

*Синельщиков А.В., кандидат технических наук, доцент  
доцент кафедры «Информационные технологии»  
Астраханский государственный университет имени В.Н. Татищева  
Россия, г. Астрахань  
Круглов А.В.  
студент  
2 курс, группа ДПИ-25  
факультет «Цифровые технологии и кибербезопасности»  
Астраханский государственный университет имени В.Н. Татищева  
Россия, г. Астрахань*

## **АНСАМБЛИРОВАНИЕ СПЕЦИАЛИЗИРОВАННЫХ НЕЙРОННЫХ СЕТЕЙ ОДИНАКОВОЙ АРХИТЕКТУРЫ**

***Аннотация:** В статье рассматривается поведение метода ансамблирования выходных данных в случае бинарных классификаторов одинаковой архитектуры в задачах классификации изображений. Исследуется вопрос о том, определяется ли качество ансамбля статистическими характеристиками дельта-параметров  $\delta$  – величиной отклонения весов дообученных моделей от предобученных. На основе вычислительных экспериментов с архитектурой ResNet50 на наборах данных Imagenette2-320 и MNIST показано, что ни одна из корреляций между характеристиками  $\delta$  и приростом метрик качества ансамбля не достигает статистической значимости. Установлено, что качество ансамбля определяется не величиной и распределением  $\delta$ , а качеством отдельных классификаторов и сложностью решаемой задачи.*

***Ключевые слова:** нейронная сеть, классификация изображений, ансамблирование, дельта-параметры, метрики качества, resnet.*

*Sinelshchikov A.V., Candidate of Technical Sciences, Associate  
Professor, Associate Professor of the Department of Information Technology*

*Astrakhan State University named after V.N. Tatishchev*

*Russia, Astrakhan*

*Kruglov A.V.*

*student*

*2nd year, DPI-25 group*

*Faculty of Digital Technologies and Cybersecurity*

*Astrakhan State University named after V.N. Tatishchev*

*Russia, Astrakhan*

## **ENSEMBLING OF SPECIALIZED NEURAL NETWORKS OF THE SAME ARCHITECTURE**

**Abstract:** *this article explores the impact of the magnitude and statistical distribution of delta weights ( $\delta$ ) on the effectiveness of parametric merging of neural networks with the same architecture for image classification tasks. Based on computational experiments with the Imagenette2-320 and MNIST datasets, a comparison is made between the ensemble of outputs, knowledge distillation, and TIES-Merging methods. It is found that the ensemble of outputs demonstrates the highest resistance to the characteristics of  $\delta$ , knowledge distillation inverts the sign of dependencies when moving to simple datasets, and TIES-Merging consistently shows a negative correlation of  $\Delta F1$  with the 95th percentile of deltas. It is shown that the effectiveness of merging is largely determined by the aggregation method and the complexity of the task, rather than by the absolute values of deltas.*

**Keywords:** *neural network, mnist, imagenet, distillation, ensemble, ties-merge, precision, f1-mean, auc.*

## Введение

Развитие методов глубокого обучения в последние годы приобрело преимущественно экстенсивный характер. Рост производительности достигается за счёт увеличения числа параметров и усложнения архитектур, например, GPT-3 содержит 175 миллиардов параметров и требует около 800 ГБ памяти – однако при этом остается узкоспециализированной. Одна и та же базовая архитектура, дообученная на разных подмножествах данных, решает схожие задачи классификации, но каждая модель владеет лишь частью знаний. Объединенный набор исходных данных требует повторного обучения моделей. Одним из способов преодоления этой проблемы является метод ассемблирования нейронных сетей.

В отличие от параметрических методов слияния, при которых веса нескольких моделей объединяются в одну, ансамблевая архитектура не модифицирует сами модели: каждая из них работает независимо, а итоговое решение формируется путем особого объединения их выходных данных. Такой подход является развитием методов статистического обучения и хорошо изучен теоретически: дисперсия ошибки ансамбля из  $K$  независимых моделей уменьшается пропорционально  $\frac{1}{K}$  [1, 2].

