks. *Distill* 2021; 6: 30. DOI: 10.23915/DISTILL.00030.
- [21] Liu K, Li Y, Xu N, Natarajan P. Learn to combine modalities in multimodal deep learning. *arXiv Preprint*. 2023. Source: <<https://arxiv.org/abs/1805.11730>>.
- [22] Lyakhov PA, Lyakhova UA, Nagornov NN. System for the recognizing of pigmented skin lesions with fusion and analysis of heterogeneous data based on a multimodal neural network. *Cancers* 2022; 14: 1819-2022. DOI: 10.3390/CANCERS14071819.
- [23] Banerjee. S. Estimation of body weight at different ages using linear and some non linear regression equations in a duck breed reared in hot and humid climate of Eastern India. *Am-Eurasian J Sci Res* 2011; 6(4): 201-204.
- [24] Tukey JW. The practical relationship between the common transformations of percentages or fractions and of amounts. In Book: Mallows CL, ed. *The collected works of John W. Tukey. Volume VI: More mathematical*. Pacific Grove, CA: Wadsworth & Brooks-Cole; 1990: 211-219.

### Сведения об авторах

**Ляхов Павел Алексеевич**, 1988 года рождения, кандидат физико-математических наук, заведующий кафедрой математического моделирования, факультет математики и компьютерных наук имени профессора Н.И. Червякова ФГАОУ ВО «Северо-Кавказский федеральный университет»; заведующий отделом модулярных вычислений и искусственного интеллекта, региональный научно-образовательный математический центр «Северо-Кавказский центр математических исследований». Область научных интересов: цифровая обработка сигналов и изображений, искусственный интеллект, нейронные сети, модулярная арифметика, параллельные вычисления, высокопроизводительные вычисления, цифровые устройства, аппаратные ускорители.

E-mail: [ljahov@mail.ru](mailto:ljahov@mail.ru)

---

**Ляхова Ульяна Алексеевна**, 1997 года рождения. Аспирант и младший научный сотрудник кафедры математического моделирования, факультет математики и компьютерных наук имени профессора Н.И. Червякова ФГАОУ ВО «Северо-Кавказский федеральный университет». Область научных интересов: цифровая обработка изображений, искусственный интеллект, обработка медицинских данных. E-mail: [uljahovs@mail.ru](mailto:uljahovs@mail.ru)

**Абдулкадиров Руслан Ибрагимович**, 2000 года рождения, студент Северо-Кавказского федерального университета с 2018 года по специальности «Прикладная математика и информатика», лаборант Северо-Кавказского центра математических исследований. Область научных интересов: машинное обучение, функциональный анализ. E-mail: [ruslanabdulkadirovstavropol@gmail.com](mailto:ruslanabdulkadirovstavropol@gmail.com)

---

ГРНТИ: 28.23.15

Поступила в редакцию 24 марта 2023 г. Окончательный вариант – 19 сентября 2023 г.

---

---

# Non-convex optimization with using positive-negative moment estimation and its application for skin cancer recognition with a neural network

P.A. Lyakhov<sup>1,2</sup>, U.A. Lyakhova<sup>1,2</sup>, R.I Abdulkadirov<sup>2</sup>

<sup>1</sup>North-Caucasus Federal University, 355009, Russia, Stavropol, Pushkin str. 1;

<sup>2</sup>North-Caucasus Center for Mathematical Research, 355009, Russia, Stavropol, Pushkin str. 1

## Abstract

The main problem of using standard optimization methods is the need to change all parameters in same-size steps, regardless of the behavior of the gradient. A more efficient way to optimize a neural network is to set adaptive step sizes for each parameter. Standard methods are based on the square roots of exponential estimates of the moments of the squares of past gradients and do not use the local variation in gradients. The paper presents methods of adaptive non-convex and belief-based optimization with a positive-negative estimate of the moments with the corresponding theoretical guarantees of convergence. These approaches allow the loss function to more accurately converge in the neighborhood of the global minimum in a smaller number of iterations. The utilization of transformed positive-negative moment estimates and an additional parameter that controls the step size allows one to avoid local extremes for achieving higher performance, compared to similar methods. The introduction of the developed algorithms into the learning process of various architectures of multimodal neural network systems for analyzing heterogeneous data has made it possible to increase the accuracy of recognizing pigmented skin lesions by 2.33–5.69 percentage points, compared to the original optimization methods. Multimodal neural network systems for analyzing heterogeneous dermatological data, using the proposed optimization algorithms, can be applied as a tool for auxiliary medical diagnostics, which will reduce the consumption of financial and labor resources involved in the medical industry, as well as increase the chance of early detection of pigmentary oncopathologies.

**Keywords:** optimization, natural gradient descent, artificial intelligence, multimodal neural networks, heterogeneous data, skin cancer, melanoma.

**Citation:** Lyakhov PA, Lyakhova UA, Abdulkadirov RI. Non-convex optimization with using positive-negative moment estimation and its application for skin cancer recognition with a neural network. *Computer Optics* 2024; 48(2): 260-271. DOI: 10.18287/2412-6179-CO-1308.

**Acknowledgements:** the authors thank the North-Caucasus Federal University for the award of funding in the contest of competitive projects of scientific groups and individual scientists of North-Caucasus Federal University. The research in section 2 was supported by the North-Caucasus Center for Mathematical Research under agreement with the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-02-2023-938). The research in section 3 was supported by the Russian Science Foundation (Project No. 23-71-10013).

---

## Authors' information

**Pavel Alekseyevich Lyakhov** (b. 1988) graduated from Stavropol State University, specialty "Mathematics" in 2009. PhD of Physical and Mathematical Sciences. Head of Mathematical Modeling department, North-Caucasus Federal University, Head of Modular Computing and Artificial Intelligence department, regional scientific and educational mathematical center «North-Caucasus Center for Mathematical Research». Research interests are digital signal and image processing, artificial intelligence, neural networks, modular arithmetic, parallel computing, high-performance computing, digital circuits and hardware accelerators. E-mail: [ljahov@mail.ru](mailto:ljahov@mail.ru)

**Ulyana Alekseevna Lyakhova** (b.1997). Ph.D. student and researcher of Mathematical Modeling department, North-Caucasus Federal University. Research interests: digital image processing, artificial intelligence, medical data processing. E-mail: [uljahovs@mail.ru](mailto:uljahovs@mail.ru)

**Ruslan Ibragimovich Abdulkadirov** (b. 2000) is a student of the North-Caucasus Federal University since 2018 with a degree in Applied Mathematics and Informatics, works as a laboratory assistant at the North-Caucasus Center for Mathematical Research. Research interests: machine learning, functional analysis. E-mail: [ruslanabdulkadirovstavropol@gmail.com](mailto:ruslanabdulkadirovstavropol@gmail.com)

---

Received March 24, 2023. The final version – September 19, 2023.

---