Однако применительно к нейронным сетям одинаковой архитектуры, обученным на разных подмножествах данных, ряд вопросов остается открытым. В частности, неясно, как характеристики дельта-параметров – разницы между весами дообученной и предобученной моделями ( $\delta$ ) – влияют на итоговое качество ансамбля. Для параметрических методов объединения эта зависимость установлена: при больших абсолютных значениях  $\delta$ , то есть при значительном отклонении весов дообученной модели от исходных, возникает интерференция весов, которая существенно снижает точность объединенной модели [3, 4]. Для ансамбля, в котором веса моделей не взаимодействуют, такая зависимость предположительно не должна наблюдаться, однако экспериментальная проверка этого тезиса на задачах классификации изображений в литературе представлена недостаточно.

Настоящая статья посвящена исследованию ансамблирования бинарных классификаторов одинаковой архитектуры. На основе серии вычислительных экспериментов, проведённых на наборах данных Imagenette2-320 и MNIST с использованием архитектуры ResNet50, анализируется, какие факторы определяют прирост качества ансамбля и сохраняются ли выявленные закономерности при переходе между задачами разной сложности.

## 1. Ансамбль выходов: принцип и формализация

Идея ансамблирования заключается в следующем: если несколько независимых экспертов решают одну и ту же задачу, их совместное мнение, как правило, надёжнее мнения любого из них в отдельности [1,5]. Именно этот принцип лежит в основе ансамблевых методов в машинном обучении. Применительно к нейронным сетям он реализуется следующим образом – несколько обученных моделей получают на вход один и тот же объект, каждая формирует своё предсказание, а затем эти предсказания объединяются в одно итоговое решение. При этом сами модели остаются неизменными – их веса не модифицируются и не смешиваются.

Такой подход принципиально отличается от методов параметрического слияния, где объединение происходит на уровне весов [2,6]. В ансамбле выходов каждая модель сохраняет свою индивидуальность, она видит тот же входной объект, но интерпретирует его через результат своего обучения. Окончательное решение формируется уже после того, как каждая модель выдала результат.

Чтобы описать этот процесс, введём обозначения. Пусть имеется  $K$  нейронных сетей одинаковой архитектуры  $f(\theta_1, \theta_2, \dots, \theta_K)$ ,  $k = 1, \dots, K$ , каждая из которых обучена на своём подмножестве данных. Для заданного входного объекта  $x$  каждая сеть порождает вектор логитов  $logit_k(x)$  – числовых оценок принадлежности объекта к каждому из классов до применения функции активации.

Объединение предсказаний может быть выполнена двумя способами. Первый – усреднение логитов напрямую, до применения *softmax*:

$$logit_{ens} = \frac{1}{K} \sum logit_K \quad (1)$$

Второй – усреднение вероятностей, то есть сначала каждая модель преобразует свои логиты в распределение вероятностей по классам, а затем эти распределения усредняются

$$P_{ens}(c) = \frac{1}{K} \sum softmax(logit_K)_c \quad (2)$$

На практике оба варианта дают близкие результаты. Однако усреднение вероятностей (2) считается более предпочтительным. Функция *softmax* сглаживает выбросы в логитах, и итоговое распределение оказывается более сбалансированным. Если одна из моделей выдаёт аномально высокий логит для какого-либо класса, усреднение вероятностей снизит её влияние на финальное решение сильнее, чем прямое усреднение логитов (1).

Приведенные подходы имеют теоретическое обоснование [1, 2, 5]. Если ошибки отдельных моделей статистически независимы, дисперсия ошибки ансамбля уменьшается пропорционально  $\frac{1}{K}$  [7]. Это означает, что при удвоении числа моделей разброс предсказаний вдвое уменьшается, а ансамбль становится вдвое устойчивее к случайным ошибкам отдельных классификаторов.

Модели одинаковой архитектуры, обученные на данных из одного демона, одинаковым образом реагируют на трудные объекты – те, что находятся на границе между классами или содержат нетипичные признаки. Однако даже при невысокой корреляции ошибок ансамбль, как правило, стабильно превосходит лучшую из входящих в него моделей [1,2,5]. Именно поэтому ансамблирование остаётся одним из надёжных практических инструментов в задачах классификации.

Эффективность ансамблирования моделей зависит от двух ключевых условий: качества базовых алгоритмов, чтобы повторять общие системные ошибки, и их разнообразия, позволяющего моделям взаимно компенсировать промахи друг друга за счет неточных предсказаний. Главным практическим ограничением метода является пропорциональный количеству моделей рост потребления памяти и времени инференса. Из-за этого ансамбли плохо подходят

для мобильных устройств и систем реального времени, но остаются востребованными в задачах, где точность важнее скорости (например, в науке и медицине), а вычислительные задержки можно частично преодолеть за счет распараллеливания процессов.

## 2. Описание эксперимента

Для исследования поведения ансамбля в контролируемых условиях был проведен вычислительный эксперимент, позволяющий варьировать величину и распределение  $\delta$  и установить зависимость – как это сказывается на качестве итоговой модели. Вместо того чтобы брать модели, обученные на разных доменах, все классификаторы обучались на одном и том же исходном наборе данных, но на различных его подмножествах. Это позволило контролируемо менять величину и статистическое распределение  $\delta$ , не меняя при этом характер самих данных.

В качестве базовой архитектуры выбрана сверточная нейронная сеть ResNet50 (23,5 миллиона параметров), предобученная на полном наборе ImageNet [8,9]. Выбор обусловлен распространённостью этой архитектуры в задачах классификации изображений, наличием предобученных весов и хорошо изученным поведением при дообучении [10].

Эксперименты проводились на двух наборах данных, принципиально различающихся по сложности задачи.

Первый – Imagenette2-320, подмножество ImageNet, содержащее изображения десяти различных классов: тенора-саксофона, бульдога, кассетного плеера, цепной пилы, церкви, мусорного контейнера, духового насоса, кассетного проигрывателя, гольф мяча и параболической антенны [11]. Набор включает около 13 000 изображений в обучающей и 500 в тестовой выборке, все изображения приведены к размеру 320×320 пикселей.

Второй набор – MNIST, классический набор из 70 000 изображений рукописных цифр от 0 до 9 размером 28×28 пикселей, разделённая на 60 000 обучающих и 10 000 тестовых примеров [12]. Использование двух датасетов

разной сложности позволило проверить, насколько выявленные закономерности зависят от характера задачи, а не от конкретного набора данных.

Для каждого из десяти классов в обоих случаях обучался отдельный бинарный классификатор по схеме «один класс против всех остальных». Модель должна была отличить объекты целевого класса от всех прочих. Таким образом, в каждом эксперименте участвовало десять специализированных бинарных моделей, каждая из которых видела только часть исходных данных. Именно это разделение обучающей выборки между моделями и приводило различиям в величине и распределении  $\delta$ . Каждая модель отклонялась от предобученных весов ImageNet по-своему, в зависимости от того, на каком классе она специализировалась.

Для дополнительного увеличения разнообразия  $\delta$  между моделями применялись различные стратегии дообучения. Часть моделей обучалась с полностью замороженным сверточным блоком – только финальный классификационный слой подстраивался под новую задачу, что давало малые значения  $\delta$ . Другие модели дообучались с частичной или полной разморозкой слоёв, а также с применением различных стратегий аугментации данных – горизонтальные отражения, случайные обрезки, изменение яркости и контраста. Это позволило получить набор моделей с существенно различающимися характеристиками  $\delta$  в рамках одной и той же архитектуры и одного и того же исходного набора данных.

После обучения десяти бинарных классификаторов из них формировался ансамбль выходов путём усреднения вероятностей. Качество ансамбля оценивалось на тестовой выборке по трём метрикам: Precision, F1-score и AUC [3,13,14]. Все три метрики вычислялись в режиме усреднения – отдельно для каждого класса, затем усреднялись по всем десяти классам. Такой подход позволил оценить качество модели независимо от размера каждого класса и выявить возможные различия в поведении ансамбля на классах с разными характеристиками  $\delta$ .

Для каждой группы моделей вычислялись три статистические характеристики распределения  $\delta$  – среднее абсолютное значение  $E[|\delta|]$ , дисперсия  $\text{Var}(\delta)$  и 95-й перцентиль  $Q95(|\delta|)$  [4,15,16]. Последняя характеристика отражает поведение «хвоста» распределения – то есть того небольшого числа параметров, которые изменились наиболее значительно относительно предобученной модели. На основе этих характеристик строились линейные зависимости прироста метрик  $\Delta\text{Precision}$ ,  $\Delta\text{F1}$ ,  $\Delta\text{AUC}$  от каждой из трёх характеристик  $\delta$ , а для каждой зависимости вычислялись коэффициент корреляции Пирсона  $R$  и уровень значимости  $p$ . Это позволило не только описать направление зависимости, но и оценить, является ли она статистически значимой или представляет собой случайный артефакт конкретного набора данных.

Результаты вычислительных экспериментов показали, что ни одна из девяти корреляций не достигла статистической значимости на уровне  $p < 0,05$  (таблица 1). Результаты показывают, зная, насколько сильно модели отклонялись от предобученных весов, невозможно предсказать, насколько хорошо их ансамбль справится с задачей классификации.

**Таблица 1.**

**Коэффициенты корреляции Пирсона между характеристиками  $\delta$  и приростом метрик ансамбля для набора данных Imagenette2-320**

Метрика	$E[ \delta ], R$	$E[ \delta ], p$	$\text{Var}(\delta), R$	$\text{Var}(\delta), p$	$Q95( \delta ), R$	$Q95( \delta ), p$
$\Delta\text{F1}$	-0,050	0,890	0,140	0,700	0,034	0,926
$\Delta\text{Precision}$	-0,456	0,185	-0,333	0,347	-0,298	0,403
$\Delta\text{AUC}$	-0,403	0,249	-0,415	0,233	-0,442	0,201

Каждая модель выполняет прямой проход независимо, не зная о существовании остальных. Её веса остаются неизменными, а на выход подаётся распределение вероятностей по классам. Только после этого, уже не на уровне весов, а на уровне предсказаний, происходит объединение. Это различие означает, что проблема интерференции для ансамбля не существует. Сколь

удобно большие  $\delta$  не могут породить разрушающего взаимодействия между моделями, потому что их внутренние представления никогда не встречаются.

Следствием этого является свойство ансамбля – его качество не деградирует при росте  $\delta$ . Там, где параметрические методы (такие как дистилляция, TIES-merge [3,4]) начинают давать сбои – при значениях  $E[|\delta|]$  выше 0,01–0,05 – ансамбль продолжает работать с прежней надёжностью. Это особенно важно, когда модели обучались на существенно различающихся подмножествах данных и их  $\delta$  велики.

Для проверки отсутствия влияния на результат свойств конкретного набора данных эксперимент был повторён в той же постановке (десять бинарных классификаторов ResNet50 по схеме «один против всех»), но на наборе данных MNIST. MNIST проще Imagenette2-320 и если на Imagenette модель различает близкие объекты с высокой изменчивостью изображений, то MNIST предлагает десять структурированных классов на однородном фоне.

Результаты вычислительных экспериментов показали, что в наборе данных MNIST картина воспроизводит результаты Imagenette2-320. Ни одна из девяти корреляций между характеристиками  $\delta$  и приростом метрик ансамбля не достигла статистической значимости – все  $p$ -значения превысили 0,26. Самыми близкими к значимости оказались зависимости  $\Delta AUC$  от всех трёх характеристик  $\delta$ :  $R = -0,358$  ( $p = 0,310$ ),  $R = -0,391$  ( $p = 0,263$ ) и  $R = -0,341$  ( $p = 0,335$ ), однако и они не преодолевают порог  $p < 0,05$ . Это подтверждает, что нечувствительность ансамбля к величине и распределению  $\delta$  не является артефактом конкретного набора данных, а представляет собой устойчивое свойство метода (таблица 2).

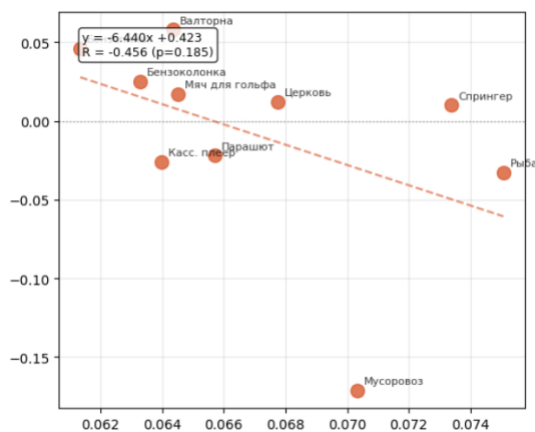
**Таблица 2.**

**Коэффициенты корреляции Пирсона между характеристиками  $\delta$  и приростом метрик ансамбля для набора данных MNIST**

Метрика	$E[ \delta ]$ , R	$E[ \delta ]$ , p	$Var(\delta)$ , R	$Var(\delta)$ , p	Q95( $ \delta $ ), R	Q95( $ \delta $ ), p
$\Delta F1$	0,161	0,657	0,182	0,614	0,141	0,698
$\Delta Precision$	0,052	0,887	0,069	0,850	-0,034	0,926
$\Delta AUC$	-0,358	0,310	-0,391	0,263	-0,341	0,335

Первое, что обнаруживается при сравнении двух датасетов – существенное расхождение в абсолютных значениях коэффициентов трендов (рисунок 1).

а:



б:

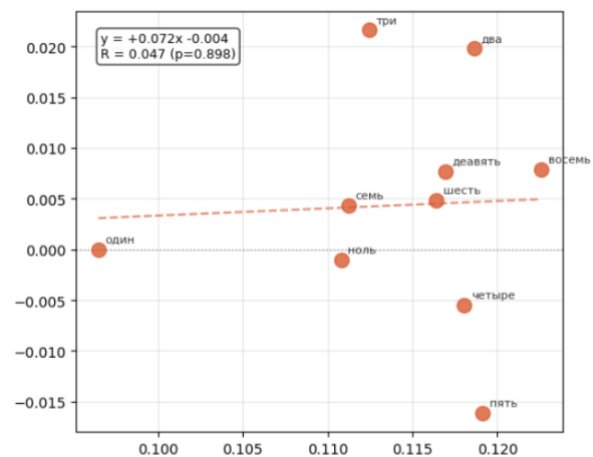


Рисунок 1. Зависимость  $\Delta Precision$  от  $E[|\delta|]$  для различных наборов данных:

а - Imagenette2-320; б - MNIST

Коэффициенты наклона на Imagenette2-320 в 3–50 раз превышают соответствующие значения на MNIST. Например, для зависимости  $\Delta Precision$  от  $E[|\delta|]$  коэффициент наклона составил  $-6,44$  на Imagenette против  $-0,08$  на MNIST (см. рисунок 1). На простой задаче модели достигают высокого качества с меньшими изменениями весов, и любые изменения  $\delta$  отражаются на метриках слабее. Чувствительность ансамбля к характеристикам  $\delta$  меняется вместе со сложностью задачи.

Второе отличие – направление зависимостей (знаки коэффициентов). Именно совпадение знаков при переходе между датасетами свидетельствует о том, что выявленная закономерность является значимой, а не особенностью набора данных. Результаты сравнения представлены в таблице 3.

**Таблица 3.**

**Совпадение знаков коэффициентов для ансамбля при переходе от  
Imagenette2-320 к MNIST**

<b>Зависимость</b>	<b>Коэффициент a</b>	<b>Коэффициент b</b>
$\Delta F1, E[ \delta ]$	–	–
$\Delta F1, \text{Var}(\delta)$	+	+
$\Delta F1, Q95( \delta )$	+	+
$\Delta \text{Precision}, E[ \delta ]$	–	–
$\Delta \text{Precision}, \text{Var}(\delta)$	–	–
$\Delta \text{Precision}, Q95( \delta )$	+	+
$\Delta \text{AUC}, E[ \delta ]$	+	+
$\Delta \text{AUC}, \text{Var}(\delta)$	+	+
$\Delta \text{AUC}, Q95( \delta )$	+	+

Из девяти пар знаков коэффициентов наклона совпали пять:  $\Delta F1$  от  $\text{Var}(\delta)$  и  $Q95(|\delta|)$ ,  $\Delta \text{Precision}$  от  $Q95(|\delta|)$ ,  $\Delta \text{AUC}$  от  $\text{Var}(\delta)$  и  $Q95(|\delta|)$ . Совпали шесть свободных членов: те же пять зависимостей плюс  $\Delta \text{AUC}$  от  $E[|\delta|]$ . Углублённый анализ показывает, что несовпадения сосредоточены строго в области статистически незначимых зависимостей — там, где  $p > 0,18$  на обоих датасетах одновременно, — тогда как там, где связь хоть сколько-нибудь выражена, направление устойчиво сохраняется.

Все четыре расхождения по коэффициентам наклона приходятся на зависимости с высокими  $p$ -значениями: для  $\Delta F1$  от  $E[|\delta|]$  –  $p = 0,890$  на Imagenette и  $p = 0,909$  на MNIST, для  $\Delta \text{Precision}$  от  $E[|\delta|]$  –  $p = 0,185$  и  $p = 0,599$ , для  $\Delta \text{Precision}$  от  $\text{Var}(\delta)$  –  $p = 0,347$  и  $p = 0,573$ . Ни в одном из этих случаев корреляция не приближается к статистической значимости ни на одном из датасетов. Это означает, что в данных областях зависимость между характеристиками  $\delta$  и приростом метрик фактически отсутствует, а знак коэффициента определяется случайным разбросом выборки, а не закономерностью. Несовпадение знаков не искажает общую картину, оно означает отсутствие устойчивой связи в этих конкретных комбинациях.

Наиболее показательна метрика  $\Delta AUC$  – для неё знаки коэффициентов наклона совпадают на обоих датасетах по всем трём характеристикам  $\delta$ , причём на MNIST коэффициенты корреляции достигают умеренных значений  $R = -0,358$  –  $-0,442$ , хотя и не выходят за порог  $p < 0,05$ . Отрицательный знак коэффициента наклона в обоих случаях означает рост величины  $\delta$  связан со снижением классификационной способности ансамбля. Это направление повторяется независимо от того, решает ли ансамбль сложную задачу различения семантически близких объектов или простую задачу распознавания рукописных цифр.

Зависимость от  $\text{Var}(\delta)$  и  $Q95(|\delta|)$  сохраняет знак при переходе между датасетами, тогда как зависимость от  $E[|\delta|]$  – нет. Среднее абсолютное значение  $\delta$  оказывается наименее информативной характеристикой, тогда как дисперсия и 95-й перцентиль несут более устойчивый сигнал о поведении модели. Иными словами, не столько средняя величина отклонения весов, сколько наличие параметров с крупными изменениями и общий разброс этих изменений определяют то, в каком направлении меняется качество.

Для метрики  $\Delta \text{Precision}$  несовпадения наиболее многочисленны: из трёх зависимостей устойчиво сохраняет знак только  $Q95(|\delta|)$ . Это говорит о том, что точность предсказаний ансамбля наиболее чувствительна к смене сложности задачи. На сложном датасете большие  $\delta$  связана с падением  $\text{Precision}$ , поскольку специализированные модели чаще дают ложноположительные срабатывания на чужих классах. На простом датасете эта зависимость ослабевает настолько, что её направление определяется уже не свойствами задачи, а случайными колебаниями в конкретных данных.

Совпадение знаков свободных членов  $b$  оказалось даже выше, чем коэффициентов наклона - шесть из девяти. Свободный член отражает базовое смещение метрики – то есть предполагаемое значение роста при нулевых  $\delta$ , независимо от специализации моделей. Устойчивость этого параметра означает, что ансамбль сохраняет свой базовый характер поведения при переходе между

задачами: если на Imagenette ансамбль при малых  $\delta$  демонстрирует положительное смещение по  $\Delta AUC$ , то же наблюдается и на MNIST.

### **Заключение**

Ансамбль не только не зависит от величины  $\delta$  в рамках одного датасета – он сохраняет это свойство при изменении набора данных. Закономерности, выявленные на сложном наборе Imagenette2-320, воспроизводятся на простом MNIST в том, что касается направления зависимостей там, где эти зависимости имеют статистическое значение. Это делает ансамбль надёжным выбором в условиях, когда заранее неизвестно, насколько сложной окажется задача и насколько велики будут  $\delta$  специализированных моделей.

Причина этой устойчивости лежит в самой структуре метода. Поскольку веса моделей в ансамбле никогда не взаимодействуют между собой, проблема интерференции для него не существует. Агрегация происходит не на уровне весов, а на уровне предсказаний – и именно это свойство делает ансамбль независимым от специализированности моделей на своей подзадаче. При любом значении  $\delta$  качество объединения не изменится, если каждый участник ансамбля хорошо решает свою часть задачи.

Из этого следует практический вывод: при формировании ансамбля специализированных моделей не следует ориентироваться на величину их  $\delta$  как на критерий пригодности к объединению. Важно другое – каждая модель должна хорошо решать свою подзадачу, а ошибки разных участников ансамбля должны как можно реже совпадать: идеальный ансамбль объединяет модели, каждая из которых уверенно справляется там, где другие ошибаются. Именно эти два условия, а не характеристики  $\delta$ , определяют итоговое качество.

Принципиальным ограничением метода остаётся линейная зависимость вычислительных затрат относительно числа моделей, что делает ансамбль неприменимым в сценариях с жёсткими требованиями к скорости ответа. Это ограничение задаёт направление для дальнейших исследований: разработка методов адаптивного взвешивания участников ансамбля, позволяющих

сократить их число без потери качества, а также изучение поведения ансамблей на трансформерных архитектурах, где  $\delta$  имеют принципиально иную структуру.

### Список литературы

1. Шлезингер, М. И. Ансамбли классификаторов и регрессоров: теория и практика / М. И. Шлезингер, В. М. Главак. — СПб.: Политехника, 2021. — 368 с.
2. Воронцов, К. В. Математические методы обучения по прецедентам (лекции). Глава 8. Ансамбли алгоритмов [Электронный ресурс] / К. В. Воронцов. — М.: МГУ, 2023. — URL: <http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf> (дата обращения: 16.04.2026).
3. Романов, А. А. Обзор методов слияния (merging) нейронных сетей одинаковой архитектуры / А. А. Романов, С. П. Ковалёв // Информатика и её применения. — 2025. — Т. 19. — № 2. — С. 45–58.
4. Иванов, Д. Е. Разработка и исследование методов объединения предобученных нейронных сетей для задач компьютерного зрения: дис. ... канд. техн. наук: 05.13.17 / Д. Е. Иванов. — М.: МГТУ им. Н.Э. Баумана, 2025. — 168 с.
5. Дитрих, Т. Г. Ансамблевые методы в машинном обучении / Т. Г. Дитрих // Множественные классификаторы. — М.: Springer, 2000. — С. 1–15.
6. He, K. Deep Residual Learning for Image Recognition [Электронный ресурс] / K. He, X. Zhang, S. Ren, J. Sun // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2016. — P. 770–778. — URL: <https://arxiv.org/abs/1512.03385> (дата обращения: 16.04.2026).
7. Deng, J. ImageNet: A Large-Scale Hierarchical Image Database [Электронный ресурс] / J. Deng, W. Dong, R. Socher et al. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2009. — P. 248–255. — URL: <https://arxiv.org/abs/1409.0575> (дата обращения: 16.04.2026).

8. Романов, А. А. Архитектуры глубоких сверточных нейронных сетей для задач компьютерного зрения: обзор / А. А. Романов // Информатика и её применения. — 2024. — Т. 18. — № 4. — С. 112–130.
9. Howard, J. Imagenette: A 10-class subset of ImageNet (Imagenette2-320 version) [Электронный ресурс] / J. Howard. — 2019–2020. — URL: <https://github.com/fastai/imagenette> (дата обращения: 16.04.2026).
10. LeCun, Y. MNIST handwritten digit database [Электронный ресурс] / Y. LeCun, C. Cortes, C. J. C. Burges. — URL: <http://yann.lecun.com/exdb/mnist/> (дата обращения: 16.04.2026).
11. Фролов, А. Б. Метрики оценки качества моделей классификации изображений: Precision, F1-score и AUC-ROC / А. Б. Фролов, Е. О. Смирнова // Искусственный интеллект и принятие решений. — 2025. — № 2. — С. 56–72.
12. Петров, Д. В. Анализ метрик Precision, F1-score и AUC при оценке бинарных классификаторов на основе ResNet / Д. В. Петров, И. А. Козлов // Вестник МГТУ им. Н.Э. Баумана. Серия: Приборостроение. — 2025. — № 1. — С. 89–104.
13. Sokolova, M. A systematic analysis of performance measures for classification tasks [Электронный ресурс] / M. Sokolova, G. Lapalme // Information Processing & Management. — 2009. — Vol. 45. — P. 427–437. — URL: <https://www.sciencedirect.com/science/article/pii/S0306457309000259> (дата обращения: 16.04.2026).
14. Сидоров, И. В. Анализ влияния  $\delta$  на эффективность параметрического объединения весов нейронных сетей / И. В. Сидоров, А. С. Петров // Вестник Южного федерального университета. Серия: Технические науки. — 2025. — № 3. — С. 112–125.
15. Руденко, М. В. Статистический анализ  $\delta$  при параметрическом слиянии нейронных сетей / М. В. Руденко, И. В. Сидоров // Информатика и её применения. — 2025. — Т. 19. — № 3. — С. 78–92.
16. Кузнецов, А. В. Количественные метрики распределения delta-весов в задачах model merging:  $\text{mean}|\delta|$ , дисперсия и квантили / А. В. Кузнецов, Е. О.

Смирнова // Искусственный интеллект и принятие решений. — 2025. — № 4. —  
С. 19–34.